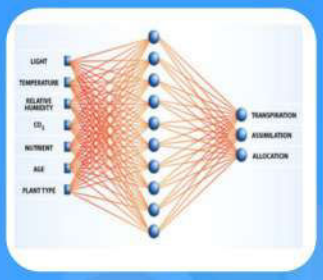
ISSN NO
2454-7719

International Journal of Computer Research & Technology (IJCRT)

Blind Peer Reviewed Half Yearly Research Journal

Vol -8 Issue -1
January-June,2023



Special Issue

Proceeding of the 8th National Conference on
"Emerging Trends in Computer and Information Technology"
held on 27th March 2023.

Editor in Chief
Dr.B.H.Barhate

।। विद्या दानम् महत् पुण्यम् ।।
Tapti Education Society's
Dept. of Computer Science and Information Technology
Bhusawal Arts Science & P.O.Nahata Commerce College,
Bhusawal - 425 201 Maharashtra
NAAc Reaccredited : 'A' Grade (*Fourth Cycle*) (CGPA:3.17)
UGC recognised ' College with Potential for Excellence for II nd Phase Effective from 2014 to 2019
Affiliated to Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon

IJCRT
ISSN 2454-7719
International Journal of Computer Research & Technology
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023

International Journal of Computer Research & Technology

** A Blind Peer Reviewed Journal **

Volume-8, Issue-1, Jan-Jun 2023

About Publisher of IJCRT

Bhusawal, as recalled and noted down in records has a prominent place on the map of the nation; proudly housing two ordnance factories, a thermal power station in the region, and itself being one of the major railway junctions of Central Railway from where, residents proudly say, you may visit any corner of India. A mixture of farmers, tribal people from adjoining are as with the servants from all over India, Bhusawal serves as a slice of the nation; and honorably has unity in diversity. It is 25kms away from the district, Jalgaon, famous as a city of gold; and few kms away from Yawal and Raver tehsils, famous all over nation for bananas. It is the only "A" graded Municipal Corporation in the district. Another identification as well as benefit of thecity is that it is situated at the bank of the Tapi river, the only river that flows from east to west. The city of Bhusawal has been ahome place for the British authorities, and it is famous for railways since British rule. It is historically remarkable for the grave of Major Robert Gill, who invented world famous Ajanta caves; and for the tomb of Sant Gadgebaba, a famous and truly a leading social reformer in Maharashtra. The world famous Ajanta caves are just 60kms away from the city. It is believed that the parental home of Rani Laxmibai (famous as Queen of Jhansi) is situated at Parola, 50kms away from the city. Bhusawal is also famous for many mythological stories like that of Shravana, coming from Ramayana who is said to be killed at Hartala, which is near to the city. Besides, the city was once famous in Bollywood for film distribution companies as well known "Rajashri" pictures.

Summing up the physiognomies of the city, Bhusawal stands as a glorious city in the eyes of everyone. However, it was the time-besides all assets of the city-when Bhusawal was a degenerated city in terms of higher education even after a long time from independence. There were few schools imparting high school level education but none of the colleges. It was only in1958, under the motivation of Late Honorable Madhukar rao Chaudhari, ex-speaker of Maharashtra Legislative Assembly, a groupof social well-wishers came together and established the Tapti Education Society in 1958. Simply having the wish in mind to provide potential students higher education facilities near their home, they started the Bhusawal College of Arts and Commerce in 1963. Their philanthropic view may be seen in the motto: *Vidya danam mahat punyam.* Yet difficulties were innumerable. The college with two faculties was started in the place of rent of a high school in the city.

It is wisely said that *vidya danam* is *mahat punyam*. The dedicated faculty, the sublime view of the management soon started to produce good academicians. Inspired by the results the trust purchased a barren land of 7 acres out of the city which is soon to be developed as a centre of imparting quality higher education in the area. The barren land with sustaining hard work, and devotion was then transferred into a naturally beautiful campus. The college is then shifted to a new place in 1972 with the introduction of Science stream. The philanthropist Late Mr. Poonamchand Nahata donated to the college, hence the college is renamed and which today itself is a brand as **Bhusawal Arts, Science and Poonamchand Omkardas Nahata Commerce College, Bhusawal.**

The college is then marching forward with a goal to **creatively contribute the society through the pursuit of learning at higher level of excellence**. The institute has contributed in many ways for economic, social and cultural uplift of the society by offering quality education. Since the inception it has been known for academic excellence, inventive pursuits and athletic dynamism. The college is a multi-stream institute catering to the needs of the young minds primarily from the rural areas.Our society runs not only the college but also the Institute of Management and Career Development and much sought Tapti Public School (affiliated to CBSE Board, New Delhi) within a minimum space of 7.3 acres. The institute is developing vertically in all of the fields.

The Tapti Education Society's, Bhusawal Arts, Science and P. O. Nahata Commerce College was accredited as **four stars (****) in 2001**, reaccredited as **"A" Grade** with CGPA **3.28 in 2008 and reaccredited 3rd cycle as "A" Grade with CGPA 3.30 in 2015,** as the **first College** in Kavayitri Bahinabai Chaudhari North Maharashtra University jurisdiction. It is also the first college to get register for the fourth cycle of accreditation in the jurisdiction of the university and reaccredited as **"A" Grade** with CGPA **3.17 in 2023 with new framework**. It is also recognized by UGC as **College with Potential for Excellence**. Recently, the society is certified as ISO 9001:2008 institute. Our institute is one of the renowned institutes in the adjoining area. We welcomed the upcoming students from rural areas who made remarkable progress and set theirs and college's image in society. Many of the students of this institute secure top position in various fields. This make us feel great. The college achieves 'A'grade in three subsequent cycles of Re-accreditations and it brings the college towards autonomous status.

Initially the college was affiliated to the Pune University, and got permanent affiliation in 1990. Since the inception of Kavayitri Bahinabai Chaudhari North Maharashtra University in 1991, the college is permanently affiliated to the same. The university spreads all over three districts: Jalgaon; Dhule; and Nandurbar, being on the boundaries of Gujrat and Maharashtra and one being the district of tribal people. The university is trying hard to uplift the downtrodden, while keeping in touch with the rapidly changing world.

Last but not least, the college has the advantages of developing youth coming from rural area, and forming them into sensible youth as they are mixed in the cosmopolitan society. The college is aware that every coin has two sides: hence students coming from rural areas have inferiority complex, their vernacular background being most disadvantage for them. The college has faced challenges to improve their communication skills, to boost their confidence to bring them into modern current while making them aware of great Indian culture. As the college has celebrated its golden jubilee, it will be a golden, in fact aplatinum moment for us when the students coming from different backgrounds will be essentially Indian serving for the welfare of humanity. With this view the college is making progress towards quality excellence so that it will be a lead college that will stand as a light house for the bewildered.

IJCRT
ISSN 2454-7719

International Journal of Computer Research & Technology

** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023

Published By
Department Of Computer Science and I.T,
TES's Bhusawal Arts,
Science and P. O. Nahata
Commerce College,
Bhusawal (M.S)

Contact: 9890966830
Email:
ijcrt.basponc@gmail.com
Website: **basponccollege.org**

The International Journal of Computer Research and Technology "IJCRT" is issued half yearly.

I N D E X

International Journal of Computer Research & Technology
** A Blind Peer Reviewed Journal **
ISSN 2454-7719
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-01

Pronoun – Antecedent Agreement For Anaphora Resolution In Marathi Language

Asst. Prof. Manoj Behere,
Head of the Department,
Department of MCA
IMRD, Shirpur
manrajdip@rediffmail.com

ABSTRACT - *In Natural Language Processing, Anaphora Resolution is used resolve anaphor with proper antecedent. This process is used in Information Retrieval, Question Answering, Machine Translation, Text Summarization in Marathi Language, Pronoun is one type of Anaphor and Noun is antecedent. The aim of this paper is to find Pronoun - Antecedent Agreement using Gender, Case, and Number in Marathi Language.*

Keywords -*Digital library, Research Student.*

I. INTRODUCTION

Anaphora resolution is mostly appears as pronoun resolution which can be resolving references to earlier items in the discourse. Basically resolution procedure consists of three steps: 1) anaphors identification 2) searching the probable antecedents, 3) Filtration using different criteria which consists of anaphora resolution factors by which identifying proper antecedent. For tracing the antecedent candidates, given text is parsed and identification of noun phrases is done. Every noun phrase is found either in the similar segment likewise the anaphor segment or within sentences nearby anaphor segment.

After finding possible antecedents, we can clean them using certain constraints to find correct antecedent. Anaphor and their antecedents must agree in number, person, and gender.

II. MARATHI LANGUAGE PRONOUN TYPES AND GRAMMAR CO REFERENCE

In Marathi Language , There are nine basic pronouns **मी(mi), तू(tu), तो(to), हा(ha), जो(jo), कोण(kon), काय(kay), आपण(aapan), स्वतः(swataa).** These basic pronouns are categorisedin six Marathi Pronoun Types as in Table 1.

Table 1: Marathi Pronoun Types

Marathi Pronoun Types	Marathi Pronouns
पुरुषवाचकसर्वनाम (Personal Pronouns)	मी, आम्ही, तो, तुम्ही, , तो, ती, ते, त्या , आपण, स्वत:
दर्शकसर्वनाम (Demonstrative Pronouns)	हा, ही, हे, तो, ती, ते
संबंधीसर्वनाम (Relative Pronouns)	जो, जी, जे, ज्या
प्रश्नार्थकसर्वनाम (Interrogative Pronouns)	कोण, काय
सामान्य / अनिश्चितसर्वनाम (Indefinite Pronouns)	कोण, काय
आत्मवाचकसर्वनाम(Reflexive Pronouns)	आपण, स्वत:

III. MARATHI LANGUAGE PRONOUNS AND GENDER RELATIONSHIP

In Marathi language , Three pronouns तो(to), हा(ha), जो(jo) are gender sensitive which has different masculine , feminine , Neuter forms and five pronouns मी(mi), तू(tu), तो(to), हा(ha), जो(jo) are number sensitive which has Singular and Plural forms as in table 2.

We observed that for Singular Pronouns are different as per masculine, feminine, Neuter But Plural Pronouns are same as per masculine, feminine,

Neuter for some Subject Pronouns like Tyans , Tyanche , Tyanchi ,Tyani , Tyana ,Hyans , Hyanche , Hyanchi ,Hyani , Hyana , Jyans , Jyanche , Jyanchi ,Jyani , Jyana

Table 2: Singular and Plural forms of pronouns in Marathi

Singular			Plural		
Masculine	**Feminine**	**Neuter**	**Masculine**	**Feminine**	**Neuter**
तो	ती	ते	ते	त्या	ती
त्यासत्याचा त्याचेत्याची त्यानेत्याला	तीसतिचा तिचेतिची तिनेतिला	त्यासत्याचे त्याचेत्याच्या त्यानेत्याला	त्यांसत्यांचा त्यांचेत्यांची त्यांनीत्यांना	त्यांसत्यांची त्यांचेत्यांची त्यांनीत्यांना	त्यांसत्यांचे त्यांचेत्यांची त्यांनीत्यांना
हा	हि	हे	हे	ह्या	हि
ह्यासह्याचा ह्याचेह्याची ह्यानेह्याला	हिसहिचा हिचेहिची हिनेहिला	ह्यासह्याचे ह्याचेह्याच्या ह्यानेह्याला	ह्यांसह्यांचा ह्यांचेह्यांची ह्यांनीह्यांना	ह्यांसह्यांची ह्यांचेह्यांची ह्यांनीह्यांना	ह्यांसह्यांचे ह्यांचेह्यांची ह्यांनीह्यांना
जो	जी	जे	जे	ज्या	जी
ज्यासज्याचा ज्याचेज्याची ज्यानेज्याला	जिसजिचा जिचेजिची जिनेजिला	ज्यासज्याचेह्याचेह्याच्या ह्यानेह्याला	ज्यांसज्यांचा ज्यांचेज्यांची ज्यांनीज्यांना	ज्यांसज्यांची ज्यांचेज्यांची ज्यांनीज्यांना	ज्यांसज्यांचे ज्यांचेज्यांची ज्यांनीज्यांना
मी			आम्ही&आपण		
माझा मला	माझी मला	माझे	आमचाआम्हास आपणालाआपला आपणासआपल्याला	आमचीआम्हास आपणालाआपली आपणासआपल्याला	आमचे
तू			तुम्ही		
तुझातुला Informal तुम्हालातुम्हास आपणाला Formal	तुझीतुला Informal तुम्हालातुम्हास आपणाला Formal	तुझे	तुमचातुम्हाला तुम्हासआपणाला	तुमचीतुम्हाला तुम्हासआपणाला	तुमचे

IV. ANALYSIS OF SAMPLE SENTENCES FOR PRONOUN – ANTECEDENT AGREEMENT FOR MARATHI LANGUAGE

Paragraph 1: ***राजहाइयत्ता४थीमध्येशिकतो. राजनेगीतगायनाच्यास्पर्धेतभागघेतला. राजचागीतगायनाच्यास्पर्धेतप्रथमक्रमांकआला . राजलातनयनेशाबासकीदिली . राजचेकौतुकतनयनेकेले.***

Paragraph1 consists 5 sentences. ***राज*** *is Masculine Singular noun in all sentences. We can replace Noun* ***राज*** *from sentence 2 to sentence 5 with Masculine, singular pronouns in Marathi language. After replacing pronouns in Paragraph1, we get*

Paragraph 2: ***राजहाइयत्ता४थीमध्येशिकतो. त्यानेगीतगायनाच्यास्पर्धेतभागघेतला. त्याचागीतगायनाच्यास्पर्धेतप्रथमक्रमांकआला. त्यालातनयनेशाबासकीदिली. त्याचेकौतुकतनयनेकेले.***

In Paragraph 2, Sentence 2 and sentence 3 there is Singular Neutor noun ***गीतगायन*** *and Sentence 4 and sentence 5 there is Singular Masculine noun* ***तनय****. We can replace Noun* ***गीतगायन*** *in sentence 3 with Singular Neutor Pronoun and Noun* ***तनय*** *in sentence 5 with Singular Masculine Pronoun in Marathi language. After Replacing pronous in Paragraph2, we get*

राजहाइयत्ता४थीमध्येशिकतो. त्यानेगीतगायनाच्यास्पर्धेतभागघेतला. त्याचात्यास्पर्धेतप्रथमक्रमांकआला . त्यालातनयनेशाबासकीदिली . त्याचेकौतुकत्यानेकेले.

Rule application on sample sentences:

Sentence 1: **तनयनेजयलामारले***Tanay ne Jayla Marale*

In the above sentence there are two nouns: *Tanay* and *Jay*. One of them can be replacing by a pronoun. Suppose we tried to replace *Tanay* (the subject) with a pronoun in Sentence 1, then we suppose we tried to replace *Tyane* a subject pronoun.

त्यानेजयलामारले*Tyane Jayla Marale*

Suppose we tried to replace *Jay* (the object) with a pronoun in Sentence 1, then we Suppose we tried to replace *Tyala,* an object pronoun.

तनयनेत्यालामारलेTanayneTyalaMarale

Sentence 2: **तनयनेरुचीलागृहपाठदाखविला***TanayneRuchilagruhpathdakhavila*

In sentence 2, *Tanay is* masculine subject noun which is *singular*, can be replaced with *Tyane*which is masculine subject pronoun of singular type.
त्यानेरुचीलागृहपाठदाखविला*TyaneRuchilagruhpathdakhavila*

Also in sentence 2, *Ruchi is* feminine object noun with singular type *can be replaced* with *Tila which is* feminine object pronoun having singular type
तनयनेतिलागृहपाठदाखविला*TanayneTilagruhpathdakhavila*
Sentence 3:तनयनेशाळेततनयचेभाषणदिलेTanayne*shalettanaychebhashandile*

InSentence 3, suppose we replace *tanay* with a pronoun *tyache. Then sentence will be*
तनयनेशाळेतत्याचेभाषणदिले*Tanayneshalettychebhashandile*
In this sentence, pronoun *tyache* refers to *tanayche.* The pronoun *tyache* is called the anaphor or referent *andTanayne* is antecedent as it appears before the pronoun and that refers to it later.

AntecedentReferent

Tanayneshalettychebhashandile.

Singular MasculineSingular Masculine

Selection of the correct pronoun for antecedents combined with and, or, or nor.

1. If singular noun antecedents are combined with and, then it form a plural antecedent.

Sentence 4: रणवीरआणिदीपिकानेत्यांच्यालग्नाचेफोटोदाखविले.

RanviraaniDipikanetyanchyalagnache photo dakhavile

Plural (1+1) Plural

RanviraaniDipikanetyanchyalagnache photo dakhavile

Antecedent Referent

Tyanchya which is plural pronoun replaces both masculine and feminine nouns.

Suppose both noun antecedents combined with and both noun antecedents are plural, then the referent pronoun must be plural.

Sentence 5: अनेकमुलांनीआणिमुलींनीसामूहिकविवाहसोहळ्यातत्यांचेलग्नकेले

AnekMulanniaaniMulinnisamuhikvivahsohalyattynchelagnakele

Plural Antecedent

Plural Referent

AnekMulanniaaniMulinnisamuhikvivahsohalyattynchelagnakele

2. Suppose antecedents are combined with or, then it will be form singular antecedent.

A. Two singular antecedents

Sentence 6: कीर्तीकिंवारुचीनेतिच्याभावालाबोलाविले
KirtikinvaRuchinetichyaBhavalaBolavile
Singular SingularSingular

KirtikinvaRuchinetichyaBhavalaBolavile

Antecedents Referent

The SINGULAR feminine pronoun replaces one of the feminine noun.

IshakinvaDipaknetyanchyaBhavalaBolavile.

B. Two plural antecedents

Sentence 7: मुलांनीकिंवामुलींनीत्यांच्याभावांनाबोलाविले

MullanikinvaMulinnityanchyabhavannaBolavile

Plural PluralPlural

MullanikinvaMulinnityanchyabhavannaBolavile

Antecedents Referent

Nouns representing groups are either singular or plural, which can be identified by meaning in individual sentences. Club, Company, Class will be describing either singular or plural.

Suppose we refer group as a whole or single unit then we consider the noun as a singular.

Sentence 8: क्लबरविवारीत्याच्यानियमानुसारबंदराहील
Club Ravivarityachyaniyamanusar band rahil.

Club Ravivarityachyaniyamanusar band rahil.

Single unit Single unit
Antecedent Referent

Suppose we are referring to the individual elements in the group, then we consider the noun plural. Hence we use a plural referent pronoun.

Sentence 9: क्लबमध्येयेतांनात्यांचेID-कार्डतपासलेजातील

Club madheyetanatyanche ID-Card tapasalejatil

Club madheyetanatyanche ID-Card tapasalejatil

More than one Plural referent pronoun

Individual plural Antecedent

Occasionally a personal pronoun may act as an antecedent.

In sentence 10, Tyane is the antecedent for the referent pronoun tyachya.

Sentence 10: त्यानेत्याच्याभावालामदतकेली*Tyanetyachyabhavalamadatkeli*

Tyachya is masculine and singular which will match with the masculine, singular antecedent *Tyane*

In sentence 11, Tine is the antecedent for the referent pronoun tichya.

Sentence 11: तिनेतिच्याभावालामदतकेली*Tine tichyabhavalamadatkeli*

tichya is both feminine and singular to agree with the feminine, singular antecedent *Tine.*

V. Pronoun - Antecedent Agreement using Gender, Case, Number in Marathi Language

After analysing, Marathi language pronoun and its type using some sample sentences we find following Pronoun - Antecedent Agreement rules.

1. A pronoun takes the place of a noun (Sometimes Pronoun)

2. Pronoun and noun agreed with following

a) A subject pronoun can be replacing a subject noun.

An object pronoun can be replacing an object noun.

b) A feminine pronoun can be replacing a feminine noun.

A masculine pronoun can be replacing a masculine noun.

c) A singular pronoun can be replacing a singular noun.

A plural pronoun can be replacing a plural noun.

These rules will help to find referent and its antecedent in resolution procedure for pronouns except interrogative pronoun and Indefinite Pronouns. For Interrogative Pronouns and Indefinite Pronouns, these rules are not suitable as Interrogative, Indefinite pronouns are not gender and number sensitive.

VI. Conclusion

For Marathi Language Rule Based Approach is most applicable as small set of rules are created to identify the antecedent of NPs of interest. For anaphora resolution process, it is essential to extract the pronouns and its proper antecedents. In this paper we studied pronoun and gender, case number relation of Marathi language and derived pronoun-antecedent agreement which is required for anaphora resolution process in Marathi language.

References

1. Imran Q. Sayed Issues in Anaphora Resolution, 2003.
2. Johan Winge, Pronominal Anaphora Resolution in NLP, Semantic analysis 2013.
3. Mitkov, Ruslan. “Discourse processing.” The handbook of computational linguistics and natural language processing (2010): 599-629.
4. SobhaLalitha Devi, Vijay Sundar Ram, Pattabhi RK Rao A Generic Anaphora Resolution Engine for Indian Languages -07 ISSN
5. Daya Shankar Yadav, Kamlesh Dutta, Pardeep Singh and PreetikaChandel Anaphora Resolution for Indian Languages: The State of the Art, CPUH-Research Journal: 2016, 1(2), 01

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-02

Application of Machine Learning Algorithms on historical Meteorological dataset for predictive solution and analysis, using modern data analytics tools and techniques.

Atul S. Patil
Assistant Professor,
Dept. of Computer Science & I.T.
JET's ZulalBhilajiraoPatil College, Dhule
atul_patils@yahoo.com

Dr. Bhojraj Hanumant Barhate
Professor,
Department of Computer Science, Bhusawal Arts, Sci and P.O.Nahata Comm. College, Bhusawal
bhbrama123@gmail.com

ABSTRACT:

Data Analytics is the scientific and statistical term for analysing raw data to renovate information for gaining knowledge. The role of analytics is to assemble, store, process and analyse data to address empirical methods in real world for decision making. Data can be in various formats, it can be in textual form, pictorial form, and audio visual form etc. This also involves acquiring knowledge from reliable data sources, rapidity in processing information, and future prediction and the main objective of this research work is to find the best prediction using machine learning tools and techniques. Machine Learning is a tool of data science which extracts meaning from data by transforming data into knowledge. Several algorithms are developed to learn patterns, acquire insights and do forecasting from previous historical datasets. For on-going research Linear Regression, Support Vector Model, Bayesian Linear Regression *algorithms are used for meteorological historical dataset (Annual Rainfall data of hundred years) with the modern tools and techniques for predictive solutions. For this we used Anaconda, Jupyter Notebook and Python as modern tools and techniques on the dataset for predictions of rain fall of upcoming years was carried out. This research is a basic attempt to go through meteorological historical datasets for extraction of patterns, and predictive solution using machine learning techniques.*

Keywords: *Data analytics, Machine Learning, Linear Regression, Support Vector Model, Bayesian Linear Regression algorithms, Anaconda, Jupyter Notebook, Python etc.*

I. INTRODUCTION

As we all Know and highlighted fact is that machine learning algorithms such as Linear Regression, Decision Tree, Random Forest and Gradient Boosting Tree etc. algorithms are basically used for analysing the various data sets. Machine Learning techniques are basically classified into supervised and unsupervised learning. Whereas supervised learning emphasises on specific predictions and unsupervised learning emphasises on concise descriptions of the data. Additionally data classification, complex pattern recognition, predictions, intelligent decisions and clustering are some of the specific characteristics of Machine Learning techniques. As we know several algorithms are developed to learn patterns, acquire insights and do forecasting from previous historical datasets. For on-going research it was discussed that Linear Regression, Support Vector Model, Bayesian Linear Regression algorithms can be used in current stage for meteorological historical dataset with the modern tools and techniques for predictive solutions.

As research progresses any suitable algorithm as well as techniques can be incorporated as per the need of research. In this context we reviewed various aspects of data analytics, machine learning tools and techniques in context with predictive solutions in terms of meteorological historical data.

II. IMPLEMENTATION

In the current stage of research we studied and tried to implement various machine learning algorithms such as *Linear Regression, Support Vector Model, Bayesian Linear Regression algorithms* for analysing the meteorological historical datasets.

Generally, data are represented in two forms (i) structured and (ii) unstructured. Structured data is the data arranged in tables and unstructured is the data with irregular form such as images, documents, text, audio and video. For on-going research in current stage we have collected and used dataset viz. "Seasonal and Annual Minimum / Maximum Temperature series from 1901 to 2019". [3] and dataset viz. "All-India-Rainfall-Act_Dep_1901_to_2019_0". [6]

Types of machine learning Algorithms: There some variations of how to define the types of Machine Learning Algorithms but commonly they can be divided into categories according to their purpose and the main categories are the following[1]:

- Supervised learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning

As mention above we used Linear Regression, Support Vector Model, Bayesian Linear Regression algorithms at current stage and tried to predict annual rainfall (in mm) forecast in upcoming year. For this we installed Anaconda. Anaconda is a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS. [4]

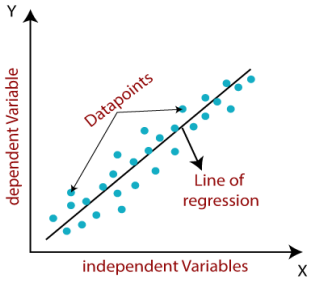
Where we used *Jupyter Notebook,* which is *a web-based interactive computing platform*. The notebook combines live code, equations, narrative text, visualizations etc. Jupyter or ipython notebook is a web application that allows you to run live code, embed visualizations and explanatory text all in one place. [5] We also used *Pandas Framework,* which is module of python.

Python is an innovative and cognitive programming platform for the modern generation. It is efficient in maintaining high level data structures, classes and object oriented approaches. Python has an elegant IPython console for mathematical and statistical computations. [7]. this is one of the nice platforms for training the Machine Learning algorithms.

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms. [13]

Linear Regression:

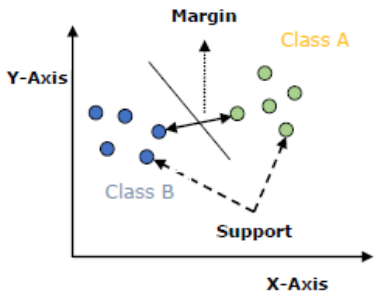
Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous real or numeric variables such as **sales, salary, age, product price,** etc. [2]



Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y (output). Hence, the name is Linear Regression. [10]

Support Vector Model (SVM):

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. [8]



An SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyper plane (MMH). [12]
The followings are important concepts in SVM −
Support Vectors − Data points that are closest to the hyper plane is called support vectors. Separating line will be defined with the help of these data points.
Hyper plane − As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.
Margin − It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin.

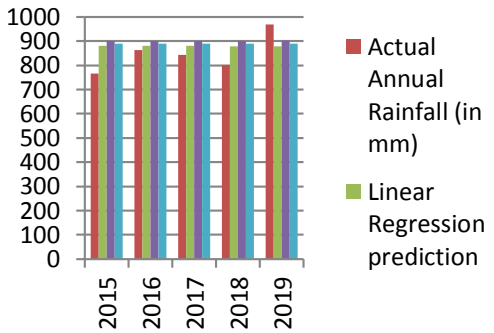
Bayesian Linear Regression*:*
Bayesian linear regression allows a fairly natural mechanism to survive insufficient data, or poor distributed data. It allows you to put a prior on the coefficients and on the noise so that in the absence of data, the priors can take over. More importantly, you can ask Bayesian linear regression which parts (if any) of its fit to the data is it confident about, and which parts are very uncertain (perhaps based entirely on the priors). [9]
Bayesian linear regression is a form of linear regression that uses the information about the variances of the input variables (whose errors have to be normally distributed as a precondition to ordinary linear regression anyway) to produce a probability distribution for outputs. While ordinary linear regression predicts a single output value on the basis of a range of input values, Bayesian linear regression predicts a probable value (which will be close to the single output value that the ordinary algorithm would have predicted) and a standard distribution around that probable value. [11]

III. RESULTS

For practical implementation and results of these three algorithms viz. *Linear Regression, Support Vector Model, Bayesian Linear Regressions,* Annual Rain fall data from year 1901 to 2019 more than 100 years dataset was used for training of these algorithms and predictions of rain fall of upcoming years was carried out. Results are as follows:

Year	Actual Annual Rainfall (in mm)	*Linear Regression prediction*	*Support Vector Model prediction*	*Bayesian Linear Regressions prediction*
2015	**765.4**	**879.89400014**	**899.78519978**	**890.07677649**
2016	**863.7**	**879.70677112**	**899.7490618**	**890.07468885**
2017	**843.7**	**879.51954209**	**899.78532308**	**890.07260121**
2018	**802.4**	**879.33231306**	**900.50276609**	**890.07051357**
2019	**969.4**	**879.14508403**	**902.13488665**	**890.06842593**

Following chart shows the comparative analysis between Actual annual rain fall and Linear Regression prediction, Support Vector Model prediction, Bayesian Linear Regressions prediction.



Going through the results it is found that results are quite satisfactory comparing with the actual annual rainfall. Some more machine learning algorithms can be more useful to expand with some other metrological factors to work on.

IV. CONCLUSION

Concluding current research work, it can be say that some more machine learning algorithms can be implemented for the available data set for prediction of annual rainfall as well other meteorological factors. Currently *Linear Regression, Support Vector Model, Bayesian Linear Regressions* algorithms were used to predict annual rainfall and comparing it with actual available dataset results are quite satisfactory; they are close nearby actual annual rainfall but can be reassessed with some specific algorithms.

REFERENCES

1. https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861
2. https://www.javatpoint.com/linear-regression-in-machine-learning
3. https://data.gov.in/catalog/all-india-seasonal-and-annual-temperature-series
4. https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)
5. https://www.youtube.com/watch?v=q_BzsPxwLOE
6. https://data.gov.in/catalog/rainfall-india
7. Ian D. S., Sasha Langford, (2008) 'Students' perceptions of python as a first programming language at wits' at ITiCSE

proceedings of 13th annual conference on innovation and technology in CSE pg: 365-365
8. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm
9. http://www.cs.cmu.edu/afs/cs.cmu.edu/project/learn-43/lib/photoz/.g/web/glossary/bayesian.html
10. https://www.geeksforgeeks.org/ml-linear-regression/
11. https://machinelearningcatalogue.com/algorithm/alg_bayesian-linear-regression.html
12. https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_support_vector_machine.htm
13. https://docs.python.org/3/tutorial/index.html.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-03

Computer Vision Based Classification System for Quality Inspection of Horticultural Products – A Review

Shubhangi K. Patil
Assistant Professor
Department of Computer Science,Dhanaji Nana Mahavidyalaya, Faizpur

Dr. Gouri M. Patil
Associate Professor,
Department of Computer Science, Bhusawal Arts, Sci and P.O.Nahata Comm. College, Bhusawal

ABSTRACT - *Computer vision based non-destructive inspection technology has found many uses in the agricultural and food industries, such as fruits and vegetables quality analysis, their grading and sorting. The quality rating of fruits and vegetables has a considerable impact on their production, pricing, and industries. This review comprehensively presents a systematic summary of the latest research articles based on computer vision in horticulture automation for classification and quality inspection of fruits and vegetables. Analyse the challenges in this field and explore the future scope along with the technologies used by many researchers with the classifiers accuracies. Through the analysis it is found that, deep learning based computer vision achieves promising results in the horticultural products identification, classification and quality analysis for large datasets whereas machine vision technology is more promising for small datasets. In future general and robust performance of agricultural automation systems can be obtained by using the computer vision and intelligent system for the real life scenario.*

Keywords -*Computer vision, classification, Grading, Sorting, quality, Agricultural Products*

I. INTRODUCTION

Agriculture is a backbone of global economy. Day by day the rise in population demands increase in agricultural products production and quality improvement. Grading agricultural goods helps the farmers to adhere quality criteria for their products, preventing them from cheating by dealers and earning a better price for their produce[1].Quality inspection of horticultural product plays a vital role for farmers as well as consumers. Fruit and vegetable variety has an impact on both the export market and quality assessment. The market value of vegetables and fruits is a crucial sensory element that influences customer preference and decision [13]. Recently, computer vision becomes a fast growing research field in the automation of agricultural application for analyzing the crops at pre harvesting and post harvesting. Computer vision has been widely employed for the fruit, vegetables and flowers quality inspection and grading. It presents the opportunity to automate the manual grading methods, thereby standardizing techniques and eliminating the time-consuming and error prone inspection tasks. Computer vision allows machine to visualize and understand the real life scenario like humans. Automatic grading and sorting of agricultural products based on artificial intelligent and computer vision is more efficient, more accurate and faster than the existing manual examination. In recent years many researchers work on classification and quality inspection of fruits, vegetables, flowers, herbs, nuts,

seeds etc. by using computer vision technology. This paper presents a detail review on different horticultural product datasets and classification models proposed by various researchers along with the accuracy achieved by them.

II. FRUIT, VEGETABLE CLASSIFICATION

Computer vision has been widely used in fruit grading system. Grading of a fruit is based on its color, size, shape, and texture features. Fruit screening for disease identification, yield prediction, grading are the applications which are automated widely by computer vision along with machine learning and deep learning technologies. As well as the vegetable plants are available in lots of different species with most varying nature. Each of the fruit and vegetable is having different nutritional components as per the different variety. Hence for the balance diet it is very important to classify them according to their variety and quality.

Most popularly CNN classifier is used by many researchers for the fruit and vegetable with different networks like AlexNet, EfficientNet ,VGG, ResNetfor large datasets [3],[4],[5],[8],[9],[10]. Also, YOLO is used in fruit recognition and classification system to obtain the region of interest [2],[5],[9],[11]. The machine learning classifiers SVM, KNN, LDA, Naïve Bayes, Random Forest are employed by the researchers for the quality analysis and classification of vegetables or fruit for the small dataset [6], [7], [12], [14].

Author and year	Application	Dataset	Classifier	Result
Suharjito , Franz A deta Junior et al.2023[2]	Ripeness grading of oil palm fruit	Image and Video dataset of 6 categories of oil palm fruit with the maturity level unripe, under-ripe, ripe, overripe, empty bunches and abnormal fruit	YOLOv4	0.99
Yu et al., 2023 [3]	Classification of 13 different varieties of Apples	Fruit-360	CNN model with AlexNet, VGG-19, ResNet- 18, ResNet-50, ResNet-101	All models achieved 96.1% accuracy for dataset A and 89.4–93.9% for dataset B. VGG-19 achieved the highest accuracy of 100.0% on dataset A and 93.9% on dataset B.
Nurrani et al., 2023 [4]	Classification of vegetable based on quality measures	The Pumpkin, Eggplant, Tomato, and carrot vegetables1021 images collected in supermarket withFresh, Withered, and Rotten class labels.	CNN-SVM classifier (SVM classifier employed to classify the feature extracted by the CNN fully connected layer with small dimensions.)	The 69.66% average accuracy of testing process and 92.51% accuracy of prediction process achieve by using CNN-SVM classifier.
Phan et al., 2023 [5]	Classification of tomato fruit based on three categories as ripe, immature, and damaged.	A tomato dataset with 1508 images captured at three different times of a day.	Yolov5m and Yolov5m combined with ResNet50, ResNet-101, EfficientNet-B0 deep learning framework	The combination of Yolo5m with ResNet-101 achieved 100% for ripe and immature tomatoes whereas for damaged tomatoes 94% accuracy achieved using

				Yolo5m with the Efficient-B0 model. The testing accuracies of the networks ResNet-50, EfficientNet-B0, Yolov5m, and ResNet-101 achieved 98%, 98%, 97%, and 97%, respectively
Saha et al., 2023[6]	Maturity based grading of star fruit with tree categories green mature, ripe, and overripe.	A dataset of 83 starfruit samples images with constant distance between camera and fruit and black background.	Linear Discriminant Analysis (LDA), Linear Support Vector Machines (SVM), Quadratic SVM, Fine K-Nearest Neighbor (KNN), and Subspace Discriminant Analysis (SDA).	By using Linear Discriminant Analysis (LDA) 96.2% accuracy achieved forcalibration and 93.3% for validation.
Natarajan and Ponnusamy, 2023 [7]	A multispectral sensor system for the classification of organic and conventional vegetables.	Multispectral sensor system is used to obtain10,800 spectral data of brinjal, green chilli and red tomato with damaged, diseased and spoiled class labels.	KNN, SVM, PCA, LDA, Neural network , Random Forest	A Neural network and random forest algorithm obtained 89% and 92% classification accuracy. By the fuzzy feature set generation andselecting best feature by the ACO algorithm 100% classification accuracy obtained.
Ismail and Malik, 2022 [8]	Real-time visual inspection system for grading apple and banana fruits.	A data set consists of 8791 apple images where 4693 defected apples images and 3946 healthy Images. The second dataset has 300 images of bananas with four categorized as unripe (green banana), yel-lowish green, mid-ripe and overripe.	CNN with ResNet, DenseNet, MobileNetV2, NASNet and EfficientNet architecture.	The 96.7% accuracy Obtained for apples and 93.8% for bananas by using EfficientNet model.
Mimma et al., 2022 [9]	An automated classification and detection of fruit system	A publically available FIDS-30 dataset consist of 971 images with 30 distinct categories of fruits and the second dataset containing 761 images with eight different classes of fruits, captured by	YOLOv3, ResNet50, VGG16	The 86% and 85% accuracies for the public dataset with ResNet50 and VGG16respectively. The 99% accuracy with ResNet50 and 98% accuracy with the VGG16 model on the custom dataset.

		smartphone camera.		
Morshed et al., 2022 [10]	Fruit classification and quality assessment	'FruitNet' dataset containing six different types of fruits19,526 images with tree types of quality (good,bad and mixed) means total 18 classes.	ResNet152 InceptionResNetV2 EfficientNetV2B0 VGG16 MobileNetV2 InceptionV3 Xception DenseNet201	The 99.67% accuracy obtained by using the CNN DenseNet model.
Fu et al., 2022 [11]	Fruit detection and freshness grading	The dataset consists of total 4000 images of six fruits classes collected from different locations with varying lightning condition and irrespective of noise and adjacent objects.	ResNet, VGG, and GoogLeNet. AlexNet, YOLO	AlexNet shows the 3.500 MSE and VGG shows lowest 3.934 error for validation.
Tata et al., 2022 [12]	Real-Time Quality Assurance of Fruits and Vegetables	The dataset created with around 2000 images for the fruit(orange, banana, and apple)and vegetable (bitter gourd, capsicum, and tomatoes)	KNN, SVM ,LR, CNN	By using CNN 95% of highest accuracy is achieved.
Li et al., 2021 [13]	Apple quality detection and classification	A dataset consist of 3600 original images of apple with three categories of premium, middle and poor grade.	CNN, Google Inception v3 model and HOG/GLCM+SVM	The 95.33% classification accuracy obtained by using the CNN model.
Narendra and Pinto, 2021 [14]	Defect detection in fruit and vegetable	A dataset of fifty fruits (apple and orange) and vegetable (tomato) images prepared.	Naïve Bayes Classifier	The 87%, 93% and 83% accuracy achieved for apple,orange and tomato respectively.
Faisal et al., 2020 [15]	Date fruit maturity level, type and weight estimation system	A dataset -1 consist of 8079 images of five date types and dataset-2 consist images of 152 bunches of 13 palm trees. Dataset-2 was developed for weight estimation.	ResNet, VGG-19, Inception-V3, and NASNet for maturity level estimation. The SVM regression is used for type and weight estimation.	For the date fruit type and maturity level estimation by using CNN ResNetarchitecture 99.175% and 99.058 % accuracy obtained respectively. Whereas for weight estimation 84.27% and 83.78% accuracy with SVM-Linear and SVM-Regression, achieved respectively
Li et al., 2020 [16]	Automatic recognition and	A dataset of ten categories of	VGG VGG-M	The 96.5% of accuracy obtained by

	classification of vegetables	vegetable image prepared with 48,000 images.	VGG-M-BN AlexNet	using the VGG-M-BN network model.
Femling et al., 2018 [17]	Identification of fruit and vegetable	A 4000 image dataset of 10 different types of fruits and vegetables including apple,avocado, banana, bell pepper, clementine, Kiwi, orange, pear, potato and tomato is prepared.	Convolutional Neural Network architectures (Inception and MobileNet)	The highest accuracy achieved by using MobileNet is 97%.

Table 1: Previous studies on different fruits/vegetables classification using computer vision

III. CONCLUSION

The use of computer vision along with deep learning technology in agriculture sector brought automation on a large scale which results in reducing the production cost and increase in profit. Thisstudy presents a comprehensively examined role of deep learning-based computer vision in the classification and quality analysis of different horticultural products.More specifically, we reviewed the literature based on computer vision for fruit classification depending on its maturity level or variety and different vegetables quality analysis. Most of the researchers used deep CNN model for the classification along with different architecturesAlexNet, VGG-19, ResNet, EfficientNet etc. whereas sometimes machine learning classifiers LDA, SVM, KNN, Naïve Bayes, Random forest are also used by the researchers and obtained higher accuracy for small datasets.

The study leads to the conclusion that deep learning based computer vision technology is more popular in the applications of agriculture as it produce the more promising, faster and reliable results on large datasets. The most challenging task to work in this field is lack of the availability of standard datasets hence mostly there need to create own dataset which is a tedious job. Also the dataset prepared by the researchers are in fixed environment which will limit the scope of work. Hence still there is a chance to work for the real life scenarios where the images are taken directly from the trees or fields, from the market places, bulk fruit images or the images captured at any time to deal with different illumination conditions. A robust system may also be designed for the classificationand quality analysis of horticultural products.

REFERENCES

1. H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, “Computer vision technology in agricultural automation —A review,” Inf. Process. Agric., vol. 7, no. 1, pp. 1–19, Mar. 2020, doi: 10.1016/j.inpa.2019.09.006.
2. Suharjito et al., “Annotated Datasets of Oil Palm Fruit Bunch Piles for Ripeness Grading Using Deep Learning,” Sci. Data, vol. 10, no. 1, p. 72, Feb. 2023, doi: 10.1038/s41597-023-01958-x.
3. F. Yu, T. Lu, and C. Xue, “Deep Learning-Based Intelligent Apple Variety Classification System and Model Interpretability Analysis,” Foods, vol. 12, no. 4, p. 885, Feb. 2023, doi: 10.3390/foods12040885.4] H. Nurrani, Andi Kurniawan Nugroho, and Sri Heranurweni, “Image Classification of Vegetable Quality using Support Vector Machine based on Convolutional Neural Network,” J. RESTI Rekayasa
4. Sist. Dan Teknol. Inf., vol. 7, no. 1, pp. 168–178, Feb. 2023, doi: 10.29207/resti.v7i1.4715.
5. Q.-H. Phan, V.-T. Nguyen, C.-H. Lien, T.-P. Duong, M. T.-K. Hou, and N.-B. Le, “Classification of Tomato Fruit Using Yolov5 and Convolutional Neural Network Models,” Plants, vol. 12, no. 4, p. 790, Feb. 2023, doi: 10.3390/plants12040790.
6. K. K. Saha et al., “Classification of starfruit maturity using smartphone-image and multivariate analysis,” J. Agric. Food Res., vol. 11, p. 100473, Mar. 2023, doi: 10.1016/j.jafr.2022.100473.
7. S. Natarajan and V. Ponnusamy, “Classification of Organic and Conventional Vegetables Using Machine Learning: A Case Study of Brinjal, Chili and Tomato,” Foods, vol. 12, no. 6, p. 1168, Mar. 2023, doi: 10.3390/foods12061168.
8. N. Ismail and O. A. Malik, “Real-time visual inspection system for grading fruits using computer vision and deep learning techniques,” Inf. Process. Agric., vol. 9, no. 1, pp. 24–37, Mar. 2022, doi: 10.1016/j.inpa.2021.01.005.
9. N.-E.-A. Mimma, S. Ahmed, T. Rahman, and R. Khan, “Fruits Classification and Detection Application Using Deep Learning,” Sci. Program., vol. 2022, pp. 1–16, Nov. 2022, doi: 10.1155/2022/4194874.
10. M. S. Morshed, S. Ahmed, T. Ahmed, M. U. Islam, and A. B. M. A. Rahman, “Fruit Quality Assessment with Densely Connected Convolutional Neural Network.” arXiv, Dec. 24, 2022. Accessed: Apr. 10, 2023. [Online]. Available: http://arxiv.org/abs/2212.04255
11. Y. Fu, M. Nguyen, and W. Q. Yan, “Grading Methods for Fruit Freshness Based on Deep Learning,” SN Comput. Sci.,

vol. 3, no. 4, p. 264, Jul. 2022, doi: 10.1007/s42979-022-01152-7.
12. J. S. Tata, N. K. V. Kalidindi, H. Katherapaka, S. K. Julakal, and M. Banothu, "Real-Time Quality Assurance of Fruits and Vegetables with Artificial Intelligence," J. Phys. Conf. Ser., vol. 2325, no. 1, p. 012055, Aug. 2022, doi: 10.1088/1742-6596/2325/1/012055.
13. Y. Li, X. Feng, Y. Liu, and X. Han, "Apple quality identification and classification by image processing based on convolutional neural networks," Sci. Rep., vol. 11, no. 1, p. 16618, Aug. 2021, doi: 10.1038/s41598-021-96103-2.
14. V. G. Narendra and A. J. Pinto, "Defects Detection in Fruits and Vegetables Using Image Processing and Soft Computing Techniques," in Proceedings of 6th International Conference on Harmony Search, Soft Computing and Applications, S. M. Nigdeli, J. H. Kim, G. Bekdaş, and A. Yadav, Eds., in Advances in Intelligent Systems and Computing, vol. 1275. Singapore: Springer Singapore, 2021, pp. 325–337. doi: 10.1007/978-981-15-8603-3_29.
15. M. Faisal, F. Albogamy, H. Elgibreen, M. Algabri, and F. A. Alqershi, "Deep Learning and Computer Vision for Estimating Date Fruits Type, Maturity Level, and Weight," IEEE Access, vol. 8, pp. 206770–206782, 2020, doi: 10.1109/ACCESS.2020.3037948.
16. Z. Li, F. Li, L. Zhu, and J. Yue, "Vegetable Recognition and Classification Based on Improved VGG Deep Learning Network Model:," Int. J. Comput. Intell. Syst., vol. 13, no. 1, p. 559, 2020, doi: 10.2991/ijcis.d.200425.001.
17. F. Femling, A. Olsson, and F. Alonso-Fernandez, "Fruit and Vegetable Identification Using Machine Learning for Retail Applications." arXiv, Oct. 23, 2018. Accessed: Apr. 12, 2023. [Online]. Available: http://arxiv.org/abs/1810.09811
18. C. Kuradeet al., "An Automated Image Processing Module for Quality Evaluation of Milled Rice," Foods, vol. 12, no. 6, p. 1273, Mar. 2023, doi: 10.3390/foods12061273.
19. D. Agarwal, Sweta, and P. Bachan, "Machine learning approach for the classification of wheat grains," Smart Agric. Technol., vol. 3, p. 100136, Feb. 2023, doi: 10.1016/j.atech.2022.100136.
20. Z. Fazel-Niari, A. H. Afkari-Sayyah, Y. Abbaspour-Gilandeh, I. Herrera-Miranda, J. L. Hernández-Hernández, and M. Hernández-Hernández, "Quality Assessment of Components of Wheat Seed Using Different Classifications Models," Appl. Sci., vol. 12, no. 9, p. 4133, Apr. 2022, doi: 10.3390/app12094133.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023

Paper ID: CS- 04

An Approach towards Unstructured Big Data Analysis Using HDFS and Map Reduce

V. P. Mahajan,
Asst. Prof. Smt. G.G. Khadse College, Muktainagar Dist. Jalgaon (MH)-India.
college Bhusawal
vpmahajan19@gmail.com

Prof. Dr. B. H. Barhate,
Vice Principal & Head of Dep. Info. Tech., Bhusaval Arts, Science and P.O. Nahata Commerce College Bhusaval (MH)-India.
bhbrama123@gmail.com

ABSTRACT - *The large data cannot be handled by the conventional database management system, then this data is called as big data. These big data can be three forms of data, structured form, unstructured form and semi structured form. Most of the part of big data is in unstructured form it is difficult to handle. The Apache Hadoop is an open source java based programming framework which provide better tools and techniques to handle huge amount of data. Hadoop and its components comprise of MapReduce and Hadoop Distributed File System (HDFS). MapReduce mechanism monitoring and execution of job using JobTracker and TaskTracker. HDFS a distributed file-system which encompass of NameNode, DataNode and Secondary NameNode for efficient handling of distributed storage purpose. In this paper, we presented our experimental work model on big data using the Hadoop distributed file system and the MapReduce. Analyzed the different usages by the Mappers and the reducers and variables for storage and processing of the data on a Hadoop cluster*

Keywords-*Hadoop, HDFS, MapReduce, JobTracker, TaskTracker, NameNode, DataNode*

I. INTRODUCTION

The data analytics Big Data contains both structured and unstructured data collected from various sources. An efficient tool is required for collecting, managing, storing and analyzing the large dataset. Open source framework Hadoop which processes large dataset. MapReduce in Hadoop is an effective programming model reduces the computation time of large scale database in a distributed architecture. [1]

To manage the huge volume of data, the proposed method will facilitates concurrent processing the data in small chunks in distributed clusters and aggregate all the data across clusters to obtain the final processed data. In Hadoop framework, MapReduce improves the reliability and speed of parallel processing and massive scalability of unstructured data stored on thousands of commodity servers. The functions of MapReduce is used to perform the task like grouping, filtering, sorting on data set, aggregation and to maintain the efficient storage structure [2][3].

Hadoop

Apache Hadoop Framework is store and process big data in a distributed environment. Hadoop's Architecture has two main components:

1. Distributed File System is used to store and process large datasets which runs on commodity hardware.
2. HDFS- (Hadoop Distributed File System) is similar to other existing distributed file system which breaks the large data into small blocks that are replicated and distributed across the cluster, however the most significant feature is highly fault-tolerant and can be deployed on inexpensive hardware.

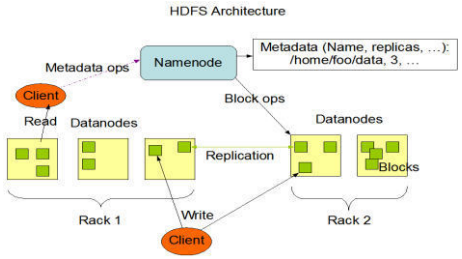


Image Source :
https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

In addition, HDFS architecture uses there are two types of nodes Master/Slave framework whereas an HDFS (Hadoop Distributed File System) cluster consists of a NameNode and a number of DataNodes. Metadata is stored in NameNode which acts as a master server responsible for managing namespace of file system. NameNode executes the namespace operation of file system, such as open, close, rename or directory, decides the mapping from data block to specific nodes and application data is stored in DataNode which acts as a slave server. DataNode is responsible for processing the read/write requests from file system client. Creating, deleting and copying the data block under the control of the NameNode. For reliability, the file content is duplicated on DataNode.

The HDFS architecture (Hadoop Distributed File System) and the MapReduce framework run on the same set of nodes because both storage and compute nodes are the same.

Due to this configuration, the framework can effectively schedule tasks on nodes that contain data, leading to support high aggregate bandwidth rates across the cluster.

HDFS accepts data in any format like text, images, videos, etc regardless of architecture and automatically optimizes for high bandwidth streaming. Hadoop-HDFS and MapReduce programming paradigm progressively being used for processing big and unstructured data sets. [4] [6]

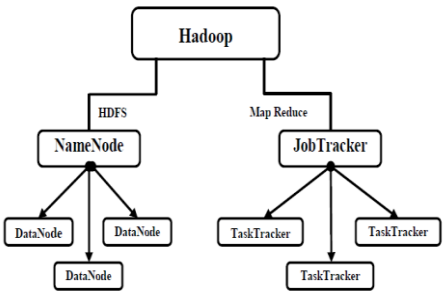


Figure: Hadoop Daemons

Master node can also play a role of slave. Thus, the master daemons, master node can run the slave daemons as well. Fundamentally, the daemons running on master node take responsibility for coordinating and managing the slave daemons on all nodes which carryout work for data storage and processing [5], [21-22].

MapReduce: MapReduce Framework is a programming model to process huge dataset in parallel in a distributed environment[7]. The huge volume of data is divided into many file blocks, distributedly stored in computer clusters. In distributed environment MapReduce is used for large datasets processing with fast realization and strong scalability.. [4][5].

Major phases involved in MapReduce paradigm, the Map function performs filtering and sorting, while the Reduce function carries out grouping and aggregation operations. The Map function splits the input data into fixed number of pieces to get evaluated in map phase. In the map phase, the data from a data blocks get split and key-value pairs are generated for the data[7]. The Reduce function is responsible for aggregating information received from Map functions, where data reduced into a single output value. The key-value pairs generated from the map function is passed as an input and clubs together the similar information in it [8][9].

II. BACKGROUND

Most of these data for big data analytics is unstructured data derived from various data sources and applications such as emails, images, text files, and social media posts, audio. Big data are meant to handle and manage unstructured data using key-value pairs [8].

Common Big Data Analytics Tools category

Traditionally, data and business analytics are performed using an integrated suite of machine learning and data mining algorithms [4]. Tools provide mechanisms to analyze small to large scale data for business decision-making process. The machine learning algorithms and tools for data analytics can be broadly categorized into:

i. **Clustering and segmentation**: Divides a large collection of entities into smaller groups that shows some similarities.

ii. **Classification** is a process of organizing data into predefined classes based on attributes that are either pre-selected by an analyst or identified as a result of a clustering model.

iii. **Regression** is used to discover relationships among a dependent variable and one or more independent variables and helps determine how the dependent variable's values change in relation to the independent variable values.

iv. **Association and item set mining** looks for statistically relevant relationships among variables in a large data set.

v. **Similarity and correlation**, which is used to inform undirected clustering algorithms. Similarity-scoring algorithms can be used to determine the similarity of entities placed in a candidate cluster.

Hadoop MapReduce programming worldview and HDFS are progressively being utilized for processing vast and unstructured data sets. Hadoop empowers collaborating with the MapReduce programming model [11]. A brief overview of the constituent technologies is now provided. The need for efficient, scale out solutions to support partial component failures and provide data consistency motivated the development of the File System [12] and the MapReduce [13]. The MapReduce is to distribute data across the commodity servers and this data is stored

as per the computation of data. This approach eliminates the necessity to transfer the data over the network to be processed. Moreover, methods for ensuring the resilience of the cluster and load balancing of processing were specified. HDFS [15] is the distributed storage component of Hadoop with master/slave architecture with participating nodes. All files stored in HDFS are split into blocks and these are replicated and distributed across different nodes called slave nodes on the cluster known as data nodes, where master node, called the name node which maintaining metadata e.g. blocks are certain storage which comprising a file, where in the cluster these blocks are located and so on. MapReduce [16] is the distributed compute component of Hadoop. MapReduce jobs are organized by the JobTracker. A job is full MapReduce program which complete the execution of Map and Reduce tasks on a dataset. The MapReduce paradigm is a master/slave architecture. The JobTracker runs on the master node and allocates Map and Reduce tasks to the slave nodes in the cluster. The slave nodes run another software daemon called the Task- Tracker that is responsible for actually instantiating the Map or Reduce tasks and reporting the progress back to the JobTracker. The extended ecosystem of Hadoop includes a growing list of solutions that integrate [17 The HBase is a distributed column-oriented database that resides on top of Hadoop's Distributed File System, providing real-time read/write random-access to big data[18]. Additionally, Hive defines a simple SQL-like query language, called HiveQL, to abstract the complexity of writing MapReduce jobs from users. Hive transforms HiveQL queries into MapReduce jobs for execution on a cluster (Apache Hive) [19]. The research on big data analytics and large scale distributed machine learning is very much in its infancy with libraries [10]. Nowadays, the massive volume of data are in an unstructured manner. It is very challenging to perform the operation in unstructured data. So the data need to be structured in order to perform some operations. Hadoop Map Reduce and collaborative filtering approach are used to structure the data and generates recommendations based on user preferences [20].

III. METHODOLOGY

Unstructured Text Data is the text written in various forms like – web pages, emails, chat messages, pdf files, word documents, etc. Hadoop framework was Map and reduce jobs, the end result is stored in Hadoop Distributed File System [HDFS]. Figure shows the execution flow of MapReduce process. Data with the use of advanced programming, we can find insights from this data [17]. Hadoop has distributed storage and distributed processing framework, and it is essential for unstructured data analysis, owing to its size and complexity.

the first designed to store and process this kind of Unstructured data is big, HDFS is stored as files. Hadoop does not carry out on having a schema or a structure to the data that has to be stored. Hadoop has different applications like Sqoop, HIVE, HBASE etc [19]. to make transactions from other popular traditional and non-traditional database forms. Hadoop is used for structuring any unstructured data and then exporting the semi-structured or structured data into traditional databases for further analysis. Apache Hadoop is a very powerful tool for writing customized codes. In the analysis of unstructured data typically involves complex algorithms. Efficiency and reliability are the benefits of the Hadoop framework. Hadoop being an open-source framework with numerous applications specific to video/audio file processing, image files analysis and text analytics [14].

A very large amount of unstructured data needs structural preparation for processing the data. Hadoop is binary compatible with Map reduce. Hadoop Map Reduce is a shuffling strategy which perform filtering and aggregation of data analysis tasks. Map technique is used for filtering the datasets. Reduce is a technique used for aggregation of data sets [21].

Knowing the volume associated with data which indirectly explores the velocity of data. The data flow from the machine or the result of human interaction or any other possible, applications. Next the variety, data may be from any source. The source may be structured data or unstructured data or a mixture of both, sources like excel sheets and databases provides structured data, in the sense unstructured data is in various other forms like video, audio, pictures [20].

Processing flow of MapReduce jobs

Map job involves three sub tasks such as Mapper, Combiner and Partitioner. Mapper involves the mapping of data, combiner combines the mapped data and partitions splits the data into small clusters, after which the shuffling key/value of map job to unique reduce job is done. Some studies show that short jobs, compose a large portion of MapReduce jobs

Reduce job involves two subtasks namely joiner and reducer. The joiner holds the joining of the intermediate results from the map jobs and reducer subtask is used for performing aggregation. After the

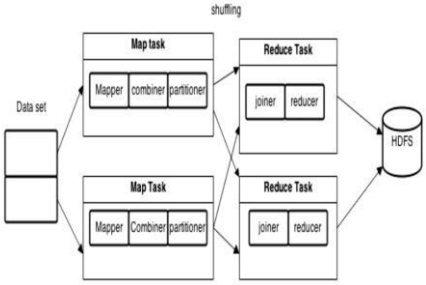


Figure: Execution Flow of a MapReduce process [2].

I) Datasets

The unstructured dataset is given as input to map-reduce process for the refinement and structuring.

II) Running Hadoop MapReduce

The Hadoop MapReduce program is supplied with the data sets as input and the MapReduce process is run for the 'N' number of data in the dataset. This process can be executed for inputs of any size. It supports fast and efficient processing of the data, by which unstructured data on any volume can be structured successfully. The output file generated by this MapReduce Technique has to be removed every time before running it in order to avoid File already exist exception.

III) HDFS Results

The Hadoop distributed File system output for a MapReduce job can be used to store the final results of map reduce process and the output can be viewed by browsing the file system in the name node log. The jobs that are completed while processing of the MapReduce process is given by job details log. NameNode log also contains information about the cluster summary, capacity of the file systems, distributed file system used and remaining and also the number of live nodes and dead nodes. The NameNode and JobTracker Details obtained as the result of the execution of Map Reduce process. NameNode log can be used to locate the output directory of the file system and the output of the map reduce job. Job Details log:

Log contains information about the kind of jobs and reports the user whether the job is completed or running or killed.

IV) Structuring of the unstructured data

After the successful execution of the map reduce program, the resulting data set is structured with a particular order according to the user requirements.

IV. CONCLUSION

Rapid growth of the data in organization both in structured and unstructured data. Hadoop is serves most excellent platforms used to store and retrieve the information in large volume at higher rate. Apache Hadoop is the solution for store and analysis the data with less cost for storage and fast processing time when compared with other technologies.

REFERENCES

1. Big Data Analytics in MapReduce: Literature Review Janani S, Dr D F X Christopher
2. Unstructured Data Analysis on Big Data using Map Reduce Subramaniyaswamy Va, Vijayakumar Vb, LogeshRc and Indragandhi Vd
3. Big Data Analytics in MapReduce: Literature Review Janani S, Dr D F X Christopher
4. Parallelizing K-means with Hadoop/Mahout for Big Data Analytics Jianbin Cui
5. Hadoop, MapReduce and HDFS: A Developers PerspectiveMohd Rehan Ghazia, Durgaprasad Gangodkara.
6. Challenges for MapReduce in Big Data, Katarina Grolinger1, Michael Hayes1, Wilson A. Higashino, Alexandra L'Heureux David S. Allison, Miriam A.M. Capretz.
7. J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," Commun ACM, 51(1), pp. 107-113, 2008.
8. Review- Big Data and Business Analytics: Trends, Platforms,Success Factors and Applications Ifeyinwa Angela Ajah * and Henry Friday Nweke Department of Computer Science, Ebonyi State University, P.M.B 053, Abakaliki 480214, Nigeria;
9. Survey on Big Data by Coordinating Mapreduce to Integrate Variety of Data N. Monica, K. Ramesh Kumar.
10. A big data methodology for categorising technical support requests using Hadoop and Mahout Arantxa Duque Barrachina and Aisling O'Driscoll.
11. REVIEW PAPER ON HADOOP AND MAP REDUCE , Nishant Rajput1, Nikhil Ganage2, Jeet Bhavesh Thakur3, IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308.
12. Ghemawat S, Gobioff H, Leung S (2003) The Google File System. In: SOSP '03 Proceedings of the 19th ACM symposium on Operating Systems Principles., vol 6. ACM, pp 10-10. http://dl.acm.org/citation.cfm?id=945450.
13. Dean J, Ghemawat S (2004) MapReduce: Simplified Data Processing on Large Clusters. OSDI' 04 Proceedings of the 6th symposium on Operating System Design and Implementation, San Francisco, CA Vol 6. pp 10-10. http://dl. acm.org/citation.cfm?id=1251264.
14. Apache Hadoop (2012), http://hadoop.apache.org/
15. Shvachko K, Kuang H, Radia S, Chansler R (2010) The Hadoop Distributed File System. In: IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST' 10). ACM pp 1–10. http://dl.acm.org/citation.cfm?id=1914427.
16. Bhandarkar M (2010) MapReduce programming with apache Hadoop. In: IEEE 24th International Symposium on Parallel & Distributed Processing (IPDPS' 10)
17. Apache Mahout (2012), https://mahout.apache.org/
18. Apache Hbase (2012), http://hbase.apache.org/
19. Apache Hive (2012), http://hive.apache.org/
20. Unstructured Data Analysis on Big Data using Map Reduce Subramaniyaswamy Va, Vijayakumar Vb, Logesh Rc and Indragandhi Vd
21. T. White, "MapReduce and the hadoop distributed file system", in Hadoop: The definitive guide, 1st edition, O'Reilly Media, Inc., Yahoopress, 2012.
22. A.Elsayed, O. Ismail, and M.E. El-Sharkawi, "MapReduce: state-of-the-art and research directions", International Journal of Computer and Electrical Engineering, vol. 6, no. 1, pp. 34-39, Feb. 2014
23. Digital Libraries Impact on Students' Learning Behaviour. Case Study: Medical Students by Rodica Volovici in Qualitative and Quantitative Methods in Libraries (QQML) 5: 23-29, 2016
24. 2. Digital library as learning resources by R Ilahi in Journal of Physics: Conference Series , 1402 (2019) 077044, IOP Publishing, doi:10.1088/1742-6596/1402/7/077044
25. S. C. Gupta and V.K.Kapoor (2006). Fundamentals of Mathematical Statistics, Sultan Chand & Sons Educational Publishers, New Delhi
26. S. C. Gupta and V.K.Kapoor (2006). Fundamentals of Applied Statistics, Sultan Chand & Sons Educational Publishers, New Delhi

International Journal of Computer Research & Technology
** A Blind Peer Reviewed Journal **
ISSN 2454-7719
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-05

The Effectiveness of Block Chain and the Future of the Internet: Overview

Dr. Kishor Mahajan,
Assistant Professor, KES's B K Shroff Arts and Commerce College, Kandivli, Mumbai, MS(India).
kishor@kessc.edu.i

Miss. Aamina Qureshi,
Assistant Professor, KES's B K Shroff Arts and Commerce College, Kandivli, Mumbai, MS(India).
amina@kessc.edu.in

ABSTRACT *–Blockchain could be a data structure that could be an upgrade list of information blocks. The knowledge blocks area unit combine along, such recent blocks can't be removed or altered. Blockchain is the backbone Technology of Digital Cryptocurrency, Bitcoin. Blockchain Technology had the foremost impact on our lifestyles within the last decade. Many of us still confuse Blockchain with Bitcoin, however they're not a similar. Bitcoin is associate application that uses Blockchain technology. Cloud computing is more about delivering computing services, comprising servers, databases, storage, etc. Cloud computing uses blockchain peculiarities to increase data security. There's a large spectrum of blockchain applications starting from cryptocurrency, risk management, Internet of things (IoT), monetary services to public and social services. Blockchain technology has shown its wide ability in recent years as a spread of market sectors need ways in which of integration its potentials into their operations. Blockchain has ton of advantages like decentralization, determination, obscurity and auditability. Though different studies specialize in usage of blockchain technology in numerous application aspects, there's no comprehensive survey on the blockchain technology in each application views and different perspectives. To fill this difference gap, we have a tendency to conduct a comprehensive study on blockchain technology, blockchain scenario, blockchain application review, Cloud, Securities and technical challenges that improve performance. Moreover, this paper additionally sorts out the longer-term side of blockchain technology.*

Keywords *– Blockchain, Cryptocurrency, Cloud Computing, IoT, Security, Scalability.*

I. INTRODUCTION

Blockchain is term that trend in future scope and future of the internet. Blockchain is an especially promising and revolutionary technology because it helps to reduce security risks, eliminate fraud, and bring transparency to a scale that has never been seen before. It was originally associated with cryptocurrency and non-fungible tokens, but blockchain technology has since evolved into a management solution for all types of global industries. Blockchain, although unnamed as such at that point, was introduced to the world in a whitepaper, in 2008, its utilisation in the advanced shared money framework, Bitcoin. Bitcoin is a type of organisation convention, similar to HTTP or TCP layers, which supports worldwide web framework and utilised each time we peruse the worldwide Web. A blockchain is a record of advanced exchanges, it is decentralised and not heavily influenced by any individual, gathering or organisation. The blockchain technology is organised and it is difficult to change the guidelines or its substance without the agreement among the individuals who are utilising it. In blockchain more uptodate impedes are connected to the more seasoned ones, shaping a chain, subsequently the term blockchain. This structure guarantees that lone the passages can be included information base, information can never be changed or taken out because changing a solitary section in a more established square would mean modifying the whole history of exchanges resulting to that block. All the

more explicitly, blockchain is an unchangeable, shared record of companion peer exchanges put away in a computerised record which is made from connected exchange blocks.

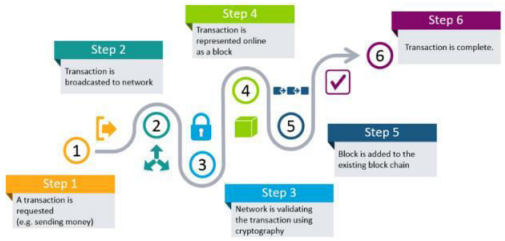


Fig: Blockchain Technology-How to works

Blockchain, is a technology that safely keep up consistently developing arrangements of information records and exchanges. Blockchain depends on set up methods of cryptography to permit every one of the members of an organisation to cooperate for store, trade, and view data. In a blockchain framework, there is no brought together position; rather, than it, exchange records are put away and dispersed over the organisation. Above all, all information passages are stepped with date and time. Associations with the blockchain medium become known to all members and require confirmation by the organisation before adding the data, empowering trust less joint effort between network members while recording an unchangeable review trail of the obvious multitude of communications. For security, clients can refresh just the square to which they have the entrance, and those updates get reproduced over the organisation. Blockchain, though not named per se at the time, was conferred to the planet during, its use within the digital peer-to-peer currency system, Bitcoin. Bitcoin could be a variety of network protocol, like HTTP or protocol layers that underpin international web infrastructure and used on every occasion we have a tendency to browse the planet Wide internet. The blockchain technology is structured and its extraordinarily troublesome to vary the principles or its content while not the agreement among the those who square measure victimisation it. In blockchain newer blocks square measure coupled to the older ones, forming a sequence, so the term blockchain. Bitcoin is that the terribly 1st application of blockchain, it's a form of digital currency supported blockchain. thanks to the success of Bitcoin, individuals currently will utilize blockchain technologies in several field and services, like money market, IOT, offer chain, election balloting, medical treatment, document handling and pursuit in office, insurance pursuit record and cybercriminals. We use these tools or services in our way of life, cybercriminals and cybercrime may be eradicated through this blockchain technology.

II. THE CONCEPT OF BLOCKCHAIN

Blockchain technology is not using one single technique but contains Cryptography, mathematics, Algorithm and economic model, combining peer-to-peer networks and using distributed consensus algorithm to solve traditional distributed database synchronization problem. The following six key elements of blockchain are:

I. Decentralized: Blockchain doesn't have to rely on centralized node anymore, the data can be recorded, stored and updated distributively.

II. Anonymity: Blockchain technologies solve the faith problem between one node to other node, so data transfer can be unidentified, only person's blockchain address need to know.

III. Autonomy: The blockchain solely works according to the rules which are defined by its members. There is no central authority for the defined rules.

IV. Security: There are various ways which proves a blockchain is more secure than other record-keeping systems. Transactions must be agreed upon before they are recorded into the system. Once a transaction is approved, it is encrypted and then linked to the previous transaction. In any trade wherever the protection of sensitive information is crucial money services, government, care blockchain has a chance to vary however the vital data is shared by helping to prevent frauds and unauthorized activity.

V. Transparency: The data's record by blockchain system is transparent to each node, it is also transparent on update of data that is why blockchain can be trusted. Changes to public blockchains are publicly viewable by all parties creating transparency, and all transactions are unchangeable.

III. BLOCKCHAIN ARCHITECTURES

There are different kinds of blockchain architecture, each of them has different design and architecture.

A. Public Blockchain

In such blockchain, everybody in the organisation can approve the exchange and can partake during the time spent achieving agreement. It guarantees decentralisation by setting up a square of distributed exchanges. Every exchange is joined with the blockchain before it goes to the framework. Subsequently, it very well may be affirmed and adjusted with each hub in the organisation. Anyone with a PC and web association can be selected as a hub and can be given the total blockchain history. It states that everyone can check the exchange and confirm it, and can likewise partake during the time spent

getting agreement. Instead, a company or organization initiates, verifies and validates each transaction. This gives a higher level of efficiency in the verification and validation of transactions. The benefit of private blockchain is that a company can select the access rights to individuals and permit a higher level of privacy when compared with public blockchains.

B. Private

A private blockchain is suitable to a traditional and governance model-based business. Using a privately-run version of blockchain can bring the organization into the current century. Private blockchains are more prone to acceptability by the private sector or government-based companies as they allow a central authority to be present with a more secure, more efficient and faster technology.

C. Consortium Blockchain

Consortium blockchain is a combination of public and private blockchain and can be interpreted as partly decentralized. These blockchains are open to public but not the entire data is available to all the participants. User rights vary and blocks are validated based on the predefined rules. Consortium blockchains are hence "partly decentralised". Consortium Blockchains are the ones in which the consensus process is controlled by a preselected set of trusted nodes. A block is added to the chain after consensus is achieved through the transaction validation by a group from the preselected set of nodes. In a consortium Blockchain, the right of reading the blockchain can be public or made restricted only to participants. In addition to this, consortium Blockchains are considered to be partially decentralized unlike private Blockchains.

IV. BLOCKCHAIN TECHNOLOGY APPLICATIONS

I. Internet of Things (IoT)

In an IoT biological system, the greater part of the correspondence is as like Machine-to-Machine (M2M) cooperation. Hence inaugurating trust between the partaking machines is that the major challenge that IoT innovation still has not been met wide. Nonetheless, Blockchain might set about as associate degree impetus in such manner by empowering upgraded ability, security, dependableness, and confidentiality. This can be accomplished by conveying Block chain innovation to follow billions of gadgets associated with the IoT eco frameworks and used to empower and additionally arrange the exchange processing. Relating Block chain within the IoT circle can likewise build reliability by surgical operation the only purpose of Failure. The cryptologic calculations used for cryptography of the block info even as the hashing strategies might provide higher security. In any case, this will request all the more processing force which IoT gadgets presently experience the ill effects of. In this manner, supplementary investigation is obligatory to defeat this contemporary restriction. A portion of the instances of block chain IoT are Smart Appliances, Supply Chain Sensors and so on.

II. Finance

- Financial services. The emergency of blockchain systems like Bitcoin and has brought an enormous impact on ancient monetary and business services. Blockchain has the potential to disrupt the globe of banking. Blockchain technology may well be applied to several areas together with clearing and settlement of monetary assets etc.

- Risk management

Risk management framework plays a major role in monetary technology (FinTech) and currently it may be combined with blockchain to perform higher. Pilkington (Pilkington, 2016) provided a unique risk management framework, during which blockchain technology is employed to analyse investment risk within the Luxembourgish situation. Investors who nowadays hold securities through chains of custodians tend to face the risk of any of these failings.

V. SECURITY AND PRIVACY OF BLOCKCHAIN

Security ideas and principles are listed below:

A. Défense in Penetration: This is a strategy which uses numerous corrective measures to protect the data. This principle of protection of data in multiple layers is more efficient than single security layer.

B. Minimum Privilege: In this strategy data access is reduced to the lowest level possible to reinforce elevated level of security.

C. Manage vulnerabilities: In this strategy we check for vulnerabilities and manage them by identifying, authenticating, modifying and patching.

D. Manage Risk: In this strategy we process the risks in an environment by identifying, assessing and controlling risks.

VI. CHALLENGES OF BLOCKCHAIN

A challenge can be defined as an implicit demand for proof. Some of the major challenges currently faced by blockchain technology are listed as below.

A. Scalability: It provide better scalability in number of ways that provides blockchain with cloud using effective utilization of internet. With regular volume growth of blockchain utilization and therefore the

come the sheer range of exchanges on a daily basis, the blockchain is ending endlessly stupendous in size. All transactions are stored in each and every node to get validated. The current transaction should be validated first before the other transactions to be validated. Meanwhile, the size of the blocks in blockchain may create an issue of transaction delay in the event of little transaction, as diggers transaction fees, miners would prefer to validate transactions. The adaptability issue of blockchains can be categorized in two classes: storage optimization and redesigning of blockchains.

B. Privacy Leakage: The blockchain is mainly vulnerable to transactional privacy leakage due to the fact that the details and balances of all public keys are visible to everyone in the network. The proposed solutions for accomplishing anonymity in blockchains can be extensively classified into mixing solution and anonymous solution. Mixing is a service that offers anonymity by transferring assets from numerous info delivers to various yield addresses.

C. Selfish Mining: Selfish mining is another challenge faced by blockchain. A block is susceptible to cheating if a small portion of hashing power is used. In selfish mining, the miners keep the mined blocks without broadcasting to the network and create a private branch which gets broadcast only after certain requirements are met. In this case, honest miners waste a lot of time and resources while the private chain is mined by selfish miners.

VII. CONCLUSION

The blockchain, incorporates in large amount things related cloud computing and additionally referred to as distributed ledger technology, is actually a digital information managed by a redistributed system, consisting of variety of various computers, in part of one centralized server. These completely different computers area unit noted as nodes and every one of them area unit connected in a very randomised approach. Nevertheless, varied forms of analysis round the blockchain area unit safeguarded by Bitcoin. In any case, it is critical to reminder that blockchain and Bitcoin isn't a similar object. In this article, we've got search the ideas of the blockchain technology that consists of basic definitions, characteristics, key ideas, advantages, limitations, agreement algorithms and in conjunction with security challenges and the future work. We shall take associate thorough exploration of better accept the long run which includes each the centralized and decentralized models. Like any new innovation, the blockchain is a notion that originally interrupts, and over time it could endorse the improvement of a superior community that incorporates both the recent method as well as the innovative invention.

REFERENCES

1. Trimborn, S.; Peng, H.; Chen, Y. Influencer Detection Meets Network Auto Regression–Influential Regions in the Bitcoin Blockchain. Available at SSRN. 2022. Available online: https://www.ml-quant.com/42546603-d6fd-493a-8736-5130928e127b.
2. A. Biryukov, D. Khovratovich, and I. Pustogarov, "Deanonymisation of clients in bitcoin p2p network," in Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, New York, NY, USA, 2014, pp. 15–29.
3. F. Tschorsch and B. Scheuermann, "Bitcoin and beyond: A technical survey on decentralized digital currencies," IEEE Communications Surveys Tutorials, vol. 18, no. 3, pp. 2084–2123, 2016.
4. NRI, "Survey on blockchain technologies and related services," Tech. Rep., 2015. [Online]. Available: http://www.meti.go.jp/english/press/ 2016/pdf/0531 01f.pdf.
5. "Antshares digital assets for everyone," 2016. [Online]. Available: https://www.antshares.org
6. M. Vukolic, "The quest for scalable blockchain fabric: Proof-of-work ´ vs. bft replication," in International Workshop on Open Problems in Network Security, Zurich, Switzerland, 2015, pp. 112–125.
7. Rose, C. The evolution of digital currencies: Bitcoin, a cryptocurrency causing a monetary revolution. Int. Bus. Econ. Res. J. 2015, 14, 617–622. [CrossRef]
8. Vinay Shrivastava, A Survey on Blockchain Technology & Future Scope. UIJRRT Volume- 02, Issue 01, 2020 | PP.4-9.
9. A.Gervais, G. O. Karame, V. Capkun, and S. Capkun, ―Is bitcoin a decentralized currency?,‖ IEEE Security Privacy, vol. 12, pp. 54–60, May 2014.
10. J.Singh, Cyber-attacks in cloud computing: A case study,‖ International Journal of Electronics and Information Engineering, vol. 1, no. 2, pp. 78–87, 2014.
11. A.S.Elmaghraby, M. M. Losavio, Cyber security challenges in smart cities: Safety, security and privacy, Journal of Advanced Research, 5 (2014), 491–497.
12. Wikipedia,―Bitcoin‖,https://en.wikipedia.org/wiki /Bitcoin.
13. [12] The Future of Blockchain Technology, https://yorksolutions.net/the-future-of-blockchaintechnology/.
14. I. Eyal, A. E. Gencer, E. G. Sirer, and R. Van Renesse, "Bitcoinng: A scalable blockchain protocol," in Proceedings of 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), Santa Clara, CA, USA, 2016, pp. 45–59.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-06

Impact of Space Debris and How to Control On Debris

Chaudhari Nayana Aravind,
Department Of Computer Science, Arts, Science And P. O. Nahata Commerce College, Bhusawal

ABSTRACT - *Humans have littered not only the earth but also space. Humans have sent many satellites into space in search of space, little by little these satellites have become so many that some of them are still orbiting the earth and we call them space garbage. Two thousand satellites are currently orbiting the earth and three thousand dead satellites and millions of pieces of such equipment are orbiting in space. This satellites debris created in space can pose a huge threat to you going forward. Managing these dead satellites is a challenge for researchers around the world. As a solution to this, we can create a remote control machine or robot using artificial intelligence that we can control from the earth. Detecting space debris and bringing it back to Earth will help you make new equipment by recycling it.*

Keywords- *Collision, Debris etc.*

I. INTRODUCTION

October 4, 1957 has a special significance in the history of human being. This year, the first satellite, Sputnik 1, was launched into space with a weight of 83. Today, more than 2,000 satellites are orbiting the Earth in orbit. This has made our lives a lot easier today, with features like high speed, mobile phones or GPS. Every 2-3 months, some country sends its satellites into space. But just as everything has its advantages, it also has its disadvantages. The world is watching this danger, but no one is doing anything special. The extent to which this waste is growing, if not addressed in time, no satellite will survive in the future .In space, there are 128 million.

Pieces of space junk larger than 1 mm lying around the earth. In 2009 and March 2021, China's satellite was destroyed by space debris. This means that the danger is that the new satellite we have sent could be destroyed by a space debris and that the country would have to bear the loss of millions of dollars. It has been proved that Apollo 15, 16 and 17 laps of the United States are still on the moon. In addition, the photos taken by the astronauts show that golf balls and other materials are still on the moon. [1]

However, we need to find a solution so that we do not have to face any dangers and difficulties in the future. We can collect space junk and bring it back to earth using today's technology to help reduce space debris.

Specification and Description:-

(Fig:1) Earth without space debris

(Fig:2) Earth with debris

We have all seen at least one photograph of the earth taken from space, how calm, beautiful and clean the view would have been. Everyone thinks this is a real photo of the earth as shown in (fig 1), but this is not the case .Otherwise, if you really look, the photo of the earth looks like the one shown in (fig 2).As shown in that figure, the debris that appears around the earth is not already there, but it is a man-made. We sent many satellites into space for our benefit and the useless satellites are now orbiting the earth as garbage.

When we launch any satellite or spacecraft, we use a launching vehicle. Normally, these launching vehicles were a rocket and their job was to propel the satellite or spacecraft out of the earth's atmosphere. As the spacecraft exits the Earth's gravity, these launches separate from the spacecraft and remain there as they run out of energy. Similarly, a satellite has an age when it stops working when it reaches the end of its life and due to the failure of its machine, it becomes out of the control of the scientists and hence they travel in space as well. Similarly there are smaller satellites called cube satellites which are relatively less expensive. They are working to deliver the goods to the space station. When they deliver and finish their work, they too become part of the space junk. This directional waste often collides with each other, causing the waste to disperse into many smaller pieces.

II. IMPACT OF SPACE DEBRIS:-

- Meteorological satellites, messaging satellites and geosynchronous satellites orbit to the earth. According to a report, there are 600 active satellites in lower earth orbit surrounded by more than 1500 satellites in debris. You know that you have no control over this waste so this debris can hit the active satellite and make it useless.Every year, at least one satellite fails to hit this space debris.
- When it comes to the full orbit of the Earth, there are more than 3,372 currently active satellites surrounded by more than 5 million space debris. In 2014, the International Space Station had to change its location 3 times due to this debris.[2]
- In 1996 the first documented satellite involving space debris happened, when a French satellite was struck and damage by debris from a French rocket that had exploded a decade before.
- In 2009 U.S. Commercial satellite Iridium 33 collided with the debris which is an inactive Russian communication satellite.[2]
- Cosmos 2251 which is at speed of 35,890km/h which is collide , the collision created a 2,300 pieces of space shrapnel which now threat to other spacecraft in low earth orbit .

This was the first case of two satellites collide catastrophically in space. Now with tens of thousands satellite travelling 27,360 km/h the chances of these satellite getting close to each other are lot a higher. Collision between high speed object in orbit can create thousands of pieces of debris .When more and more object collide and create new space junk in the process to the point where earth orbit becomes unusable.[3]

Fig (3) Satellite Collision in Space

The number of satellite increase drastically over the past few years .In 2020 alone 1,283 satellite were launched into lower earth orbit. This is the highest number of satellite launches in a year to date, now the low earth orbit is about to become even more crowded. Over the next few decades Spacex is hoping to send a constellation of 42,000 satellites even "Amazon's Project Kuiper" , Oneweb Corporation , China's Hongyan and Canada's Telesat they are also planning of placing thousands of satellites of their own in the earth orbit.

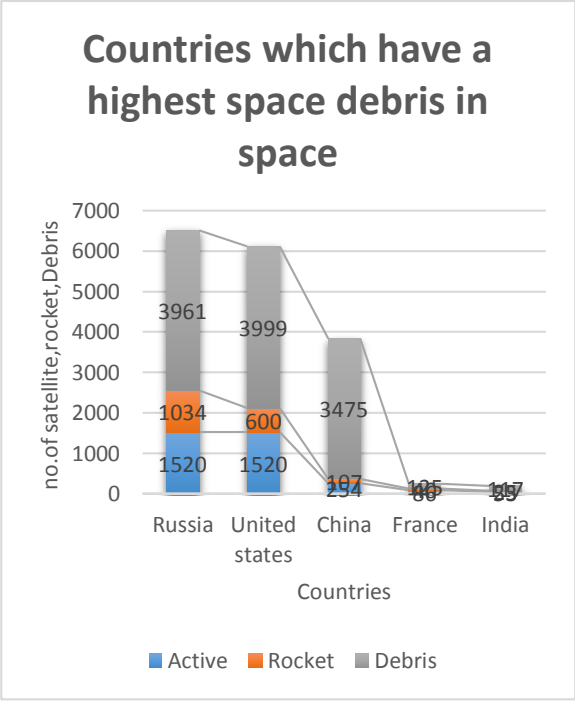
Countries Which Are Responsible For Highest Space Debris:-

Russia, the United States and China are the three largest producers of space debris. As shown in the graph, the satellite launches of these three countries are largely visible to us as well as their space debris.Due to the Indian rocket launch and the ASAT in 2019, it has been criticized that there are about 220 parts around the earth which are accumulated due to India. More than 400 pieces of piles came out, but only one orbit remained. And the rest of the Earth's atmosphere has re-entered and burned up. In this way we can see that Russia, U.S.A. and China have a huge contribution to make in space debris. [4]

Graph 1: Countries which responsible for highest space debris:-

Countries	Active Satellite	Rocket	Space Debris
Russia	1520	1034	3961
U.S.A	1520	600	3999
China	254	107	3475
France	86	40	125
India	53	25	117

(Table: 1)



Graph: Countries per no of satellite in space

III. METHODOLOGY:

Garbage growing in the Earth's atmosphere is a huge threat and we need to find a solution as soon as possible so that you do not have to deal with it again. Using today's technology we can find solutions to this problem, such as we can call it a remote control machine or we can call it a remote control robot in which we can fit a program that detects junk in space, collects it and returns it. We can bring it to earth and we can control it from the ground. When we bring space waste to earth with the help of a machine, it will have two advantages, one is that it will help to reduce the waste in space and the other is that we can recycle the returned satellites and create new objects from them. So how does this machine work, when we build a remote based machine we can put different sections in it like

1) A large section in which we can collect all the inactive satellites.

2) We can fit the robotic arm to the machine so that we can grab the space junk detected in front and put it in its collective section.

3) The third section is that the machine will have a robotic arm but that arm will have a net when you cannot catch a small piece of garbage. We can catch him with the help of this net.

4) In the fourth section we can fit the magnet to the robotic arm. The space junk that we can't catch with a robotic hand, a robotic hand with a net, we can catch by attracting an object with a magnet.

5) In the fifth section we can keep it as a fire block. Which we can use for objects that we cannot capture. We can destroy them by direct fire.

This way we can fit that robotic machine with 4 arms + inactive satellite collector.

IV. CONCLUSION

In the last few years, science and technology have flourished, man has made many discoveries not only on earth but also in space. Different satellites have been sent into space and this has made our life very convenient today but we are also facing many dangers due to the progress made by human beings. We call this space debris. If these are not disposed of as soon as possible, we will not be able to send satellites into space in the next few years. We can clean this space debris with the help of a remote control machine and bring it back to Earth and recycle it. So that you do not have to deal with bad situations.

REFERENCES

1. How rising space debris impact on isro budget. By vigneshradhakrishnan on 4th april 2022 https://www.thehindu.com/data/data-how-rising-space-debris-will-impact-isros-budget/article65289373.ece
2. ISS moves to avoid space debris on September 23,2020

https://phys.org/news/2020-09-iss-space-debris.html
3. Kessler Syndrome and the space debris problem. By Mike Wall published November 16, 2021
4. https://www.space.com/kessler-syndrome-space-debris
5. These are the countries on Earth with the most junk in space By Dave Mosher,AndyKierszon Oct 21, 2017.
6. https://www.businessinsider.in/tech/thspace/articleshow/61158510.cmsese-are-the-countries-on-earth-with-the-most-junk-in-space/articleshow/61158510.cms

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-07

Effect of Global Warming and Solution

Miss.Vishakha Murlidhar Kolte

Department Of Computer Science, Arts,Science And P.O.Nahata Commerce College, Bhusawal

Miss.Nayana Aravind Chaudhari

Department Of Computer Science, Arts,Science And P.O.Nahata Commerce College, Bhusawal

***ABSTRACT**:*

Global warming means an increase in global temperature . Although global warming does not seem to be a major problem, its far-reaching effects can be seen in the human race. The ice at the poles is melting faster as global temperatures rise. We have seen global warming rise in a few years .We are also seeing huge changes in the seasons due to global warming and adverse effects on humankind such as diseases due to climate change, crop failure due to unseasonal rains, The only solution is to create awareness among the people, and all countries must work together to ban factories and industries that cause pollution. At the individual level we should also try to stop deforestation and we should try to plant as many trees as possible, reduce the use of personal vehicle, and also use more solar or wind energy, conserve forests, plant one tree each. It is not impossible to reduce global warming if it is nurtured. As a solution to this we can make an app to plant as many trees as possible so that we can plant as many trees as possible.

I. INTRODUCTION:

After the invention of the steam engine in the eighteenth century, real industrialization began. Large scale use of coal in industrialization has increased the levels of carbon dioxide and methane in the atmosphere, so the sun's rays reflected from the earth remain in the earth's atmosphere, increasing the earth's temperature. The growing population and the increasing use of modern technology have made the situation worse, with the result that global warming is what we call global warming.

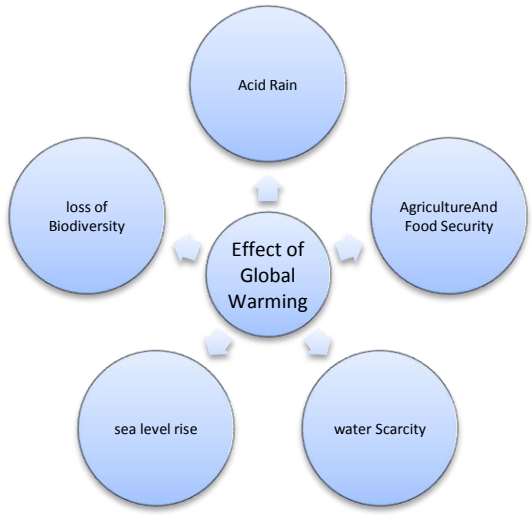
The result of climate change is that it never rains. Countries like India depend on regular rainfall for drinking and agricultural water but due to rising temperatures this rainy season is delayed. Rising global temperatures and rising carbon dioxide in the air are increasing the acidity of ocean water, which is affecting aquatic organisms, leading to the gradual destruction of ecosystems. Of course, limited use of plastic, recycling, and natural resource use can certainly make our generation the last generation to see the effects of environmental change and do something about it.

II. SPECIFICATION AND DESCRIPTION:

Factors such as increasing industrialization, increasing number of personal vehicles and their dust and resulting pollution as well as increasing deforestation,excessive use of plastic are responsible for global warming.Climate change due to extreme heat disturbs the balance of nature and has adverse effects on humankind such as extreme heat can lead to heat-related illnesses, extreme heat can cause forest fires.Global warming causes the temperature to rise as harmful gases such as carbon dioxide, methane, nitrous

oxide, hydrofluro carbon, sulfur, hexafluroide are absorbed by the sun's radiation to the earth, causing global warming. 2016 is the year with the highest temperature recorded, and the previous two decades, 1980 and 1990, are also recorded as the highest temperatures. As global temperatures rise, so does the amount of carbon dioxide in the atmosphere, as science and technology continue to evolve in every country, leading to the development of industry and the release of large amounts of carbon dioxide.In the last 6.5 lakh years, the amount of gas was only around 280 per 10 lakh units, now it has increased to 379 by 2005 and is still increasing. As it continues to rise, so does the prediction of an increase in atmospheric temperature.In the last 6.5 lakh years, the amount of gas was only around 280 per 10 lakh units, now it has increased to 379 by 2005 and is still increasing. As it continues to rise, so does the prediction of an increase in atmospheric temperature.bbbIn the last 6.5 lakh years, the amount of gas was only around 280 per 10 lakh units, now it has increased to 379 by 2005 and is still increasing. As it continues to rise, so does the prediction of an increase in atmospheric temperature.[4]

Impact of Climate change due to global warming :



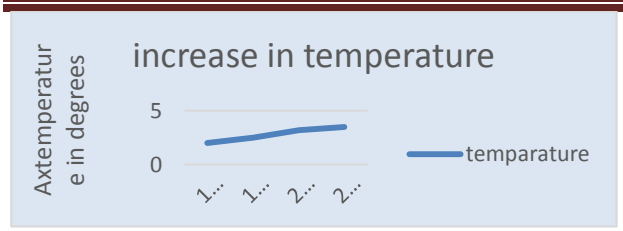
Acid Rain : When both the constituents of nitrogen oxides and sulfur dioxide are mixed with water, acid rain is obtained. Acid rain is very dangerous for us and also for agriculture.

Agriculture and food security: large part of india is rain fed depend on rainfall .If the rainfall is low or high, then the current situation may be very low or sometimes very high, but moderate rainfall is required for agricultural crops, but not due to climate change, which can lead to loss of crops and increase in food prices. If the rains do not come properly, the crop will not grow well and this will reduce the productivity of the soil and hence the demand for chemical fertilizers may increase. Excessive consumption can lead to greenhouse gases and affect the ozone layer. The ozone layer is already cracked and it will get bigger and we will have to face various consequences[1].

Water Scarcity : The source of Himalayan rivers is fresh water which never dries up. That is, when other rivers dry up in the summer, glacial glaciers in the Himalayas melt and form the same water that flows in the form of rivers. This means that the Himalayan rivers that we have flow for twelve months are called "perennial rivers". But due to global warming, the ice that used to melt in summer will now start melting in winter, which means that the glacial melting will be much higher and the glacial will disappear.this means that if the glacial disappears, the rest of the rivers will start drying up in the summer. Most of the people depend on this Himalayan water. When these rivers disappear, there will be barrenness. The Kaveri River, the Perennial River, never dried up, but it is now slowly drying up.It is also very dangerous for human beings.[5]

Sea Level Rise :India's coastline is very large and any change in sea level will affect India. Floods, cyclones can come and it can destroy infrastructure. Mumbai is on such a big coast and there is a lot of infrastructure there, if it is destroyed then tourist infrastructure there could be destroyed and all this will affect our economy.[3]

Loss Of Biodiversity :From the above points, we now know the consequences of global warming on humankind. Such as flooding, acid rain, dehydration etc can destroy your biodiversity. Birds, animals, various medicinal plants can be destroyed and the harmony of nature can be disturbed and we may have to face very bad consequences.[2]



Graph : 1 Temperature increasing per two decades in degrees

We are the first generation to see the effects of global warming and we are last generation have to find a solution as soon as possible and implement it.The main reason for the rise in global warming is deforestation and we need to increase the number of trees to increase the amount of oxygen in the atmosphere which help reduce the temperature by absorbing carbon dioxide.We need to focus on what the government is planning for this and also on an individual level.

As a digital solution, we can create an app that can detect the number of plants in a particular area and the amount of oxygen and other harmful gases like carbon dioxide in that area. If the oxygen level is low, try to plant as many plants as possible so that the carbon dioxide is absorbed and the climate is conducive.

The second solution is that we can distribute prizes to the specific area where the number of trees is more or the balance, so people from other areas will also try to plant trees."

we can store a carbon dioxide gas by using carbon sink , there are two types of carbon sink one is natural sink and second is artificial sink.

III. METHODOLOGY:

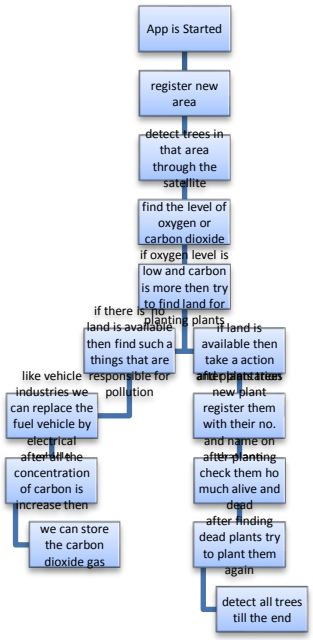


FIGURE:

In a nutshell, this app can work like a google Map, the only difference being that it will only detect trees in a specific area so that we can plant trees where we need to.

When you start the app, you have to register in that specific area. After registration, the number of trees in that area will be seen through satellite. The second step is that along with the plants, we can also check the amount of carbon dioxide and oxygen in that place. If there is less oxygen and more carbon dioxide then there is a need to plant more trees then we can check if there is land below for planting. If so, we should plant trees that produce maximum oxygen.Then you have to record how many and which trees you have planted on that app. Planting trees does not mean that all the trees will survive. Considering how many trees have died, we should replant them.And till the end we can keep track of those trees through that app.

On the other hand, if there is no place to plant trees, then we need to look at the carbon dioxide content of the area and see what factors are responsible for producing other gases like carbon dioxide, such as fuel vehicles that produce carbon dioxide. Instead, we can use electrical vehicles, factories that pollute the air. In this case, we can store carbon dioxide gas.We can store carbon dioxide using a carbon sink. We can store

carbon dioxide in two ways, one naturally and the other artificially.

Natural sink : Oceans , Soil , Forests etc.

Artificial sink : mines , depleted oil , reserve etc

IV. CONCLUSION

Today we are experiencing the changing climate of increasing heat, erratic rainfall, sometimes very low and sometimes very cold, and the reason for this is human beings. The pollution we have caused is responsible for increasing global warming,However, we need to find a solution on this problem as soon as possible and make sure that pollution is minimized.As a solution to this we can use an app to plant as many trees as possible so that we can help reduce global warming by reducing the amount of carbon dioxide in our atmosphere.

REFERENCES

1. UmairShahzadAugust 2015 "Global Warming: Causes, Effects and Solutions"https://www.researchgate.net/publication/316691239_Global_Warming_Causes_Effects_and_Solutions.
2. By AlinaBradford , Stephanie Pappas published ,February 03, 2022 "What are the effects of global warming." https://www.livescience.com/37057-global-warming-effects.html.
3. Connie Roser-Renouf, Justin Rolfe-Redding Engaging Diverse Audiences with Climate Change: Message Strategies for Global Warming's Six Americas
4. https://www.climatechangecommunication.org/journal-articles/ Ding, D., Maibach, E., Zhao, X., Roser-Renouf, C. &Leiserowitz, A. (2011). Support for climate policy and societal action are linked to perceptions about scientific agreement. Nature Climate Change, 1, 462-466. Doi: 10.1038/NCLIMATE1295
5. What 2021 looked like for global climate change on Thursday, April 21, 2022.https://indianexpress.com/article/research/what-2021-looked-like-for-global-climate-chang-7698219/

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-08

A Review: Comprehensive study of Electroencephalography (EEG) signal processing of human brain

Rahul Dinkar Chaudhari
Department of Computer Science
S.S.V.P.S.L.K.Dr.P.R.Ghogrey Science College, Dhule
merahul_chaudhari@rediffmail.com

Dr. Rakesh Shamrao Deore
Department of Computer Science
S.S.V.P.S.L.K.Dr.P.R.Ghogrey Science College, Dhule
rakeshsdeore@gmail.com

ABSTRACT:

A brain-computer interface (BCI) system has usually been developed using electroencephalography (EEG). Human brain is a highly complex system. Many of the properties that characterize complex and dynamical systems are relevant in the context of the brain. Electroencephalography (EEG) is a non-invasive neuro imaging method that measures the electrical activity of the brain. EEG signals are complex, extremely varied, and contain rich information about brain function. EEG signal processing involves a number of filtering, artifacts removal, feature extraction, and classification algorithms, is a crucial stage in the interpretation of EEG data. This paper provides an overview of the techniques used in EEG signal processing. We also discuss the applications of EEG signal processing, uses of Electroencephalography and the study of brain-computer interfaces. Overall, EEG signal processing is a fast developing field that has enormous potential to further our knowledge of brain function and enhance human outcomes.

Keywords: *EEG signal processing, Brain-Computer Interface(BCI), EEG Electrodes, EEG Acquisition.*

I. INTRODUCTION

Brain–computer interface

A brain–computer interface (BCI) is a computer-based communication system that analyses signals produced by the central nervous system's neural activity. It is a very effective communication technology that does not rely on neuromuscular or muscle pathways to accomplish communication, command, and hence action. While thinking with intention, the subject generates brain signals that are converted to commands for an output device. As a result, a new output channel is available to the brain[1]. The basic goal of a BCI is to detect and assess the features of signals in the user's brain that indicate the user's intention. These features are then transmitted to an external device that executes to fulfill the user's desired intention [1,2].The four successive elements of a BCI-based system are signal acquisition, preprocessing, translation, and feedback or device output.

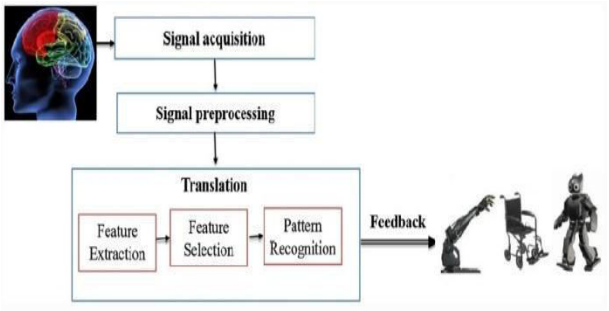


Fig:- BCI Components[1]

Signal acquisition, the first BCI component, is primarily responsible for receiving and recording the signals produced by neural activity, as well as sending these data to the preprocessing component for signal enhancement and noise reduction. Brain signal acquisition methods can be categorized as invasive and non-invasive. In invasive methods, electrodes are neuro-surgically placed either inside or on the surface of the user's brain. Brain activity is recorded using external sensors in non-invasive technology [1,3].

After preprocessing, the important signal's different characters (such as the signal's characteristic connected to the user's intention) are extracted from irrelevant data and presented in a way that allows them to be translated into output instructions. This component creates selective features for the improved signal, reduces the size of the data that can be sent to the translation algorithm, and then converts characters into the relevant instructions that the external device needs to complete the task (for example, instructions that complete the user's intent). The output device is guided and controlled by the instructions acquired by the translation algorithm. It assists users in achieving their goals, such as selecting alphabets, controlling a mouse, operating a wheelchair, moving a robotic arm, and moving a paralyzed limb with a neuroprosthesis. Computers are currently the most often utilized output device for communication [1,4].

II. Materials and Methods

1. Human Brain Anatomy

The cerebellum, cerebrum, and brainstem are the three major components of the human brain. The cerebral cortex, brain nucleus, and limbic system make up the cerebrum. Cognitive and higher-level emotional functions are principally controlled by the cerebral cortex. It is found on the human brain's outermost layer, with a thickness of around 1-4 mm, and is primarily made up of grey matter, with white matter below[5]. The cerebrum is responsible to perform the brain functions like touch interpretation, vision analysis and hearing. It also acts for controlling the speech, learning, reasoning, emotions, as well as the movement control. The cerebellum which is located below the cerebrum controls the muscle movements, body posture, and balance. The third part, named the brainstem acts as connecting part between the cerebrum and cerebellum to the spinal cord and controls heart rate, breathing, body-temperature. It also works to control the digestion process, swallowing, vomiting, sneezing, coughing, and wake and sleep cycles[6].

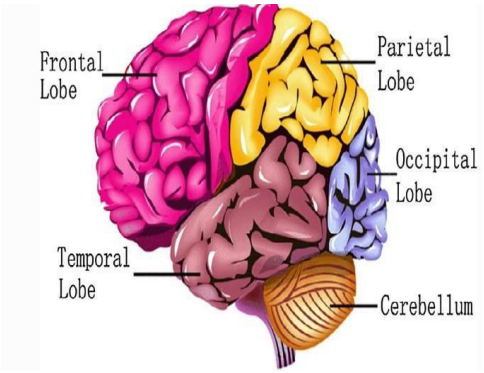


Fig:- Physiological Structure of the cerebral cortex[1]

2. History of EEG

The brain works by transferring electrical signals between neurons. One method to study the brain's electrical activity is to record the potential of the scalp caused by brain activity. The signal that is recorded, i.e., the potential variations between two placements, is called an electroencephalogram (EEG). EEG is one of the most efficient methods to monitor brain activity, often known as brain wave. Hans Berger recorded the first human EEG in 1929 and published the first human EEG paper.

As a major in the field, it was he who devised the term "ectroencephalogram". Richard Caton's early research on animal brain activity in the nineteenth century were the foundation for his work. Electrophysiologists and neurophysiologists gradually verified his results, allowing EEG research in clinical medicine and brain science to advance quickly. The changes in emotion can be understood by studying the EEG signals. The central nervous system's (CNS) functional and physiological changes can be reflected in neuronal potentials. The EEG does not just represent the electrical activity of a single neuron, but rather the electrical activity of a group of neurons in the brain area where the EEG measuring electrode is positioned[6].

As a result, the EEG signal includes a wealth of useful and meaningful psychophysiological information. In medicine, EEG signal classification, processing, and analysis can give an objective basis for detecting some diseases. In neuro-engineering, disabled people can use EEG signals produced by motion imagery or mind to control wheelchairs or robotic limbs. This is a popular topic right now that is known as Brain-Computer Interface (BCI). Analysis and processing of EEG signals is always problematic in brain research because of the non-stationarity of EEG data and the numerous environmental influences[1].

3. Basics of EEG signals

The EEG is a test that measures electrical activity inside the brain. In EEG procedure an electronic instrumentation called EEG-instrumentation is used to acquire the brain electrical signal. The EEG-instrumentation is interfaced with the brain through small conducting discs called EEGelectrodes which are attached to the scalp (Fig. 3) which is the outer part of the human head developed with the layers of skin and subcutaneous tissue. The brain cells which are active all the time, even during our asleep, communicate via electrical impulses which collectively appear as the electrical signal called electroencephalogram. Electroencephalograms are found as the time varying signal which can be analyzed to extract many important information related to brain anatomy, physiology and the brain disorder and brain diseases [6].

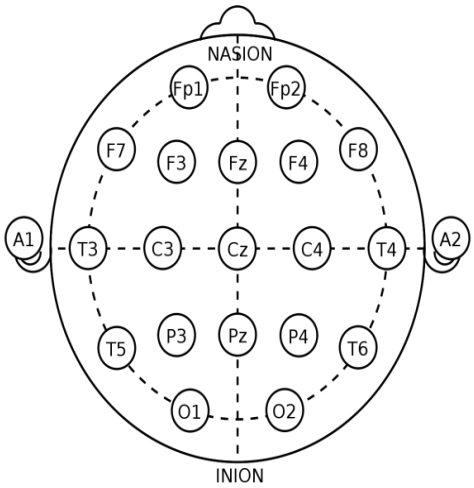


Fig 3 :Positions/location of the EEG-electrodes for the international 10-20 system for electroencephalography-recoding [6]

According to the difference of the EEG signal frequency bands, EEG signals can be divided into five types: the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (>30 Hz), as shown in figure 4. In the different studies, the specific range of each frequency band is slightly different [7].

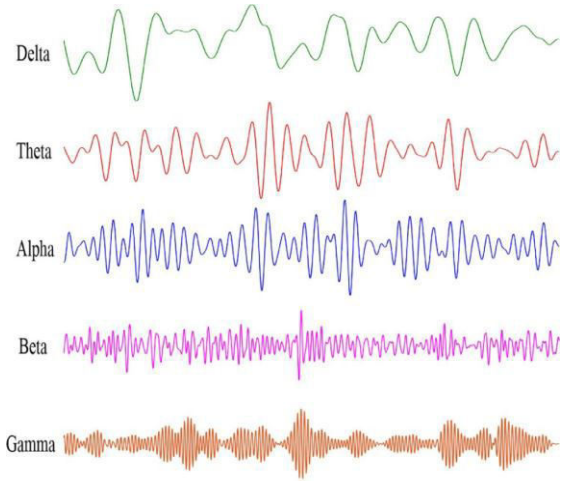


Fig 4 : EEG Rhythm waveforms [7]

The EEG signals of the different frequency bands are related to conscious human activities [9]. Delta waves often occur in the unconscious state of deep, dreamless sleep. Theta waves appear in sleep, dreaming, and sleepiness and are associated with sub consciousness. When the positive emotions are evoked, the Theta waves on the frontal midline will increase[7].

Alpha waves arise when a person is relaxed but conscious. The asymmetry of the alpha waves in the frontal lobe reflects the valence of the emotions, and the midline sagittal channel plays an essential role in the study of the EEG signals [10]. In neutral and negative emotions, alpha waves have higher oscillatory energy than beta and gamma waves[7].

Beta waves occur when the human mind is active and highly concentrated. The significant beta waves in the frontal lobe can reflect emotional valence. The average power ratio of the beta and alpha waves can reflect the active state of the brain. Gamma waves are associated with hyperactivity in the brain [11,12]. Studies have shown that simultaneous use of the alpha, beta, and gamma waves for emotion recognition is more reliable [7][10].

4. EEG Electrodes and Electrode Placement

In EEG procedure, a number of EEG electrodes are attached to the head-scalp with a standard electrode placement protocol such as international standard 10–20 system [40-42]. The 10–20 system which is an internationally recognized EEG-acquisition process to place the EEG signal recording sensors to determine the locations of the electrodes on the scalp surface. The 10–20 protocol based electrode placement system for EEG recording has been proposed and developed to maintain a standard testing procedure so that the

clinical test results obtained from the EEG procedures could not only be reproduced but also could be effectively analysed and compared with some standard data sets to extract conclusive information and research finding[6].

The 10–20 electrode-array system is basically based on the relationship between the location/position of the EEG-electrodes attached to the headscalp and the underlying area of cerebral-cortex. In the 10-20 system, the electrodes are placed on the scalp surface in such a way that the actual distances between two adjacent electrodes are either 10% or 20% of the total front–back or right–left distance of the skull and hence "10" and "20" in the "10–20 system" actually refer to inter electrode distance as a fraction of the total distances. Therefore, during an EEG experiment, the technician would place the EEG electrodes after measuring the dimension of the head and making the markings on the scalp surface to fix the location of the electrodes to be placed. As the patient's head generally contain lot of hairs the electrodes are placed individually with the help of a certain type of adhesive through an array of surface electrodes are placed on the scalp as a head net arrangement housed on an elastic cap[6].

5. EEG Acquisition

It is first part of BCI system, which sense and measures the signals of brain. This component is mainly accountable for receiving & registering the signals generated by neuronal activity and transferring these signals to the next component of BCI (preprocessing part) for signals improvement and electrical noise attenuation[8] . There are three common methods used for signal acquisition:

Invasive: In this type of BCI, very small electrodes are directly implanted into the brain i.e. cortex for assess the neuronal activities exa. Invasive intracranial electrodes [8].

Semi-Invasive: In this type of BCI, electrodes are implanted either in epidural or in arachnoids space exa. Electrocorticogram (ECoG) [8].

Non-Invasive: In this type of BCI, electrodes are placed on the scalp. In non-invasive system, EEG (electroencephalography) uses most commonly, although MEG (magneto encephalography), PET (positron emission tomography), fMRI (functional magnetic resonance imaging) and fNIRS (functional near-infrared spectroscopy) have been started to use recently [8].

III. METHODS

Common methods for preprocessing EEG data that have been employed in various studies include independent component analysis (ICA) , principal component analysis (PCA) , common average reference (CAR) and common spatial patterns (CSP) . When employing multi-channel recordings, PCA and ICA tools use blind source analysis to remove noise from the source signals, allowing them to be utilized to remove artifacts and reduce noise. The CSP method identifies spatial filters that can be used to identify signals that correlate with muscular motions. For noise reduction, the CAR is ideal [1].

IV. CONCLUSION

The invention of EEG signal processing and BCI technology has transformed the study of neuroscience by enabling researchers to monitor and manipulate the electrical activity of the brain. In this review paper we attempted to summarize information on BCI and EEG signal processing with human brain, EEG electrodes system, EEG signal frequency, EEG electrodes placement, EEG acquisition and EEG processing methods with will help full for researcher for their future study. EEG Signal processing and BCI have enormous potential for the future of neuroscientific research and human-computer interaction.

REFERENCES

1. Houssein, E.H., Hammad, A. & Ali, A.A. Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. *Neural Comput & Applic* 34, 12527–12557 (2022).
2. Mudgal SK, Sharma SK, Chaturvedi J, Sharma A (2020) Brain computer interface advancement in neurosciences: applications and issues. Interdiscip Neurosurg 20:100694
3. Abdulkader SN, Atia A, Mostafa M-SM (2015) Brain computer interfacing: applications and challenges. Egypt Inf J 16(2):213–230.
4. Jerry JS, Dean JK, Jonathan RW (2012) Brain–computer interfaces in medicine. In Mayo Clinic Proceedings, volume 87. Elsevier, pp 268–279.
5. Yin Z, Zhang J (2018) Task-generic mental fatigue recognition based on neurophysiological signals and dynamical deep extreme learning machine. Neurocomputing 283:266–281.
6. T. K. Bera, "A Review on The Medical Applications of Electroencephalography (EEG)," *2021 Seventh International conference on Bio Signals, Images, and Instrumentation*

(ICBSII), Chennai, India, 2021, pp. 1-6, doi: 10.1109/ICBSII51839.2021.9445153.

7. Jiang Wang, Mei Wang, Review of the emotional feature extraction and classification using EEG signals, Cognitive Robotics, Volume 1, 2021, Pages 29-40, ISSN 2667-2413, https://doi.org/10.1016/j.cogr.2021.04.001.
8. Shiv Kumar Mudgal, Suresh K Sharma, Jitender Chaturvedi, Anil Sharma,Brain computer interface advancement in neurosciences: Applications and issues,Interdisciplinary Neurosurgery,Volume 20,2020,100694,ISSN 2214-7519,
a. https://doi.org/10.1016/j.inat.2020.100694.
9. Z. Liu, Q. Xie, M. Wu, W. Cao, D. Li, S. Li Electroencephalogram emotion recognition based on empirical mode decomposition and optimal feature selection IEEE T. Cogn. Dev. Syst., 11 (4) (2019), pp. 517-526.
10. E.S. Pane, A.D. Wibawa, M.H. Pumomo Channel selection of EEG emotion recognition using stepwise discriminant analysis 2018 Int. Con. Comput. Eng. Net. Intel. Multimedia (2018), pp. 14-19.
11. G.H. Zhang, et al. A review of EEG features for emotion recognition Sci. Sin. Inform., 49 (9)(2019), pp. 1097-1118.
12. D.O. Bos EEG-based emotion recognition the influence of visual and auditory stimuli Emotion (2012).

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID:CS-09

Image PreProcessing for melanoma skin images combination of techniques – a survey

Mr. Sunil Dhondu Mone
Assistant Professor,
R. C. Patel ACS College, Shirpur, Dist. – Dhule
sunil.d.mone.research@gmail.com

Dr. B. H. Barhate
Professor, Vice Principal
TESS Bhusawal Arts, Science and P. O. Nahata Commerce College, Bhusawal, Dist – Jalgaon
bhbrama123@gmail.com

ABSTRACT -*In this paper we have consider selected papers for review purpose, all papers are working with melanoma skin cancer detection. A paper varies with respect to techniques used for each stapes of processing of image,segmentation, extraction of feature as well as classification purpose.*

For each stages of melanoma skin cancer detection various methods have been used. In this paper our study is limited to stages consider in preprocessingof skin images for melanoma skin cancer detection and altimetry our survey is limited to various techniques or methods used for preprocessing, may be various techniques used for each stage of preprocessing or may be simply technique used to processes skin image so that after preprocess image feed for segmentation shall be more clear and effective to find region of interest.
For preprocessing purpose various techniques have been used which include Separate gray scale contraction algorithms, Filtering morphological closing and contrast enhancement, Dull razor method, Gaussian filtering for image smoothing, Median filtering for noise filtering Etc.

Keywords - *lesion, mole, Melanoma, Non-melanoma, Skin Cancer.*

I. INTRODUCTION

Traditionally dermatologist see or observer skin mole by naked eye or by magnifying manually followed by after observing mole he judge whether the mole is melanoma skin cancer or not. For judgment purpose dermatologist considers some criteria like color change, Asymmetry, Border, diameter, Texture etc. He may consider all or few of these criteria for judgment and accordingly take judgment. Criteria consider for judgment vary from dermatologist to dermatologist, ultimately judgment for same skin mole also vary from dermatologist to dermatologist.

Melanoma is melanocyte-based (pigment cells). Melanoma might often occur on any skin surface since most melanocytes are in the skin. It frequently appears on the skin of the head, neck, or the area between the shoulders and the hips in the case of men. In the case of women, it is often found on the skin on the lower legs or between the shoulders and the hips. Though melanoma is rarely found in people with dark skin, and for dark skin person, The palms of the hands, soles of the feet, under the fingernails, and toenails are where it is most frequently discovered.. Melanoma may appear as a new spot or as an existing spot, or mole that changes color, size, or shape with time. It usually has an irregular shape size and color. (ALISON J. BRUCE, May 2020)

Over the past few decades, the number of cases has been rising quickly, of melanoma which is one of the most aggressive forms of skin cancer, It causes the bulk of skin cancer-related deaths.

Fortunately, early detection of melanoma skin cancer can help in the patient's recovery.. However, still it is challenging to classify melanoma in its early stages.

There are numerous methods for predicting and categorising melanoma skin cancer..

For detection of melanoma skin cancer using digital image processing need to go thru four steps

- Feature Extractions
- Classification

In preprocessing step image is smoothen by removing unnecessary part or by doing enhancement. For removing unnecessary part of image like hairs or any unwanted part various techniques or algorithms are used also for image enhancement purpose various techniques or algorithms are used which include Separate gray scale contraction algorithms, Filtering morphological closing and contrast enhancement, Dull razor method, Gaussian filtering for image smoothing, Median filtering for noise filtering Etc.

Purpose of segmentation is to obtain area of interest (AOI) or lesion from given whole image. In Segmentation step image is segmented using available segmentation algorithms or methods, for segmentation as per need one or combinations of more than one algorithm are used. Few of segmentation algorithms can be listed as masking, flood fill operation, single erosion, color based K-mean filtering, K-mean thresholding in histogram, segmentation method like Otsu, Modified Otsu, Water shade Etc.. Since dermic image obtain shall be irregular in shape and size and it is challenging to segment image so that one can obtain lesion of area of interest. (Bhumika Parmar, 2019)

Purpose of feature extraction is to know various feature of segmented image. Based on value of segmented image it shall be efficient to classify given image is as a melanoma skin cancer or not.

Among commonly used features we can list two sets as 1) Color, Shape, size, Texture and 2)Asymmetry, Border, Color, Diameter (ABCD)

In classification step various techniques are used for classification which include 1) calculate TDV (Total Dermescopy Value) by multiplying value of parameter. 2) Using ‘multiclass support vector machine’ (MSVM) 3) ‘Support Vector Machine’ (SVM) 4) Supervise Classification (Training, Testing and Classification) 5) Neural Network 6) Conventional Neural Network (CNN) (Farzam Kharaji Nezhadian, 2017)

1. Normal flow for melanoma skin cancer detection:-

Normally with the help of computer to detect and classify whether dermic image is melanoma or malign goes thru four steps (Steven Lawrence Fernandes*, 2016) as shown in figure 2.

- Pre-processing
- Segmentation

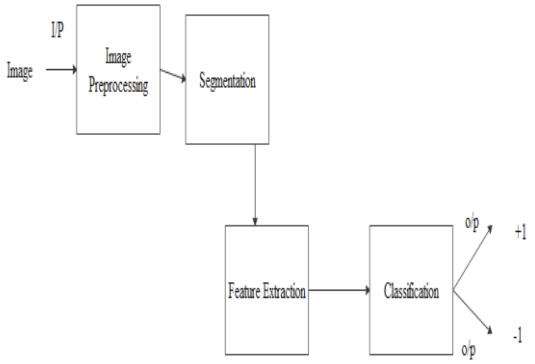


Figure 2: Computer-Aided melanoma skin cancer or lesion detection system

Image given as input will be any skin suspected image. Preprocessing will improves the quality of the image by reducing or removing unwanted parts.Various algorithms or mixing of algorithms used by the existing researcher for smooth the image.

In segmentation stage on can try to find an Area of Interest (AOI) or lesion. To segment images, algorithms or methods used by the existing researcher will use to improve the quality of segmented images, so that the next step may be efficient.

For feature, extraction selection provides the measurement vectors for various parameters. Parameters may ABCD, Texture, etc. for better judgment of classification.

For classification model will be developed may be by using the concept of machine learning or deep learning depends on the need for quantitative and qualitative analysis of skin lesion.

2. Pre-processing Technique sub techniques

Since it reduces noise and enhances the quality of the original image, image pre-processing is a crucial step in the detection process. It was necessary to use it to focus the look for background abnormalities that affected the result. The main objective of this stage is to improve the melanoma image's quality so that it may be processed further by removing irrelevant and unnecessary background elements. A wise choice of pre-processing methods can significantly increase the system's accuracy. The following figure illustrates the overall structure of approaches employed in the pre-processing step of medical image processing.

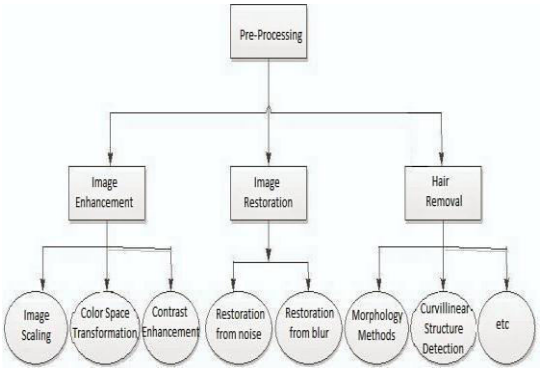


Fig. Framework techniques used in preprocessing stage of image processing

The three phases of image enhancement, image restoration, and hair removal can all be used to achieve the preprocessing stage's objective.

3 Image Enhancement

Picture augmentation is a vital technique for enhancing the image's aesthetic appeal. The three categories of picture enhancement are as follows:

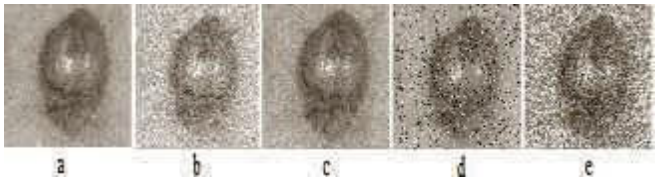
3.1 Scaling of Image

Due to the lack of uniform and standard image sizes, image scaling techniques are used. Because skin cancer image may come from a variety of sources and sizes, the first step is to scale the images so that the width pixels are set but the height pixels are changeable.

3.2 Image Restoration

The process of recovering a deteriorated image from a blurry and noisy one is known as image restoration. Several methods can be used to recover the damaged photos. Defects in the imaging system, poor focusing, motion, and other factors can all lead to picture degradation, which typically results in noisy or blurry images. To choose the best denoising technique, it is crucial to be aware of any noise present in a picture because corrupted images make defect identification more difficult.

Such noises have been simulated to create the sample.



a) Noice free Imageb) noice using Gussian c) noice using Poison d) noice using Salt and Peppee e) noice using Speckle

II. SURVEY OF MELANOMA SKIN CANCER DETECTION PREPROCESSING TECHNIQUES

[1] The work of Nadia Zghal presented his work by dividing image processing of the dermic image into 4 steps, in the first step of preprocessing separate grayscale, used filtering morphological median and closing and contrast enhancement. Author try to achieve accuracy up to 90% M. Krishna Monito (Nandia Zghal, 2020)

[2]Sanjay Jaiswar, MehranKadri, VaishaliGatty use technique of histogram equalization so that color values are distribution between large set of extreme color used in the palette, which help to detect. Additionally, authors noted that image resolution normalization, scale fitting, color range normalization, and image illumination equalization are all possible with the aid of pre-processing approaches. For example image normalization for resolution matching. They stated also it is possible to normalize other parameters like color palette normalization, color saturation normalization, and color components normalization also known as histogram equalization. (Sanjay Jaiswar, June 2015)

[3] Nilkamal S. Ramteke and Sweta V. Jain used various techniques for pre-processing. Various sources of bitmap or JPEG images are transfer in indexed images. The initial image is changed from RGB to grayscale and then to binary. It transforms an image into one that is suitable for a certain use. Image enhancement (edge highlighting, sharpening, de-blurring, brightening, change in contrast, masking, hair removal, cropping or resizing, and/or noise reduction) is the second stage of pre-processing. To detect the edge of a skin lesion, they used the Canny Edge Detection method. Author success to obtain detection accuracy up to 90%. (Nilkamal S. Ramteke, Aug 2013)

[4] M. Krishna Monito attempt to use Gaussian filter for image smoothing and median filtering for noise filtering and preserve the colorof lesion. For classification used multi-class support vector machine (MSVM) and achieved 96.25% accuracy. (M. Krishna Monito, 2020)

[5] Enakshijane who surveyed various methods used for each step she highlighted algorithms used for preprocessing step for the various purpose include hair removal, denoise, sharping, resize etc. Also highlight different classification which includes SVM, feed-forward artificial neural network, and deep convolution neural network etc. (Enakshi Jana, 2017)

[6] XiaojinYuan dividedcomplet task in two stages Texture feature extraction(Texture analysis and Texture feature Extraction) and classification. For texture analysis used algorithms ‘local auto regration’ and ‘Globle filter banks’ results in generation of set of texture feature vector. Classify - use SVM for remove error and better accuracy. Author claim up to 86% accuracy. (Yuan, Sep. 2006)

[7] Vijayalakshmi M M. created model with three phases. In phase 1 collect data set from ISIC dataset, for

preprocessing task done for hair remove, glare remove and shading remove. Author success to reach up to 85% accuracy (M., 2019) (Dr.R. Pon Periyasamy, May 2017)

[8] Agung w Setiawan demonstrated the performance of algorithms CLAHE for contrast enhancement and stated CNN is not much suitable or effective for contrast enhancement, CLAHE gives the same accuracy in training and validation, also stated image contrast enhancement is not required for skin cancer screening purpose and tried to achieve 92% accuracy. (Agung W, 2020)

[9] H. M. Ahmaed used following steps in their paper, Pre-processing for this type of image requires obtaining a color image and extracting the three bands RGB. Applying the same procedure to each band, first using a 3x3 median filter, and then using an morphological operation in connect to image closed with the help of three masks of 0 degrees, 45 degrees, and 90 degrees For classification purpose used SVM. Try to achieve accuracy up to 97% (H M Ahmed, 2018)

[10] Dr.R. PonPeriyasamy used bi-linear interpolation method to reduce image to 25% for reducing image size, followed by Frangivesselness algorithm used to remove hairs in preprocessing step. For segmentation k-mean clustering segmentation method is used. Followed by features are extracted for parameters Asymmetry, Border, Color, Diameter and Evolving (ABCDE), for classification 4 classifiers are consider for compare performance, classifier (maximum performance) as NN(80.00%) NMC(70%), LINEAR(100%) and SVM(83.33%) (Dr.R. Pon Periyasamy, May 2017)

Sr. No	Reference	Preprocessing Techniques / Task	Result
01	(NandiaZghal, 2020)	separate grayscale, used filtering morphological median and closing and contrast enhancement	90%
02	(Sanjay Jaiswar, June 2015)	image illumination and equalization, normaliztion of color range, fitting of image scale, normalization for image resolution, equalization for histogram.	
03	(Nilkamal S. Ramteke, Aug 2013)	Sharpening,Edge highlighting, de-blurring, contrast change, brightening, masking, follwoed by hair removal, resizing or cropping , and or reduction of noise) Edge Detection with a Twist	90%
04	(M. Krishna Monito, 2020)	Gaussian filter, median filtering for noise filtering and preserve the color of lesion	96.25%
05	(Enakshi Jana, 2017)	hair removal, denoise, sharping, resize etc	Survey paper
06	Yuan, X. (Sep. 2006).	Task is divided in two steps 1)classification and 2)Extraction of Texture feature(analysis for Texture and Extraction for feature of Texture).	86%
07	(M., 2019; Zeeshan Abbas, 2019)	task done for hair remove, glare remove and shading remove	85%
08	(Agung W, 2020)	CLAHE and CNN are used for contrast enhancement	CLAHE is more suitable than CNN 92%
09	(H M Ahmed, 2018)	Three band of RGB is extracts, median filter is applies having size 3x3, then image is consider for processing. Operation of type morphological with Closed having masks of 90 degrees, 45 degrees, and 0 degrees.	97%
10	(Dr.R. Pon Periyasamy, May 2017)	bi-linear interpolation method to reduce image, followed by Frangivesselness algorithm.	NN(80.00%) NMC(70%), LINEAR(100%) and SVM(83.33%)

III. CONCLUSION

1. Preprocessing is very initial stage for detection of melanoma skin cancer, further accuracy is fully depend on preprocessing.
2. Algorithms or methods as well as combination of methods used for preprocessing vary from paper to paper.
3. Most papers used more than one combinations of techniques for preprocessing melanoma skin.
4. Overall performance is depending on proper combination of methods used for preprocessing, step.

REFERENCES

1. Agung W, S. (2020). Effect of color enhancement on early detection of skin cancer using convolutional neural network. IEEE Xplore, 100-103.
2. Albahar, M. A. (2016). Skin Lesion classification using convolution neural network with novel rugularizeer. IEEE Access Vol 4, 1-8.
3. ALISON J. BRUCE, M. A. (May 2020). Overview of Skin Cancer Detection and Prevention for the Primary Care Physician. Mayo Foundation / or Medical Education and Research, 491-500.
4. Bhumika Parmar, B. T. (2019). Automated Melanoma Types and Stages. 978-1-5386-8190-9/19/@31.00/© 2019 IEEE, 2019 innovations in Power and Advanced Computing Technology i-PACT, 1-7.
5. Dr.R. Pon Periyasamy, V. (May 2017). Melanoma Detection through K-Means Segmentation and Feature Extraction. International Research Journal of Engineering and Technology (IRJET) Volume 4, Issue 5, 1301 - 1305.
6. Enakshi Jana, D. R. (2017). Research in Skin Cancer Cell Detetin using Image Processing. IEEE, 50-53.
7. Farzam Kharaji Nezhadian, S. R. (2017). Melanoma skin cancer detection using color and new texture features. IEEE - 2017Artificial Intelligence and Signal Processing (AISP), 1-5.
8. H M Ahmed, R. J.-a. (2018). Evaluation Methodology between Globalization and Localization Features Approaches for Skin Cancer LesionsClassification . IOP Conf. Series: Journal of Physics: Conf. Series 1003 (2018) 012029 doi :10.1088/1742-6596/1003/1/012029, 1-12.
9. Jianfeng He, Q. D. (2019). Prediction of Skin Cancer Based on Convolutional Neural Network. Springer Nature Switzerland AG 2019 https://doi.org/10.1007/978-3-030-00214-5_150, 1223-1229.
10. M. Krishna Monito, N. A. (2020). Skin Cancer detection and Classification using machine Learning. Material Today Processing, 40-45.
11. Nandia Zghal, N. D. (2020). Melanoma Ski Cancer Detection Based on Image Processing. Current Image Imagin , 50-58.
12. Qurrat Ul Ain, B. X.-S. (2018). Genetic Programming for Feature Selection and Feature Construction in Skin Cancer Image Classification. Springer Nature Switzerland AG 2018 https://doi.org/10.1007/978-3-319-97304-3_56, 732–745.
13. Sinha, S. A. (2016). A Low cost FPGA based SVM Classifier fro malanoma detection . IEEE.
14. Steven Lawrence Fernandes*, B. C. (2016). Early Skin Cancer Detection Using Computer Aided Diagnosis Techniques. Journal of Integrated Design and Process Science 20 (1) DOI 10.3233/jid-2016-0002, 33-43.
15. Yuan, X. (Sep. 2006). SVM based Texture Classification and Application to Early Melanoma Detection. Proceding of 28th IEEE EMBS International Conferance, 70-73.
16. Zeeshan Abbas, M.-u.-R. S. (2019). An Efficient Gray-Level Co-Occurrence Matrix (GLCM) based Approach Towards Classification of Skin Lesion. IEEE, 317 - 320.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-10

The Role of Image Quality Enhancement for Anthracnose and Stem End RotDisease Detection.

Priyanka V. Barhate[1]
Assistant professor, DNCVPS Shirish M. Chaudhari College, Jalgaon, India,
[1]priyankavbarhate@gmail.com

Dr. B.H. Barhate[2]
Associate professor, BASPONC College, Bhusawal, India
[2]bhbrama123@gmail.com.

Abstract:

Mango is the most significant fruit in agriculture around the world and India's most widely cultivated fruits. However, it is susceptible to diseases that can reduce its quality and quantity and suffers from the major losses in marketing. This proposed approach considers Anthracnose and Stem End Rot fruit diseases and uses two techniques to obtain clear, highquality images. The high quality images will facilitate easy disease identification. Digital Image processing is used to identify and assess problems. Image quality is one of the crucial factors that influence the subsequent steps in the disease detection process.Two statistical quality metrics MSE and PSNR are used to assess the quality of diseased images. Filtering and image contrast enhancement techniques are used to enhance the quality of images.

Keywords: Anthracnose, Stem End Rot, Filtering, MSE, PSNR, Contrast Enhancement.

I. Introduction:

Food is essentialpart of human's life and one of the best natural sources of food is fruit.Mango is the most leading fruit in our country which is highly demanded in local as well as international market. As per statistics given in the year 2014–2015, India exported around 42,998.31 million tons of mangoes[1]and in the year 2021-2022, 26,868.71 MT Mangoes.**[**Available from: https://agriexchange.apeda.gov.in/**]** However this production is greatly diminished because of the fruits affected by diseases, pests and make them non-edible or waste[1].

Hence there is a need to automatically detect Mango fruit diseases. To detect diseases, image quality is one of the crucial factors that influence the subsequent steps in the disease detection process. Images suffer from poor and bad contrast and noise. It is necessary to enhance the contrast and remove the noise to increase image quality[2].Noise is added into the image when it is being captured. Image quality is degraded by noise. Filtering techniques are used to reduce noise.In this experiment we used median and gaussian filter techniques.Two statistical quality indicators MSE and PSNR values are compared to evaluate image quality and determine the best filter[2].The primary purpose of image contrast enhancement is to improve image visual quality so that it can be seen by both humans and machines.Contrast is a visual contrast that distinguishes an object from the background and other objects[3].The image enhancement technique improves an image's contrast, brightness features, noise content reduction, and so on. This is useful to easily recognize the different features of image[4]. Anthracnose and Stem End Rot diseased images are used for experimental work. According to literature review these diseases are destructive, causing significant production and marketing

losses, as well as degrading the quality and quantity of fruit.

In Anthracnose **brown to black spots appear on fruits** later on coalesce to form large irregular botches or even cover the entire fruits. In Stem End Rot, **light brown spots near pedicel**, spot enlarge form brownish black patches. A dark brown to black rot develops at the stem end as a dark brown ring and progresses to the opposite end[5].

II.Problem Statement:

- Fruit disease identification using digital image processing is the challenging research area of agriculture.
- Images capture by cameras include noise which lowers the visual image quality.
- A low-quality image may produce inaccurate findings.
- Design and develop a proposed image quality enhancement framework.

III.Proposed Framework:

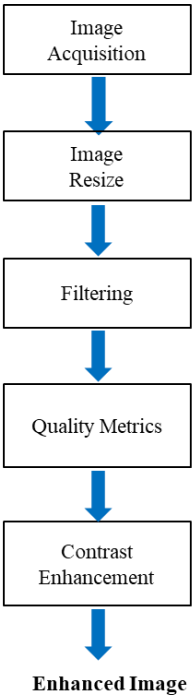


Figure 1: Image Quality Enhancement Framework

IV.Methodology:

Experimental work is performed in Python3.8 Programming Language. The IDE Spyder 4.2.5 under anaconda distribution is used for implementation.

Data Collection:

- 200 samples were collected for each disease from various fields and ripening rooms in Jalgaon.

Image Acquisition:

- Images are captured using **high resolution mobile camera** in natural sunlight.
- Images are stored in **jpg** format in database folder.

Image Resizing:

- The original images were captured using a mobile high pixel resolution camera.
- Images are resized to reduce processing time in subsequent steps.

Image Filtering:

- Image noise, often known as electronic noise, is a random variation of intensity or color information in images. It can be generated by a circuitary of scanner, digital camera's image sensor or particles of dust.
- This is the process for enhancing some information and cleaning up an image. The median and gaussian filters are used in this experiment.

Median Filter:

Image noise is removed using the median filter. The median filter is particularly important in image processing since it is well known for preserving edges during noise removal. It is *used to eliminate salt and pepper noise.*The median could be a midpoint pixel value chosen from the neighbourhood sorted distribution values. It does not generate a new, erroneous pixel value. At each pixel element position, the median operator arranges the values within the pixel element neighbourhood. This decreases edge blurring and image detail loss[6].

Gaussian Filter:

The Gaussian filter blurs images while removing noise and detail.The outcome of blurring an image with a Gaussian function is a Gaussian blur (also known as Gaussian smoothing).This blurring technique produces a smooth blur that resembles seeing a picture through a translucent screen. Gaussian smoothing is frequently employed as a pre-processing step in computer vision algorithms to improve image quality[7].

Filter Results:

Figure 2: (a)Anthracnose(b) Median filter (c) Gaussian filter

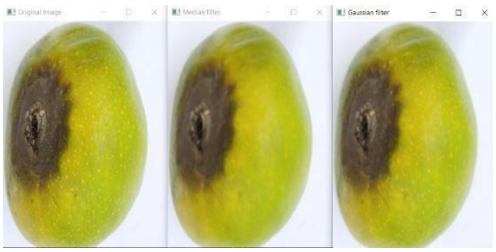


Figure 3: (d) Stem End Rot (e) Median Filter(f) Gaussian filter

PSNR and MSE are statistical measures used to assess image quality.

PSNR (Peak signal-to-noise ratio): The PSNR is the ratio of an image's maximum probable power to the power of corrupting noise that influences the quality of its representation.

$$\text{PSNR} = 10.\log_{10}\left(\frac{MAX_I{}^2}{MSE}\right)$$

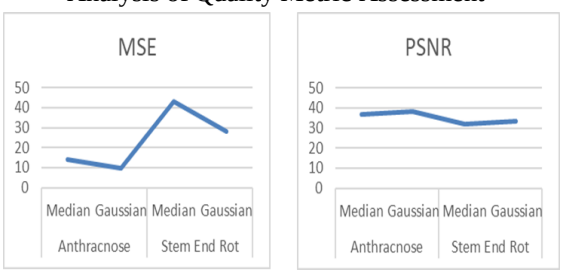
Here, MAX_I is the image's greatest possible pixel value.

MSE: It is the most commonly used image quality metric estimator. It is a comprehensive reference metric, and the values closer to zero are the better[8].The MSE is a measure of the cumulative squared error between the compressed and original images, whereas the PSNR is a measure of the peak error.

For a good quality image, the value of PSNR should be high and MSE should be low[2].

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[f(i,j) - g(i,j)]^2$$

Analysis of Quality Metric Assessment



Contrast Enhancement:

Contrast enhancement is essential to improve the perceptibility of objects in an image[4].

There are various approaches for improving images as histogram Equalization, Adaptivehistogram equalization and contrast limited adaptive histogram equalization[9].This research focused on contrast limited adaptive histogram equalization (CLAHE) because it works on color images[10]. CLAHE is a sophisticated method for equalising the histogram, in which the entire image is broken into tiles with sizes ranging from 8X8, 4X4 and so on. CLAHE focuses on small sections of an image known as tiles rather than the entire image. To erase the false boundaries, the neighbouring tiles are combined using bilinear interpolation[11].

Figure 4 (a)AnthracnoseGaussian image (b)Stem End Rot Gaussian image (c)Contrast Enhanced Anthracnose (d) Contrast Enhanced Stem End Rot

V. Conclusion

In comparison to the median filter, Gaussian filteryields better results because it generates lower MSE and higher PSNR valuesfor both diseases. Lower MSE and higher PSNR values create good quality images. CLAHE technique is used to enhance contrast of images which shows the clear disease portion. The proposed approach obtained clear, high-quality images for further processing. The aforementioned framework focuses the results of the enhancement process in disease detection, demonstrating that the visual representation of diseased images improved.In the future, improved fruit disease images will be taken and apply segmentation method to extract the affected region. These images could be used to detect anthracnose and stem end rot diseases.

References:

[1] A. Diana Andrushia and A. Trephena Patricia, *Artificial bee colony based feature selection for automatic skin disease identification of mango fruit*, vol. 150. Springer International Publishing, 2019.

[2] J. P. P. Janani, "Image Enhancement Yechnique: A Study," *Indian J. Sci. Technol.*, vol. 8, no. 12, pp. 83–89, 2015, doi: 10.17485/ijst/2015/v8i.

[3] V. S. Padmavathy and R. Priya, "Image contrast enhancement techniques-a survey," *Int. J. Eng. Technol.*, vol. 7, no. 2.33 Special Issue 33, pp. 466–469, 2018, doi: 10.14419/ijet.v7i1.1.10146.

[4] S. Kaur and Ritika, "Contrast Enhancement Techniques for Images – A Visual Analysis," *Int. J. Comput. Appl.*, vol. 64, no. 17, pp. 20–25, 2013.

[5] R. subba Reddy, *Mango:Inscect Pests Diseases, Nutritional Disorders*. New Delhi: Ooford & IBH Publishig Co. Pvt. Ltd.

[6] K. Padmavathi and K. Thangadurai, "The Role of Image Enhancement in Citrus Canker Disease Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 9, pp. 2014–2017, 2016, doi: 10.14569/ijacsa.2016.070941.

[7] V. Devishree, S. N. Bhavya, A. Anjalinair, and M. Thanuja, "Diagnosis of Liver Disease Using Gaussian Blur Algorithm," vol. 3, no. 2, pp. 995–1007, 2015.

[8] U. Sara, M. Akter, and M. S. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study," *J. Comput. Commun.*, vol. 07, no. 03, pp. 8–18, 2019, doi: 10.4236/jcc.2019.73002.

[9] S. Arora, M. Agarwal, V. Kumar, and D. Gupta, "Comparative study of image enhancement techniques using histogram equalization on degraded images," *Int. J. Eng. Technol.*, vol. 7, no. 2.8, p. 468, 2018, doi: 10.14419/ijet.v7i2.8.10487.

[10] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," *Proc. 2014 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2014*, pp. 2392–2397, 2014, doi: 10.1109/ICACCI.2014.6968381.

[11] S. Sood, H. Singh, and M. Malarvel, "Image quality enhancement for Wheat rust diseased images using Histogram equalization technique," *Proc. - 5th Int. Conf. Comput. Methodol. Commun. ICCMC 2021*, no. Iccmc, pp. 1035–1042, 2021, doi: 10.1109/ICCMC51019.2021.94180

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-11

Improving Wireless Sensor Network Reliability without Affecting QoS Attributes

Dipak E. Chavan[#1],
[#]1 Assistant Professor DOCS R. C. Patel ACS College Shirpur
**Address Including Country Name*
[1]dipak.chavan@rcpasc.ac.in

Dr. Gouri M. Patil[*2]
2 Assistant Professor DOCS P. O. Nahata College Bhusawal
[2] gourimpatil@gmail.com

Abstract Now a day there is trend of wireless medium where everyone is adopting the wireless communication for sensing data. The technique of clustering involves assembling comparable nodes into groups. The suggested effort entails creating a way to send messages with higher priority, such as information about hospitals, defence systems, forecast or environmental messages. A cluster is required to construct a trustworthy network in which all nodes/motes are distributed and the (CH)cluster head for each cluster is determined. The best way to increase network longevity and energy efficiency is by clustering. Now days the monitoring whole network is done through software defined network (SDN) in wireless communication. This SDN based proposed work can ensure to provide reliable network for essential communication. While increasing the reliability of network in cluster-based approach, we must focus on QoS attributes i.e latency, throughput, residual time etc.. The simulation results show that our strategy increases the number of living nodes and rate of packet delivery while decreasing end-to-end latency and usage of energy.

Keywords — ***WSN- Wireless sensor network, Clustering, CH- Cluster Head, QoS- Quality of service, SDN- Software defined Network, reliable network, Latency, Packet delivery ratio.***

INTRODUCTION

Wireless sensor networks (WSNs) are made up of nodes that are densely dispersed over an area and carry out a single activity (Chen Q, 2018, April). In WSNs, sensor nodes are given an irreversible finite energy resource, so that when it runs out, they die and the lifespan of the network decreases (Laufer R, 2016). Thus, one of the challenges facing WSNs is lowering energy usage (Goyal N K. K., 2018).

A scenario where a sender provides important information that is very sensitive and valuable and needs a high-reliable transmission medium that might be both secure and dependable may arise given the millions of data transmissions that occur across wireless networks today. (Santosh Ananad, 2021). In cases such as patient information at a hospital, essential information needed by our defense team, and natural disasters in remote areas of the country. Rather than efficiency, there is a need for trustworthy communication. The suggested SDN-based approach would provide a loyal and efficient network while also providing significant strength to impartant data transfer. Communication dependability is a critical element of a network that must meet numerous internal requirements.Researchers often created a network that is not dynamic and not energy efficient in order to provide network resilience. Here, the goal is to maintain network dependability without compromising service quality. The deployment of the sensor nodes must take into account a number of characteristics related to WSN QOS. Range of the node: It describes, in terms of distance value, the capability of the sensor to detect the data or respond to user command (Laufer R, 2016). There are two separate ranges: one is the transmission range, when data sensing findings are at their best, and the other is the interference range, where there

is no guarantee of data sensing or any response but it could occasionally make sense (Zhu, 2017).

Data Accuracy:-

It refers to how correctly a node is detecting by describing whether or not the records it makes of its surroundings are correct. Due to environmental changes and security threats including denial of service, masquerade, relay, and replay, node may occasionally detect unexpected data (Duan, 2017).

SDN base WSN can assist to improve dependability in order to overcome all of the mentioned scenarios. In essence, it functions as a node positioned across networks that has a global graphical representation that is centrally controlled by user-defined instructions and algorithms. This offers a single point complete control over a network. The problem can be solved more effectively if we group network nodes into clusters, locate cluster heads according to the node's remaining energy or range, and then use a software-defined network to manage the clusters.

Clustering is the most effective strategy for increasing network longevity and energy efficiency (Goyal N D. M., 2018). Clustering divides sensor nodes into clusters, each with its own cluster head (CH) and members. CH gathers data from members and sends it to the base station (BS). The information acquired should be correct, complete, and fulfil the requirements of the project. Clustering's main purpose is to accurately choose a group of CHs spread around the network. (Bhuyan B, 2016, March). CHs should be chosen such that they span the entire network, and participants should be capable of connecting directly to CHs through one-hop routing. Because data transmission requires a significant amount of energy, clustering enables most sensor nodes to send data over a shorter range while saving energy. CHs burn energy more quickly than cluster members. Hence, clustering should be done on a regular basis, with the most appropriate CHs chosen at each interval.

II. LITERATURE REVIEW

Wireless sensors are widely employed in industries, but they struggle to be reliable and secure at the physical layer, which is why the authors of this study have developed solutions that network designers might consider. Link failure, route loss, fading from many paths, and interference range are all factors that are considered for dependability. They research jamming attacks and eavesdropping for security purposes and take precautions as necessary. Eventually, they even established a relationship between dependability and security to efficiently utilise sensor resources and even how energy might be dispersed over an industrial wireless network *(Elappila, 2018)*. The authors of this paper presented an SDN-based network named SDSense and demonstrated its dynamic nature. This work was continued in two models, one of which uses the topology of the network to address performance-related scenarios, and the other of which addresses network congestion. They then concluded that their work maximised the utility of the network under two separate situations for congestion and topology based on their testing findings of energy consumption, delay, and throughput *(Haque, 2018)*. The authors of this paper suggested localising nodes depending on RSSI value. They utilised the ns2 simulator to put their ideas into practise, and it began with putting the reference nodes and obtaining their RSSI values, followed by the placement of four anchor nodes with respect to the target node, which involved measuring the distance and positioning them in an effective location. According to their simulation findings, their mathematical model achieved 95% accuracy, and their future work will focus on identifying hostile activities in networks. (Xu, 2018, October). The authors presented a cluster-based wireless network in which the cluster head is dynamically chosen based on the amount of energy it possesses at the time of the round. This cluster-based network was developed using different energy-efficient forms of the LEACH protocol, including heterogeneous, homogeneous, and centralised versions, and the work was assessed based on energy consumption and the number of active nodes after the simulation. This outcome shows how their labour extended their lives *(Anand, 2020, February)*. This paper proposes a study that use the AODV protocol for testing purposes in order to discover a defective connection by taking into account residual energy and end-to-end latency of each node. The network's fault tolerance is shown here. Before, they set four anchor modes at the network's corners and two in the middle at random to monitor defective nodes in the network. As their AODV protocol tests show, the node with the most latency and the least residual energy is the one that has to be taken care of *(Arun P Prabhan, 2019)*.

The authors of this study assert that they improve the performance of WSNs by developing their own proprietary load balancing algorithm that determines the busiest node and assigns the task to a less active node, resulting in optimistic performance from that path. As compared to the performance of round robin and active clustering, this method produced superior results since it required less response time and data center processing time (Nair, 2019). The authors of this paper suggested a paradigm for constructing a

reliable network throughput in a wireless sensor network. They discovered a reliable link between scalability, mobility, responsiveness, and power-efficiency as they followed the task. To create a dependable network, they compared the efficiency of cooperative ARQ and non-cooperative ARQ using the sequence and discovered that cooperative ARQ has less latency but uses more power; if power efficiency is important, non-cooperative ARQ is far superior to other methods (Bello, 2020).

III PROPOSED SYSTEM

A) Formation of Cluster: -

The nodes are dispersed throughout the network to detect environmental data, as depicted in Figure. The nodes located further away from the washbasin use more energy to transport data, resulting in a shorter lifespan.

In the first phase, nodes are grouped to minimise node energy usage. CHs are chosen based on three criteria: residual energy, centralization, and distance from the sink.

The location of nodes in a wireless sensor network is critical for establishing a viable connection between two endpoints. Because node clustering produces a predetermined number of clusters that connect networks with various cuts of nodes, we may select one parent node as the head of the cluster to collect data from sensor nodes within the boundary of the cluster by using RSSI value obtained from the node data. The received signal intensity is shown by the RSSI value, the higher the values, the greater the rate of success. Smooth transmission will occur within the cluster if the cluster head has higher values.Following the cluster head selection, information may be transmitted from the cluster head to the base station, allowing us to skip various transactions that take place at the base station and have a variety of negative effects on performance (Santhosh Kumar B.J., 2017).

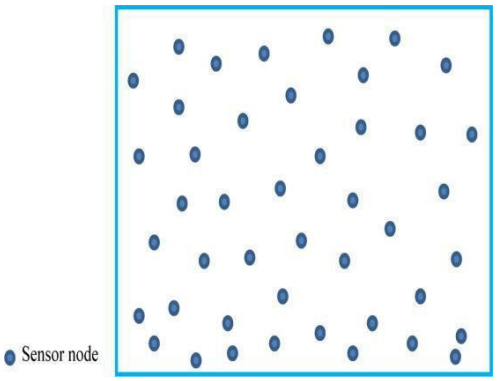


Figure 1 The Nodes Of The Sensor Network

Selecting CHs

The fitness values of nodes are compared to the fitness values of their neighbours. If a node's fitness value is greater than that of its peers, the node will promote as a CH; alternatively, the node will wait for an advert from a node with a better fitness value.

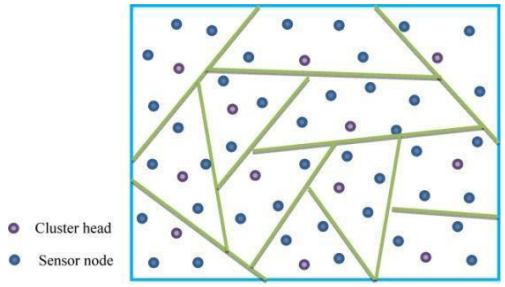


Fig ure 2 Clustered WSN With CH And Nodes

B) Algorithm for Cluster formation: -

1. ***Start***
2. ***Determine the centricity of nodes and the distance between nodes and sink***
3. ***Calculate the fitness value***
4. ***Sends fitness value to neighbour***
5. ***If the value highest, then sends CH promoted message and receive join message.***
6. ***Else Receive CH Promoted Message and Select CH.***
7. ***END***

It is a set of rules and algorithms that manage the whole network setup to enable flexible communications and connectivity between network nodes.

C) SDN (Software defined node):

In SDN, it is also referred to as a base station. It is a centered element that effectively acts as an anchor node. Depending on how many hops were utilized to convey the message and how far it was from the sensor node, different amounts of energy are needed to route a signal to the SDN node. We developed an SDN node using the washbasin node approach. In a network or diagram, normal nodes are used to create a point where lines or pathways converge or diverge. The normal nodes notice the intercepted data and route the packets to another accessible node specified in the nearby routing database (C. Gujjar, 2017). It creates a neighbour table to categorise nodes that are near together in order to construct a cluster (Karakus, 2017).

IV Designing of System

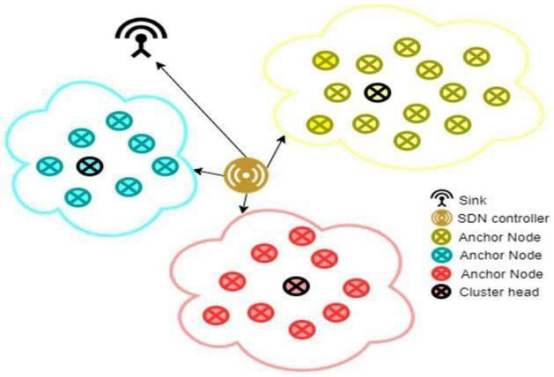


Figure. 3 Architecture of proposed system

Figure shows the architecture of proposed system, it content sink at the top and anchor nodes with cluster head in black colour. The SDN controller is situated at centre.

Workflow of system is as follows-

1) ***At first step we initialize the nodes.***
2) ***Clustering is done with the homo or heterogeneity nature of the node***
3) ***Select the cluster head with highest residual energy.***
4) ***Design SDN to Handle Cluster Head.***
5) ***Compare the work with respect to QoS Parameter.***
6) ***Performance analysed with Cooja Simulator.***
7) ***Compare the outcomes of the effort in terms of QOS measures like latency and energy left in each node.***

V ANALYSING THE RESULT

To Analyze the result, we do the SDN cluster base WSN on Simulator. For this purpose, we use Cooja simulator on Contiki OS specially designed for Wireless sensor Network. The outcomes of an SDN-based clustered network in terms of the percentage of packets delivered delay and end-to-end latency.

Following fig shows our simulation window of Coooja Simulator. Where we place 41sky motes deployed at different locations which form 02 clusters based on transmission range.

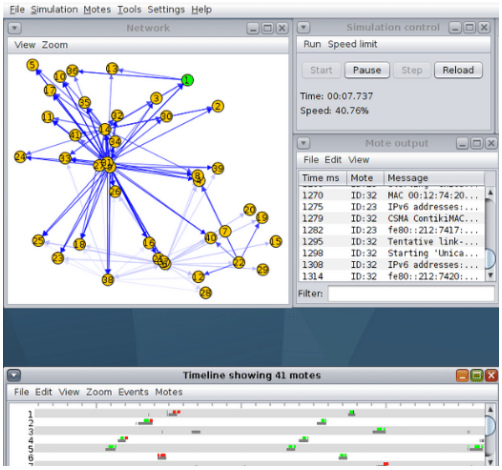


Figure 4. Simulation window of the network on Cooja

In the above network the cluster head (CH) are selected dynamically by OCHS (Optimal Cluster heads selection) algorithm. They elect cluster head (CH) based on received signal strength indicator (RSSI) value of node at any instance. At the centre we put one SDN Node which bridges the gap between two clusters which transmits the packet by looking neighbour table node in multi-hop communication way which has a path from source to sink or BS.

Following figure shows the information of each node.

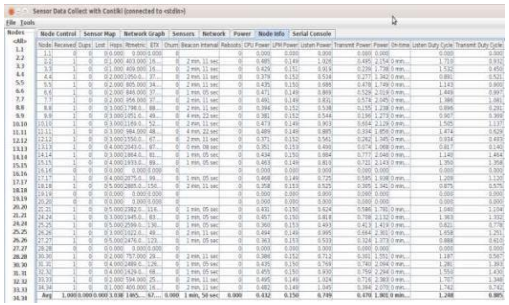


Figure 5 Node Information Table.

We chose 41 sky nodes from the above data supplied by the COOJA simulator collect view tool, and we could observe node id, received packet count, duplicate packet count, lost packet, hops, ETX range, beacon interval, listen and transmit duty cycle and many sorts of energy usage in joule and other units.

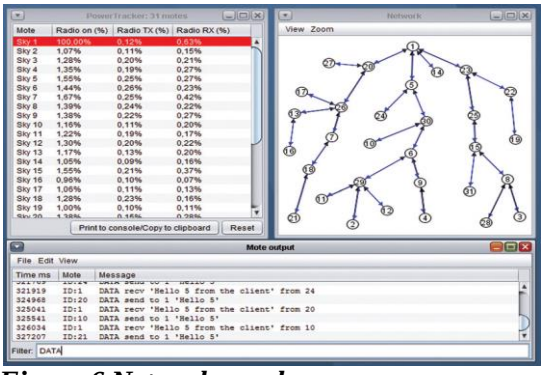


Figure 6 Network graph

Figure 6 shows the network graph which shows the connectivity between the nodes. When handling a large number of nodes through radio media and needing a visual depiction of how deployed nodes are connected to one another, this becomes handy for network developers through Cooja simulation. We compare on the simulation that the end to end delay quantity between network with SDN and without SDN. By seeing the simulation result we found that the end to end delay is less in SDN defined network even we have large numbers of nodes. And the network form without SDN has more delay and it is in direct proportion of numbers of nodes.

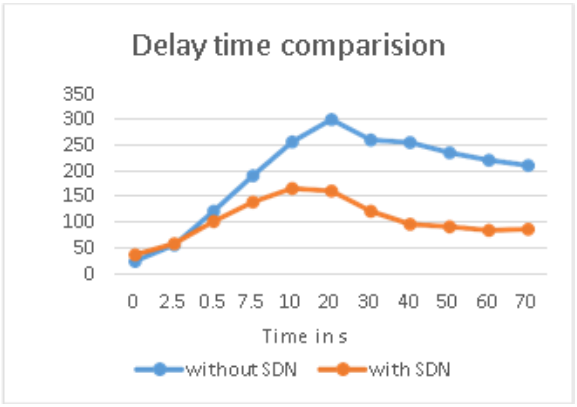


Figure 7 End to End Delay comparison graph

Now by considering energy matric after simulation with different simulation time it is observed that energy id continuously decreased in network without SDN as compared to network with SDN. In Network with SDN they can save more energy than network with without SDN which is major achievement of this approach which save the energy in critical situation as well *(Manoj, 2020)*. The comparison chart is depicted in figure 8.

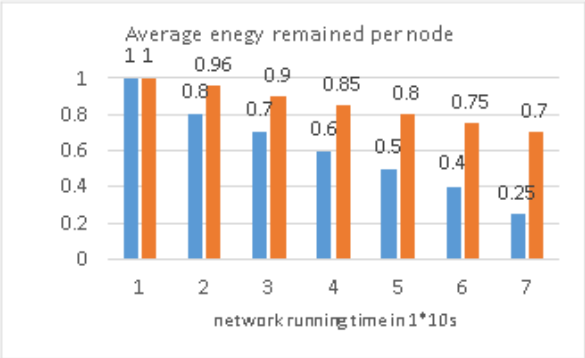


Figure 8 Total energy remained after simulation

When we compare the packet delivery ratio the simulation displays the results when we send various numbers of packets in networks with and without SDN. The packet received rate is significantly better in the SDN-controlled networks, whereas in the non-SDN networks, fewer packets are successfully delivered, making it more reliable to send data from one end to the other end. The comparison table and graph in indicated as below.

Table 1 End to End Packet delivery ratio

Numbers of packets sent over network	**Nos. of packets received without SDN**	**Nos. of packets received with SDN**
One Hundred	78	92
Two Hundred	161	183
Three Hundred	243	278
Four Hundred	314	371

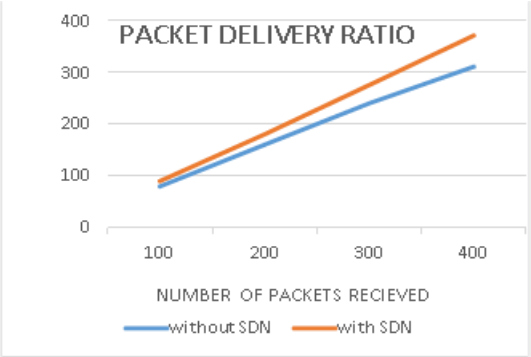


Figure 9Ratio of Packet delivery over both network

VI CONCLUSION

This study contributes to the development of a new way that organises a collection of nodes into a cluster depending on the transmission range, which will be managed by an intermediary SDN node. Our suggested work has been able to preserve more

energy after simulation, which will lengthen the life of the network. The packet delivery ratio is also greater, and the fact that the network latency has decreased so much implies that data is getting to its destination rather quickly. This work can promote our planned effort by demonstrating that the fundamental drawbacks of a network without SDN have been solved and that network dependability has improved with longer node lifetimes. The results in terms of WSN QOS parameters are significantly superior in this work when compared to works without software defined networks. In the future, we plan to increase data security, which will boost our business as a whole.

The overall packet delivery ration in Network with SDN is quite reliable as compared to network without SDN

References

1. Anand, S. &. ((2020, February)). Cluster-Based Energy Efficient Protocol for Wireless Sensor Network. *In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)* (pp. 1-7). IEEE.
2. Arun P Prabhan, S. A. (2019). Identifying Faulty Nodes in Wireless Sensor Network to Enhance Reliability. . *International Journal of Recent Technology and Engineering (IJRTE)*, 2277-3878, Volume-8 Issue-2.
3. Bello, L. L. (2020). Experimental assessments and analysis of an SDN framework to integrate mobility management in industrial wireless sensor networks. . *IEEE Transactions on Industrial Informatics*.
4. Bhuyan B, S. N. (2016, March). A QoS aware routing protocol in wireless sensor networks with mobile base stations. *International Conference on Internet of things and Cloud Computing* (p. 15). ACM.
5. C. Gujjar, S. S. (2017). A survey on wireless Ad- hoc implementations in android-based smartphones. *International Conference on Communication and Signal Processing (ICCSP)*, (pp. 0894-0899).
6. Chen Q, G. H. (2018, April). Energy-collision aware data aggregation scheduling for energy harvesting sensor network. *In IEEE INFOCOM 2018-IEEE conference on computer communications works* , (pp. (pp. 117–125). IEEE).
7. Duan, Y. L. (2017). A methodology for reliability of WSN based on software defined network in adaptive industrial environment. . *IEEE/CAA Journal of Automatica Sinica*, 74-82.
8. Elappila, M. C. (2018). Survivable path routing in WSN for IoT applications. *Pervasive and Mobile Computing*, 49-63.
9. Goyal N, D. M. (2018). A novel fault detection and recovery technique for cluster-based underwater wireless sensor networks. *Int J Commun Syst.* , 31(4).
10. Goyal N, K. K. (2018). An optimal scheme for minimizing energy consumption in WSN. *Global Res Dev J Eng.*, 1-7(8).
11. Haque, I. N.-G. (2018). SDSense: An agile and flexible SDN-based framework for wireless sensor networks. *IEEE Transactions on Vehicular Technology, 68(2), IEEE Transactions on Vehicular Technology, 68(2),* , 1866-1876.
12. Karakus, M. &. (2017). A survey: Control plane scalability issues and approaches in software-defined networking (SDN). *Computer Networks*, 112.
13. Laufer R, K. L. (2016). The capacity of wireless CSMA/CA networks. *IEEE/ACM Trans Netw (TON).* , 24(3).
14. Manoj, S. A. (2020). A Survey on Clustering Approaches to Strengthen the Performance of Wireless Sensor Network. *Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, (pp. 814-820). Coimbatore.
15. Nair, A. &. (2019). A performance booster for load balancing in cloud computing with my load balancer technique. *International Journal of Recent Technology and Engineering*, 8(1).
16. Santhosh Kumar B.J., K. V. (2017). Symmetric Key based Encryption and Decryption using Lissajous Curve Equations. *International Journal of Electrical and Computer Engineering (IJECE) Vol. 7*, 285~288.
17. Santosh Ananad, J. N. (2021). Enhancing Reliability of Wireless Sensor Network Without Affecting Other QOS Parameters. *International Conference on Signal Processing and Communication(ICPSC)*, (pp. 7-11). Coimbature.
18. Xu, H. Z. (2018, October). PRMS: Phase and RSSI based localization system for tagged objects on multilayer with a single antenna. *ACM International Conference on Modeling, Analysis and Simulation.*
19. Zhu, J. Z. (2017). Physical-layer security and reliability challenges for industrial wireless sensor networks. . *IEEE access*, 5313-5320(5).

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-12

Word Speech Recognition using Deep Learning and Object Mapping

Dr.Ranjana S. Zinjore
Assistant Professor
MTES's Smt. G. G. Khadse College, Muktainagar
rszinjore14@gmail.com

Dr.Tanuja K. Fegade
Assistant Professor
KCES's Institute of Management and Research, Jalgaon
tanujamahajan18@gmail.com

ABSTRACT - *Speech recognition models are very important in society. Speech recognition is the translation of spoken words into text. In this paper we have proposed a model for word speech recognition with convolutional neural networks. This model consists of different phases such as pre-processing, training and testing. To train the model English numbers as 0 to 9 are considered in words. The recognized word speech class label is mapped with respective objects. This model is used in the school for the students*

Keywords**:** *Speech Recognition, Spoken words, Deep Learning*

I. INTRODUCTION

Speech is the best way of communication for the human being in their native language [1]. Speech is used to convey important information quickly and precisely. The ability of a machine or programme to recognise words and phrases in spoken language and translate them into a machine-readable format is known as speech recognition [2]. There are numerous speech recognition applications available today, including speech-to-text, simple data entry, and voice dialling [2].

Statistical pattern recognition, communication theory, signal processing, combinatorial mathematics, and linguistics are only a few of the many distinct components that make up automatic speech recognition systems [2].

In automatic speech recognition, the conversion of spoken words into textremains a difficult problem. Speakers may, for instance, communicate in a variety of styles, at varying paces, and with varying degrees of emotion [3]. They may also have various accents, dialects, or pronunciations. Additional diversity is produced by background noise, reverberation, various microphones, and recording equipment [3].

In this paper we proposed a theoretical model for word speech recognition using convolution Neural Network (CNN) and object mapping i.e. displaying the class label object. We are focusing on speech recognition of English numeralsfrom 0 to 9 in words. The practical implementation of the model is used in school for the students.

II. LITERATURE REVIEW

AkhileshHalageri et al.[2] discussed different pattern matching abilities of neural networks on speech signals. Deep learning algorithms are useful for automatic speech recognition (ASR) due to the way words and phrases are spoken can be vastly altered by accents, dialects and mannerisms.

Yan Zhang suggested deep learning algorithms, including deep neural networks (DNN) and deep belief networks (DBN), for automatic continuous speech recognition [3].

A neural network-based approach based on lip-reading and acoustic speech recognition was created by Denis Ivanko et al. [4]. The first and most crucial stage in creating an automated lip-reading system is identifying a region-of-interest (ROI) that contains the mouth action. Thus, cropping this ROI (mouth region) from each frame of the movie is our initial goal. Then they handle acoustic speech processing by obtaining spectrograms of the utteredphrases from the raw audio data with its further processing by the integral CNN-LSTM network.

Jing Yu et. al. [5] proposed an automated English speech recognition using dimensionality reduction and deep learning (AESR-DRDL) approach. The proposed AESR-DRDL technique involves a series of operations, namely, feature extraction, pre-processing, dimensionality reduction, and speech recognition.

HardikDudhrejia et.al.[6] discussed different types of neural networks for speech recognition systems. AyadAlsobhaniet. al.[7] suggested a word-tracking modelby applying speech recognition features with deep convolutional neural learning. Start, stop, forward, backward, right, and left these six wordsfrom people of different ages are considered for preparing a dataset.

III. PROPOSED Architecture

Fig. 1 shows the proposed architecture of word speech recognition using deep learning and object mapping.

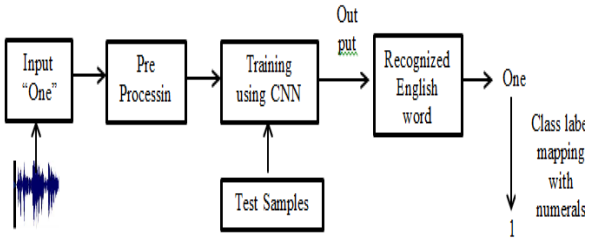


Fig.1: Word Speech Recognition using Deep Learning and Object Mapping

IV. METHODOLOGY

The proposed architecture mentioned in section 3 is implemented by following the given steps:

1) Collect the dataset (.wav files) intothe train_dir folder.
2) Pre-processed the dataset by resampling the audio file using a tested sample rate.
3) Declared the class_labels variable for storing the 10 classes of English numerals as Zero-Nine.
4) Reshape the audio file into 3-Dimension which is required for CNN.
5) Split the dataset into training (80%) and testing (20%) dataset.
6) Construct the CNN by mentioning the convolution layers with activation functions, flatten layer and dense layer.
7) Train the model and stored the training model
8) Test the model.

V. CONCLUSION

In this paper we proposed the model for word speech recognition using deep learning and object mapping. The basic steps required for the implementation of the model are also discussed. The implemented model will be used in society. In the future, to implement the model we collect the English words (Zero to Nine) from pre-primary students to recognize the voice and convert it into text. Finally in future the recognized class labels numbers are also displayed using object mapping.

REFERENCES

1. Speech Recognition: a review of the different deep learning approaches- https://theaisummer.com/speech-recognition
2. AkhileshHalageri , Amrita Bidappa , Arjun C ,. MadanMukundSarathy and Shabana Sultana, "Speech Recognition using Deep Learning", International Journal of Computer Science and Information Technologies (IJCSIT), 2015, Vol. 6 (3) ,3206-3209.
3. Y. Zhang, ''Speech recognition using deep learning algorithms,'' Stanford
4. Univ.,Stanford, CA, USA, Tech. Rep., 2013, pp. 1–5. [Online]. Available:
5. Y. Zhang, ''Speech recognition using deep learning algorithms,'' Stanford
6. Univ.,Stanford, CA, USA, Tech. Rep., 2013, pp. 1–5. [Online]. Available:
7. Y. Zhang, ''Speech recognition using deep learning algorithms,'' Stanford
8. Univ.,Stanford, CA, USA, Tech. Rep., 2013, pp. 1–5. [Online]. Available:

9. Yan Zhang, “Speech Recognition Using Deep Learning Algorithms”, Stanford University, 2013, pp.1-5.
10. AbdinabiMukhamadiyev, IlyosKhujayarov, OybekDjuraev and Jinsoo Cho, “Automatic Speech Recognition Method Based on Deep Learning Approaches for Uzbek Language”, Sensors 2022.
11. Denis Ivanko and Dmitry Ryumin, “Development of Visual and Audio Speech Recognition Systems Using Deep Neural Networks”, 31st International Conference on Computer Graphics and Vision, September 27-30, 2021.
12. Jing Yu, Nianhua Ye, Xueqin Du and Lu Han, “Automated English Speech Recognition Using Dimensionality Reduction with Deep Learning Approach”, HindawiWireless Communications and Mobile Computing, 2022
13. AyadAlsobhani, Hanaa M A ALabboodi and Haider Mahdi, “Speech Recognition using Convolution Deep Neural Networks”, Journal of Physics, IICESAT Conference, College of Material Engineering, University of Babylon, Iraq, 2021,
14. HardikDudhrejia and Sanket Shah, “Speech Recognition using Neural Networks”, International Journal of Engineering Research & Technology (IJERT), Vol. 7 Issue 10, October-2018, ISSN: 2278-0181

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-13

Crop based Disease and Pesticides Recommender System

Prof. Abhishek S. Patil
Assistant Professor
MTES's Smt. G. G. Khadse College, Muktainagar
abhishek.patil.1987@gmail.com

Dr.Ranjana S. Zinjore
Assistant Professor
MTES's Smt. G. G. Khadse College, Muktainagar
rszinjore14@gmail.com

Prof. Satish R. Damade
KCES's Institute of Management & College, Muktainagar
satish.damade@gmail.com

Abstract: *India is an agriculture based country. To increase the productivity in agriculture technology plays an important role. A farmer also frequently deals with pests and diseases that might harm the crops they raise but that they are typically ignorant of at an early stage. In this paper we proposed a model that recommended the disease name and pesticides requiredfor controlling the disease as per the crop names. A database of 30 different crops is considered for recommendation which is used in the proposed algorithm.*

Keywords: Crop Disease, Recommender system, Pesticides

I. INTRODUCTION

India is an agricultural country with over 1.2 billion people, about 70% of which depend on agriculture. Agriculture is the main source of income for the Indian people and it also impacts the Indian economy [1].Due to rapid growth of population and increase in demand for food, Agriculture plays an important role.Farmers grow varieties of crops based on their requirements but due to some climatic changes various diseases affects the growth of the plants and damage the whole crop. This disease reduces the quantity and quality of the agricultural product. In this paper basic 30 crops are considered which are mostly growing in the Maharashtra region. A recommender system is used to suggest the crop wise disease name and pesticides required.

Avinash Kumaret. al. [2] used different classifiers like Support Vector Machine, Decision Tree and Logistic Regression for identification of the best suitable crop. R. Saliniet. al. [3] proposed a rice plant leaf diseases area affected part identification and solution for recovery. Nikhil Patilet. al. [4] developed a mobile based system for pesticide recommendation and the amount ofpesticide to be used for an unhealthy crop.TanmayThoratet. al. [5] proposed insecticide recommendation and fertilizer recommendation model. In this model they used Transition Probability Function (TPF) andConvolution Neural Network (CNN) to process the pest's image discretely and continuously for applying therecommended insecticide.For the predicting, identification, and categorization of diseases and pests, Tiago Domingues et al. [6] examined the employment of machine learning-based approaches. By encouraging the development of methods that will enable farmers to reduce their use of pesticides and chemicals while conserving and enhancing their crop quality and productivity, this study hopes to advance smart farming and precision agriculture.Taranjeet Singhet. al. [7] discussed that Crop failure, a lack of sufficient information, crop

damage caused by ignorance or carelessness, a lack of professional support, and lack of access to agro-tech solutions are some of the issues that farmers in our nation are currently facing. The suggested tools will be provided by CROFED to assist farmers in resolving these issues: Crop Recommendation System, Fertilizer Suggestion System, and Crop Disease Detection System.

II. METHODOLOGY

Fig.1 shows the working of the recommender system for disease and pesticides based on crop name. Total 30 different crops are considered in the database with concerned disease name and pesticides required shown in Table 1.

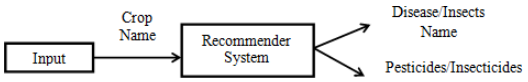


Fig.1: Recommender system for disease and pesticides

Table1: Sample crop database

Crop Name	Disease/Insects	Pesticides/Insecticides
Rice	Karpa	Copper Oxychloride 50% spray
Bajra	Khodkidi	Melotheon 50% + 10% water spray
Jowar	Karpa	Copper Oxychloride 4gm per litre water
Wheat	Karpa	Copper Oxychloride (0.2%) + Mancozeb(0.2%)
Maize	Khodkidi	Clinophos 35% 15 mL + 10L water spray
Gram	Ghateali insect	Nimboli Powder spray
Tur	Ghateali insect	HNPV
Mung	Juice sucker insects	Dimethoate 15 mL + 10L water spray
udid	Hairy insects	Dimethoate 15 mL + 10L water spray
Groundnut	Tikka	Mancozeb 25gm +10L water spray
Soyabean	Tambera	Propyconzol 1mL + 1L water
Sunflower	Ghateali	HNPV
Sesamum	Karpa	Copper Oxychloride 1.25kg + strptocyclin 50gm per 500L water spray
Safflower	Mava insect	Rogar 30% + 500L water spray
Sugarcane	Kani insect	Bavistin 0.1%(100L + 100gm)
Cotton	Karpa	Popycolyazol 500gm + 500L water spray
Mango	Karpa	50% Carboril + Carbendzim 10gm + Mancozeb 2gm per 10L water
Banana	Karpa	10gm Carbendzim + 10L water
Grapes	Bhuri	5mL denocap + 10L water spray
Guava	Mar	Trycodarma powder (100gm per tree)
Chicoo	seed burrowing insect	Deltamethrene 10mili per 10L water spray
Coconut	Koli insect	Diacofal 22 mL per 10L water spray

2.1 Algorithm:

disease_pesticide_recommnder(cname,d,p)

Input: crop name

Output: d=disease and p= pesticides

i) Stored the entered crop name

cname-> enter crop name

ii) Read the crop database and stored the first column into crop variable

[num,txt,raw] = xlsread('crop dataset.xlsx',2);

crops = raw(:,1)

iii) Compared the crop name and find the index

ans = strcmp(crops,cname)

[index,val] = find(ans);

iv) Used index to determine disease and pesticides

disease = raw(index,2);

pestricites = raw(index,3);

III. CONCLUSION

To increase the productivity in farming, which types of diseases occurs in the crops this knowledge is important. With the help of this knowledge the farmer improve the quality of crop. A recommender system is proposed in this paper using the disease_pesticide_recommnder algorithm to determine the disease name and pesticides amount by entering the crop name. In future along with the disease the images of that crop diseases also shown.

REFERENCES

1. Rani Holambe, PoojaPatil, PadmajaPawar, SaurabhSalunkhe and Hrushikesh Joshi, "IOT based Crop Recommendation, Crop Disease Prediction and Its Solution" , International Research Journal of Engineering and Technology (IRJET), Vol. 7 Issue: 10, Oct 2020, e-ISSN: 2395-0056.
2. Avinash Kumar, SobhangiSarkar and ChittaranjanPradhan, "Recommendation System for CropIdentification and Pest Control Technique inAgriculture", IEEE International Conference on Communication and Signal Processing, April 4-6, 2019
3. R. Salini, A. J. Farzana and B. Yamini, "Pesticide Suggestion and Crop Disease Classification using MachineLearning", International Journal of Engineering Science and Computing (IJESC), Vol. 11, Issue 4, April 2021
4. Nikhil Patil, Rajab Ali, VaibhavWankhedkar and DeepaliNayak, "Crop Disease Detection using Deep ConvolutionalNeural Networks", International Journal of Engineering Research & Technology (IJERT), Vol. 8 Issue 03, March-2019.
5. TanmayThorat, B.K. Patle and Sunil Kumar Kashyap, "Intelligent insecticide and fertilizer recommendation system based on TPF-CNN for smart farming", Journal of smart agriculture technology, September 2022.
6. Tiago Domingues, TomásBrandao and Joao C. Ferreira, "Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey", Agriculture journal, September 2022.
7. Taranjeet Singh, SaurabhAnand, AnmolSehgal, SiddheshMahajan and PranotiKavimandan, "CROFED - Crop and Fertilizer Recommendation and Disease diagnosis system using Machine Learning and Internet of Things", International Journal Of Innovative Research In Technology, Vol. 9, Issue 2, July 2022, ISSN: 2349-6002

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-14

Supervised Machine Learning Algorithm for Optical Character Recognition.

Asst. Prof.Vasundhara R. Fegade[1]
[1]Department of Computer Science,Dhanaji Nana Mahavidyalaya,Faizpur.
[1]vrfegade@gmail.com

Dr. G.R.Wani[2]
[2]HOD,Department of Computer Science, BhusawalArts,Science and P.O. Nahata Commerce College,Bhusawal.
[2]Gajuwani03@gmail.com

Machine Learning is one of the important application of Artificial Intelligence. It provide machine an Intelligence i.e. the ability to learn from data automatically. Supervised learning is category of Machine Learning and Artificial Intelligence. In supervised learning machines are trained using labelled training dataset to classify data and predicts outcome. In Supervised Learning first machines are trained using training dataset .once the machine become trained , testing is applied on it on the basis of test data. There are two problems in supervised learning Regression and Classification. This paper studies classification algorithms and how they work on Optical Character Recognition.

Keywords:Machine Learning, Supervised Learning, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbour, Naive Bayes

I. INTRODUCTION

Machine learning is subfield of Artificial Intelligence that focuses on design of system that can learn from and make decisions and predictions based on the experiences which is data in the case of machine. Machine Learning enables computer to act and make data driven decisions rather than being explicitly programmed to carry out a certain task. These programs are designed to learn and improve over time when exposed to new data .Most of the people think that Artificial Intelligence,Machine Learning and Deep Learning all are same. They are wrong. Artificial Intelligence is a broader concept of machine being able to carry task in smarter way. Itcovers anything which enables the computer to behave like humans. Machine Learning is a subset or a current application of AI. It is based on the idea that we should be able to give machine the access to data and let them learnt from themselves. Machine Leaning deals with extraction of pattern from dataset. This means that machine can only find the rules for optimal behaviour but also can adopt to the changes in the world. Deep Learning is subset of Machine Learning where similar Machine Learning algorithm are used to trained Deep Neural Network so as to achieved better accuracy. In those cases former wa not perform up to the map.
There are three types of Machine Learning:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning.

In this paper we focused on Supervised learning. Supervised learning is where you have input variable X and output variable Y and you used an algorithm to learn the mapping function from the input to the output.
Y=f(X) .
The goal is to approximate the mapping function. The goal is to approximate the mapping function. Whenever you have a new input data X you could predict the output variable for that data.

Risk Management, Image Classification, Fraud Detection, Spam Filtering are some applications of Supervised learning.

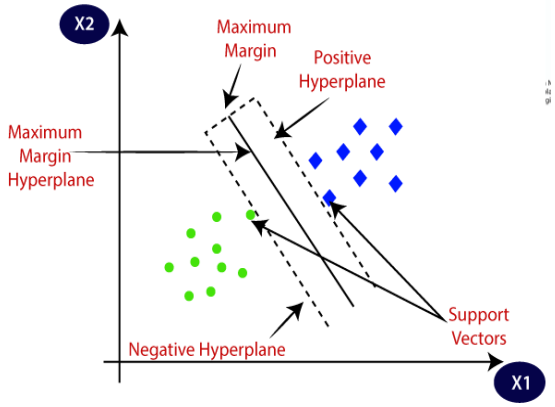
Out of these applications we focused on Image Classification. In this paper we study various supervised machine learning algorithms those can be used for character recognition.

1. SUPPORT VECTOR MACHINE (SVM) CLASSIFIER:

SVMs, or support vector machines, are at the forefront of ranking algorithms and have drawn particular interest from the global scientific community. There are numerous effective SVM-based applications in a variety of knowledge domains, including text categorization, digital image analysis, character recognition, and bioinformatics. [1].

The image from the training dataset decompose into sequence of character. Each sub image is used to calculate width and height at a particular font size [2]. SVM has primarily two class classifier. Here the optimization criterion is width of the margin between classes, means the distance to the nearest training pattern is the empty area around decision boundary [3].

SVM is computation algorithm that construct hyperplane or a set of hyperplane to Separate two linearly separable classes. Misclassification is ignored using hyperplane. e.g. All the points of Class A is labelled as +1, and all the points belong in class B labelled as -1. There may be many hyperplanes that classify the same set data as shown in figure 1.



Objective of SVM is to find best separation hyperplane using highest margin. Highest margin is the distance between the nearest points of the two classes. Larger the margin, Lower is the generalization error of the classifier.

2. DECISION TREE CLASSIFIER:

Decision tree is a binary tree. It is made of nodes and edges. At every internal node you are asking a question and that question has yes/ no answer. Leaf node contain decision. In decision tree classifier, we will use this type of binary tree structure to make decisions based on the values of different features.

When the input data is well structured decision free classifier is suitable for OCR [4].Decision tree algorithm is based on human thinking ability while making decision, so it is very easy to understand.

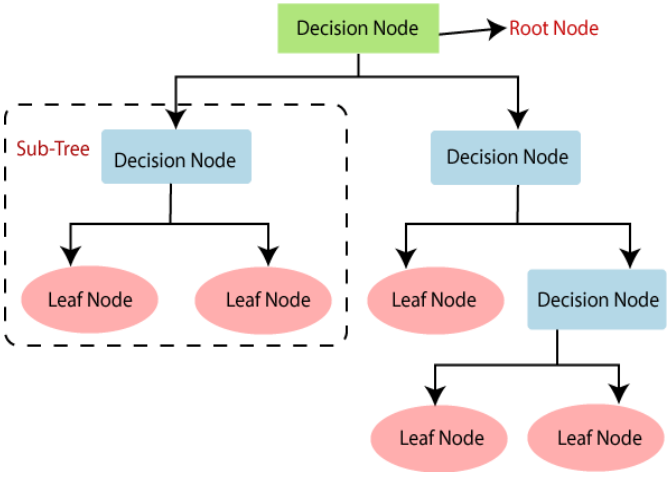
Terminology [6]:

Root Node- It is starting point of decision tree. It represents entire dataset. That is further classified into two or more homogeneous set.

Leaf Node – It is a final output node and the tree cannot be further classified.

Spitting - It is the process of dividing root node or any internal node into sub-nodes as per condition.

Purning - In this process all the unwanted branches are remove from the tree.



Decision Tree algorithm -

Step 1: Starting the tree with root node S. It is a complete dataset.

Step 2: Using Attribute selection method (ASM) is used to find best attribute in the dataset.

Step 3: Divide root node S into subset that contains values for best attributes.

Step 4: The node which has best attribute is the decision node.

Step 5: Following these steps repeatedlyuntil a stage is reached where you cannot further classify the nodes. It is final node or leaf node.

The main challenge in decision tree algorithm is to select the best attribute for sub nodes and root node. So for this Attribute selection measure (ASM) is the technique using this we can easily select the best attribute for the node.

ASM Techniques

1. Information Gain
2. Gini Index

3. K-NEAREST NEIGHBOUR CLASSIFIER:

KNN is the non-parametric machine learning algorithm used for classification. The general idea is very simple. To classify a new character, the

algorithm, find out k-nearest neighbour from the training dataset. To find out the nearest neighbour distance function is required [7] . Euclidean distance is used to find distance between the test point and all the referenced points. After that all the distances are arranged in ascending order. The reference point having k- smallest Euclidean distance is the k-nearest neighbour of testing point. The testing point is labelled with the same class label as the label of majority of its k-nearest neighbour.KNN is the better computational efficiency than other classifier[5].

Euclidean distance:

Euclidean distance is a distance metric works on the principle of the Pythagoras theorem. It gives shortest distance between two points using the formula:

$$d = ((p_1 - q_1)^2 + (p_2 - q_2)^2)^{1/2}$$

4.Naive Bayes Classifier:

Naïve Bayes is supervised machine learning probabilistic classifier, means it predict on the basis of probability. The classifier work on probability of occurrences of certain features using Bayestheorem[9,10].

Bayes theorem is used to determine the probability hypothesis with prior knowledge. It used conditional probability. Bayes formula is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

P(A/B) is Probability of hypothesis A on observed event.

P(B/A) is Probability of evidence given that the probability of a hypothesis is true.

P(A) is probability of hypothesis before observing the evidence.

P(B) is probability of evidence.

II. CONCLUSION:

All the classifiers describe above are well suited for character recognition. SVM can handle high-dimensional data that are robust to noise. Decision tree and K-Nearest Neighbour classifiers are easy to understand and well suited where input data is well structured. Naïve Bayes are well suited for multiple classes.

REFERENCES:

1. X. Ding, Ed., Advances in Character Recognition. InTech, 2012. doi: 10.5772/2575.

2. Kavya N, Madhushree N, Maheshwari S, Manasa H C, Anandhi G, 2019, Handwriting Recognition Using SVM, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCRACES – 2019 (Volume 7, Issue 10),

3. Mahesh Jangid, RenuDhir, Rajneesh Rani, Kartar Singh," SVM Classifier for Recognition of Handwritten Devanagari Numeral", 2011 International Conference on Image Information Processing (ICIIP 2011)

4. Thaker, Hetal&Kumbharana, Chandresh. (2014). Structural Feature Extraction to recognize some of the Offline isolated Handwritten Gujarati Characters using Decision Tree Classifier. International Journal of Computer Applications. 99. 10.5120/17452-8381.

5.Admassu, Tsehay& Nair, Pramod. (2019). Handwritten digits recognition with decision tree classification: A machine learning approach. International Journal of Electrical and Computer Engineering (IJECE). 9. 4446. 10.11591/ijece.v9i5.pp4446-4451.

6. BahzadTaha Jijo1*, Adnan Mohsin Abdulazeez2,Classification Based on Decision Tree Algorithm for Machine Learning",Journal of Applied Science and Technology Trends,Vol. 02, No. 01, pp. 20 – 28 (2021) ISSN: 2708-0757

7.Wang, Wei. (2014). Optical Character Recognition, Using K-Nearest Neighbors.

8.OndrejBostik∗ Jan Klecka ,"Recognition of CAPTCHA Characters by Supervised Machine Learning Algorithms",IFAC PapersOnLine 51-6 (2018) 208–213

9. Chen, H., Hu, S., Hua, R. *et al.* Improved naive Bayes classification algorithm for traffic risk management. *EURASIP J. Adv. Signal Process.* **2021**, 30 (2021). https://doi.org/10.1186/s13634-021-00742-6

10. Rish, Irina. (2001). An Empirical Study of the Naïve Bayes Classifier. IJCAI 2001 Work Empir Methods ArtifIntell. 3.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-15

An Overview Of Text Steganography Techniques

Mr.Yogesh Narayan Chaudhari,
Assistant Professor,
Department of Computer,
KCES's Institute of Management and Research, Jalgaon.
yogeshchaudhari@imr.ac.in ,
chaudhariyogn@gmail.com

Dr.Bhojraj H. Barhate
Associate Professor, Vice-Principal,
Department of Computer,
Bhusawal Art, Science & P. O. Nahata Commerce College, Bhusawal.
bhbrama123@gmail.com

ABSTRACT:

Steganography is one of the types in information hiding used to hide message to ensure it cannot be recognized by human vision. Hiding secret information in text file is known as text steganography. This paper explores steganography and its techniques that primarily focus on text steganography. Types of steganography classifications were discussed, that consist of the natural key used, and the embedding techniques. Text steganography can then be further separated as per the technique used.

Keywords: Information security, Information hiding, Steganography, Text steganography Techniques, StegaAnalysis

I. INTRODUCTION

Steganography is a Greek word which means hidden writing. The word "steganos" means "hidden" and "graphical" means "writing". Steganography is an art of hiding data inside any digital medium like audio, image, video, text, protocol etc. In Steganography various terminologies are used these are:[1]

1. COVER OBJECT: Cover Object is the medium to embed the data. Text, audio, video, images are used for embedding data. It is also known as vessel object.
2. SECRET DATA/MESSAGE: The data which is to be embedded in a cover object is known as secret message.
3. STEGO OBJECT: It is the output result obtained after embedding the secrete message into the cover object.

In Text Steganography various techniques are used to hide the information but primarily all the techniques are categorised under three main categories.

1) Changes Format of Text
2) Replaces Text with Synonyms
3) Change the position of Text

II. TEXT STEGANOGRAPHY TECHNIQUES

Here are the Techniques which primary based on basic techniques Text Steganography

Word-level technique: This technique can be implemented in several ways. One common

method is to use a thesaurus or other lexical resource to generate a list of synonyms for each word in the original message. [2][3]

The steganographer can replace certain words in the message with their synonyms, according to a predetermined scheme or algorithm.

For example, consider the following sentence:

"The quick brown fox jumps over the lazy dog.“

Using the word-level technique, the steganographer might replace certain words with their synonyms to create the following hidden message:

"The swift brown fox leaps over the indolent dog."

Sentence-level technique: This technique involves embedding secret data within the structure of a sentence, such as by rearranging words or changing the order of clauses. The resulting modified sentences are still grammatically correct and appear similar to the original sentences.[3]

One way to implement the sentence-level technique is to use sentence rearrangement. The steganographer can reorder the words or phrases in a sentence in a way that still preserves the sentence's grammatical correctness and meaning but also embeds a hidden message.

For example, consider the following sentence:

"The cat is sleeping on the windowsill. “

Using sentence rearrangement, the steganographer might change the sentence to:

"On the windowsill, the sleeping cat is."

Grammar-based technique: In this technique, secret data is embedded within the grammatical structure of the cover text, such as by using synonyms, antonyms, or homonyms that have a specific meaning in the context of the text.

The grammar-based technique involves using various techniques to modify the grammar of a sentence to encode a message. For example, a steganographer could use homophones, which are words that sound the same but have different meanings, to create a message. In this case, the steganographer would replace certain words in the original sentence with homophones that together form a hidden message.[4]

Another approach is to use a homonymic technique, where words that sound the same but have different meanings and spellings are used to encode a message. In this technique, the steganographer would replace certain words in the original sentence with homonyms that together form a hidden message.

A third approach is to use a punctuation-based technique. In this technique, the steganographer uses specific punctuation marks or combinations of punctuation marks to encode the message. For example, the steganographer could use a series of commas, semicolons, or parentheses to encode a message.

Steganography using whitespace: This technique involves hiding the secret message within the whitespace between words or lines in the cover text. The whitespace can be manipulated by changing the font size, kerning, or line spacing to create a pattern that represents the secret message.[5][6]

The process of hiding a secret message using whitespace typically involves the following steps:

Encoding the secret message: The first step is to encode the secret message as binary data. Each character in the message is represented by a sequence of bits, which can be 8 bits (1 byte) for ASCII characters or more for Unicode characters.

Choosing a whitespace encoding scheme: The next step is to choose an encoding scheme that will be used to represent the binary data using whitespace. For example, one scheme might use the number of spaces between words to represent a 0 bit and the number of spaces and tabs to represent a 1 bit.

Embedding the binary data: The binary data is then embedded within the cover text by manipulating the whitespace. For example, if the first character in the secret message is represented by the bit sequence 01100001, the number of spaces between the first two words in the cover text can be changed to represent the first bit (0), and the number of spaces and tabs between the second and third words can be changed to represent the second bit (1).

Decoding the binary data: To retrieve the secret message, the recipient needs to know which whitespace encoding scheme was used to encode the binary data. The recipient can then decode the binary data by analyzing the whitespace in the cover text and converting it back into binary data.

Steganography using font manipulation: This technique involves hiding the secret message by changing the font of specific characters in the cover text. The different fonts can be used to

encode binary data, which can then be decoded to retrieve the secret message.[7]

The process of hiding a secret message using font manipulation typically involves the following steps:

Encoding the secret message: The first step is to encode the secret message as binary data. Each character in the message is represented by a sequence of bits, which can be 8 bits (1 byte) for ASCII characters or more for Unicode characters.

Choosing fonts: The next step is to choose a set of fonts that will be used to represent the binary data. Each font is assigned a unique binary code, such as 00 for Arial, 01 for Times New Roman, and so on.

III. STEGANALYSIS

Steganalysis is the process of detecting the presence of hidden messages in stenographic media. In the case of Text Steganography, there are several tools that can be used for steganalysis to detect the presence of hidden messages. Here are some examples:

StegExpose: StegExpose is a steganalysis tool that can detect hidden messages in text, images, and audio files. It uses statistical analysis to detect anomalies in the media file that may indicate the presence of hidden data.[8]

Gargoyle: Gargoyle is a steganalysis tool that can detect hidden messages in text files. It works by analysing the frequency distribution of characters in the text file to identify any unusual patterns that may indicate the presence of hidden data.[9]

Stegdetect: Stegdetect is a steganalysis tool that can detect hidden messages in images and audio files. It uses a variety of detection methods, including statistical analysis, checksums, and pattern recognition, to identify the presence of hidden data.[4]

SteganoG: SteganoG is a steganalysis tool that can detect hidden messages in text, images, and audio files. It uses statistical analysis, frequency analysis, and pattern recognition to detect anomalies in the media file that may indicate the presence of hidden data.[10]

IV. CONCLUSION

Text steganography is a technique of hiding secret data within a text message to maintain

confidentiality and integrity. Each technique has its own advantages and disadvantages, and the choice of technique depends on the requirements of the application. Steganalysis tools such as Gargoyle, SteganoG, and Stegdetect are used to detect the presence of hidden data. To ensure secure communication, it is essential to use strong encryption techniques along with steganography.

V. REFERENCES:

[1] S. R. Yaghobi and H. Sajedi, “Text steganography in webometrics,” Int. J. Inf. Technol., vol. 13, no. 2, pp. 621–635, Apr. 2021, doi: 10.1007/s41870-020-00572-z.

[2] M. A. Majeed, R. Sulaiman, Z. Shukur, and M. K. Hasan, “mathematics A Review on Text Steganography Techniques,” 2021, doi: 10.3390/math9212829.

[3] M. Taleby Ahvanooey, Q. Li, J. Hou, H. Dana Mazraeh, and J. Zhang, “AITSteg: An innovative text steganography technique for hidden transmission of text message via social media,” IEEE Access, vol. 6, pp. 65981–65995, Aug. 2018, doi: 10.1109/ACCESS.2018.2866063.

[4] S. Shakir Baawi, M. Rosmadi Mokhtar, and R. Sulaiman, “A COMPARATIVE STUDY ON THE ADVANCEMENT OF TEXT STEGANOGRAPHY TECHNIQUES IN DIGITAL MEDIA,” vol. 13, no. 5, 2018, Accessed: Jun. 24, 2021. [Online]. Available: www.arpnjournals.com.

[5] L. Y. Por, K. Wong, and K. O. Chee, “UniSpaCh: A text-based data hiding method using Unicode space characters,” J. Syst. Softw., vol. 85, no. 5, pp. 1075–1082, 2012, doi: 10.1016/j.jss.2011.12.023.

[6] S. Roy and P. Venkateswaran, “A Text based Steganography Technique with Indian Root,” Procedia Technol., vol. 10, pp. 167–171, 2013, doi: 10.1016/j.protcy.2013.12.349.

[7] N. Naqvi, A. T. Abbasi, R. Hussain, M. A. Khan, and B. Ahmad, “Multilayer Partially Homomorphic Encryption Text Steganography (MLPHE-TS): A Zero Steganography Approach,” Wirel. Pers. Commun. 2018 1032, vol. 103, no. 2, pp. 1563–1585, May 2018, doi: 10.1007/S11277-018-5868-1.

[8] E. Olson, L. Carter, and Q. Liu, “A

Comparison Study using Stegexpose for Steganalysis.," Int. J. Knowl. Eng., pp. 8–12, 2017, doi: 10.18178/IJKE.2017.3.1.079.

[9] M. Taleby Ahvanooey, Q. Li, J. Hou, A. R. Rajput, and C. Yini, "Modern Text Hiding, Text Steganalysis, and Applications: A Comparative Analysis," Entropy, vol. 21, no. 4, p. 355, Apr. 2019, doi: 10.3390/e21040355.

[10] R. Kumar and H. Singh, "Recent Trends in Text Steganography with Experimental Study," in Handbook of Computer Networks and Cyber Security, Cham: Springer International Publishing, 2020, pp. 849–872.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-16

Artificial Intelligence based Smart ATM

Prof. PaurnimaN.Patil
DOCS and IT,BAS P.O. Nahata College Bhusawa
patilpaurnima291@gmail.com

Prof.Vishakha M. Kolte
DOCS and IT,BAS P.O. Nahata College Bhusawal.
koltevishakha5@gmail.com

ABSTRACT - *In the modern world we witness ATM system for making financial transactions with withdrawal money, money transfer, checking savings account etc. above current ATM system uses ATM card and PIN for this validation purpose. This system makes many people suspicious Security issues. Therefore, in this paper we propose a advanced ATM security system to prevent ATM theft. Using fingerprint authentication for user access account using it's unique ID given in the bank details. In this process, bankers will collect customer's fingerprints, mobiles number, e-mail and name. Specific to these personal details the user will be stored in the back end which is in banks database. This developed system offers more secure insurance instead of the current system that relies solely on identity, we are proposing some additional features such as automatic door system during theft with message policing and alarming at the same time. Image processing is used CCTV as well as in the cash tray. There is an artificial innocence program used to eliminate illiteracy and facilitate communication. The main objective of the proposed system is to prevent theft multiple without relying on PIN number as security levels are enhanced and enforced.*

Keyword-: *ATM, biometric, GSM, image processing, artificial intelligence.*

I. INTRODUCTION

An ATM is an automated teller machine that a computerized telecommunication device that provides customers with access to public financial transactions place that does not require a human clerk or bank teller. The first ATM was installed in Enfield, London in June27, 1967 by Barclays bank. ATM's are given various names Other names like Automated transaction machine , Automated banking machines, Cash Points (in Britain), Hole in the wall, Ban Comet (in Europe and Russia), Ever Paisa (India). In the modern world, money is needed anytime or anywhere for travel, shopping and health. We need to carry money in times of emergency. That increases the risk robbery at any moment, where the bank is the safest place keep our money. It provides automated tailoring machines (ATM) which can dispense money at any time location. So ATM has given us all the benefits. We adding some more security is encouraged here features in the current ATM system and thereby here we propose further security by adding paper a second phase of deregulation of the existing ATM system and on existing security systems using biometric authentication and AI for easy communication. Validation and verification is always part of the concern customer security and privacy. Rapidly maintaining integrity is not easy in a changing world a person's authenticity. We need to prevent fraud some full proof security solutions that we can use together with the currently available technology. We use artificial intelligence biometric authentication, automatic door lock system, process the image inside the cash tray to detect insufficient funds and currency scanning so torn notes will be scanned and removed. Also using this feature we will search and limit access to ATM room's at large number of heads. A combination of all these technologies ATM fraud can

help reduce and therefore can improving the security level of other financial transactions.

II. LITERATURE SURVEY

- **Biometric authentications to control ATM theft**

This paper proposes that route banking and the transaction system is changing in the world, authentication, validation and confirmation of a person is too much should be important and of greater concern. So biometric is one of the best solutions for system and can security added to maintain authentication and privacy. Instead of just going for PIN identification and verification fingerprint scanning, iris scan, palm scanning can be used.

Even we can use voice recognition. Such a combination technology can help reduce ATM fraud and it can improve other financial security levels transactions. There are many ways of biometric scanning these are

- ✓ Fingerprint verification
- ✓ Face recognition
- ✓ Scanning of retina
- ✓ Scanning of hand geometry
- ✓ Scanning of vein geometry
- ✓ Iris scanning
- ✓ Signature biometrics
- ✓ Voice analysis

- **Card less ATM system:**

In this paper it is proposed to add an embedded system that uses biometrics or pin identification. ATM system is susceptible to many types of fraud such as ATM card theft, lebanese loop, skimming etc. This system will perform personal maintenance back-end user information in the bank's database end .Fingerprint of the entire user who wants access account taken. The system first acquires a fingerprint. If a particular user wants to transact, if the fingerprint matches, the user gets direct access without PIN process or any other process. If password or fingerprint matches then it can be transacted.

III. EXISTING METHOD

When the ATM card is inserted into the card slot, the information is received two card readers read the magnetic strip present in the card slot. One card reader searches specifically a code that confirms that the card is genuine. Second card reader what captures the account number and password to check you entered. If the authentication is successful, the ATM and telephone connects to the bank server through the network. Now user can do bank transaction and when once the transaction is completed, the card comes out through the ATM slot and the user is automatically logged out. There is a counting machine attends to count the number of notes and receives receipt through the printer that informs you transaction completed.

User needs to perform one of the following transactions:

- Cash transfer
- Cash Withdraw
- Balance Enquiry
- Password change

This is just general method of what happens.

IV. PROBLEM STATEMENT

There are many methods that individuals can hack into the customer's account, can steal money from the customer's account. It is as follows.

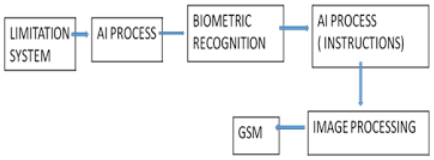
Skimming: A device is installed in the card slot and all the information is read on the card. A small pinhole camera is fixed above the keypad that lifts all this information. With the help of this information, a duplicate card is created by a criminal and can be accessed customer account.

- **Lebanese Loop:** In this type of fraud, the criminal establishes a device that locks the card into the slot the time when the machine completes the transaction card is now pushed out as the card is locked. A customer moves away from his place to complain and by then the criminals will come out and unlock the card and will withdraw all money from users account.
- **Card lost or stolen:** If the card is lost or stolen at the time it is also possible that user's accounts can be accessed by criminal. Criminals can gain access to you through many such techniques if the account uses biometrics with some additional security so this fraud can be reduced.

V. PROPOSED METHOD

- ✓ In our proposed system, apart from biometric we are accessing some more features in the ATM system like
- ✓ Automatic door lock
- ✓ AI based image processing and currency scanning

- ✓ Voice accessible machine using AI
- ✓ Insufficient deposits may be detected in the tray
- ✓ Inform the police if you find anything wrong using GSM module



Block Diagram of Proposed Method

- **Limit System-** This system has CCTV as part of hardware and image processing as a software part in which the system finds the numbers limit the principals and persons present in the ATM room access at will as per bank regulations.
- **AI Process –** CCTV sends signal to ATM the machine where the AI responds to the extra head to go out and wait and it also welcomes the user.
- **Biometric identification –** as information the user is already stored on the cloud by the bank. By using the technology will save time and be accessible to the user access with biometrics only and not required follow the old long process.
- **AI Notification-** Here without touching the ATM machine or buttons the user uses voice commands withdrawal or any other action. The user can access by speaking the local language.
- **Image Processing –** This technique is used in ATM machine cash tray to find insufficient funds and torn notes.
- **GSM -** The GSM module will send messages to the user and bank for withdrawal.

Block Diagram for Avoiding Theft

- ✓ **Auto door lock mechanism -** during theft a CCTV or motion sensor will detect illegal activity. A signal will be given to the stepper motor to lock the door.
- ✓ **GSM-** As soon as illegal movement is detected message or ring the nearest police station sent.
- ✓ **Alarm-** As soon as illegal movement is detected an alarm will sound for nearby residents.

Moreover, the machine can also be operated by blind people. If found disabled, the mechanism could improve the accessibility of ATM's for handicapped people.

Required Ingredients:

- R305 Fingerprint Sensor
- Arduino UNO
- GSM module
- Motion sensor
- Camera
- Stepper motor
- Buzzer alarm

VI. FUTURE SCOPE

- QR code recognition.
- Scanning of microchip on ATM card
- Release of chloroform through air conditioner during burglary whenever illegal moments are found.
- A person who cannot speak will find a webcam and transact using actions that the AI will respond to.

VII. CONCLUSION

The problem of ATM theft and misuse of ATM cards has become a big problem in the society. From the above conceptual model explains that the biometric ATM system is highly secure as it provides authentication. Multimode biometric can be applied to enhance the security level of ATM organization. This paper identifies high level model for modification of existing ATM system with biometric authentication, automatic door lock, AI based image updating process and currency scanning system.

REFERENCES

1. Multi authentication ATM theft prevention using ibeacon-Manoj V, Moulishankar R, Sasipriya S, Usha Devi E, Devika t.UG student knowledge institute of technology, Salem.
2. ”Biometric authentications to control ATM theft” Ahmad TasnimSiddiqui, College of computers and IT Taif University, Kingdom of Saudi Arabia.
3. ATM security using fingerprint authentication and OTP-Aruna r, Sudha v, ShruthiG,Usha Rani R, Sushma V ,Department of electronics and telecommunication Engineering, ShriSairam College of Engineering, Anekal, Bangalore.
4. A Biometric based ATM Security System using RFID & GSM Technology, India
5. Cardless ATM system-RushikeshKulkarni (B.E.) | MuzammilMadki (B.E.) | TejasMapari (B.E)

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-17

Education System Using Cloud Computing

Prof.Paurnima.N.Patil.
DOCS and IT,BAS P.O. Nahata College Bhusawal.
patilpaurnima291@gmail.com

ABSTRACT-*Cloud computing is an emerging topic in today's lifestyle publications and users. In cloud computing the educational sector is going beyond classrooms as an essential service. Higher education, distance learning, online education etc. use cloud computing services. Cloud computing offers a new platform for innovative teaching practice. In this paper, we have analyzed the various academic fields that use cloud computing as a service. We have found that the above mentioned academic areas are benefiting from the service of cloud computing. Some remedial action is required to use it properly as a service.*

Keywords-*Cloud Computing, SAAS, PAAS, IAAS, Education*

I. INTRODUCTION

The world is changing fast. Day by day people are getting more and more involved in ICT. Education is not outside the phenomenon. Students of most educational institutions are now relying on web based education system. Everything from lectures to assignment submission is now online based. There comes the necessity of cloud.The idea of computing in the "cloud" traces back to the origins of leveraged computing, a concept that computer scientist John McCarthy publicized in 1961: If the computers I've suggested become the computers of the future, computing might happen someday. Computer leverage can become the basis of a new and important industry.

Cloud computing is a buzzword in the ICT industry these days. It means distributing computing resources over the Internet. In a word, cloud computing is nothing but a group of highly scalable computers that work together somewhere and serve the end user over the Internet. It can also be visualized as one's computing task using hardware resources from another location. The term "cloud computing" was first coined by Compaq in 1996 and gained popularity when Amazon.com introduced Elastic Compute Cloud. Cloud users can take advantage of other organizations delivering services related to their software, data and other computing needs without running their usual computer software.

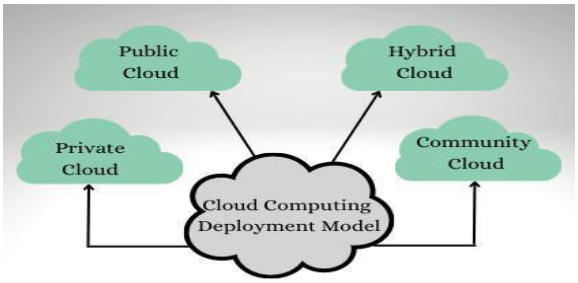
Cloud computing is a modern approach to managing business needs more efficiently. With the advancement of the internet and technology, this term has only evolved and become more popular. You can easily find cloud computing applications in many domains. This article explains how cloud computing has spread like a virus and how it affects educational technology. Before going deeper, it is important to understand what cloud computing means.

What is Cloud Computing in Education?

Cloud computing in education helps students, teachers and administrators alike. Cloud computing allows students to access homework anywhere they have an Internet connection, allows teachers to instantly upload learning materials, and allows administrators to easily collaborate with each other and save money on data storage.

II. CLOUD COMPUTING MODELS

In this age where internet is working like today a tool for users where data can be easily accessed, the cloud computing can be used to manipulate data, configure it access systems and applications online very easily. we so no need to worry about the storage in our devices everything is being stored in the cloud if we have it configuration and application required to handle it.



Deployment Model

Different types of access to the cloud can be defined by the deployment model, i.e., the location of the cloud. The cloud access can be any of four types: public, private, hybrids and communities.

- **Public Cloud**

Accessibility for systems in the public cloud and the service is relatively simple. It is less secure because the data is stored in the data center of the provider and is the provider Responsible for managing and maintaining data center, e.g, e-mail.

- **Private Cloud**

Accessibility in private cloud for systems and the service is in the organization. It is also known as a Internal or Enterprise Cloud. This introduces increased security because of its private nature.

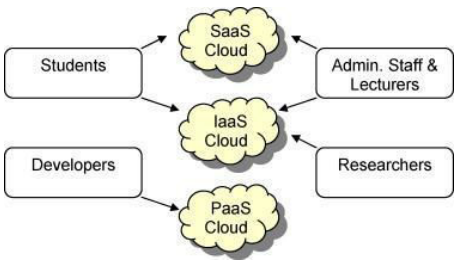
- **Community Cloud**

Accessibility to systems in the community cloud and services are by group of organizations.

- **Hybrid Cloud**

A hybrid cloud is mix of public and private clouds. Here, the critical activity performed using private cloud while non-critical activities are performed using public cloud.

III. SERVICE MODELS



Service models define three categories.These are following,

- **Software as a Service (SAAS):**

It represents the largest cloud market. The cloud market is still growing rapidly. SAAS uses the web deliver to distribute applications managed by third-parties vendor and whose interface is accessed from the client side. Most SAAS applications can be run directly from the web browser without any downloads or installations, although some require plugins. Today SAAS offers companies like Google, Salesforce, Microsoft, etc

- **Platform as a Service (PAAS):**

In this type, the level of the software or development environment is summarized and other higher levels at which services are provided a service can be created. They were used for applications, and other development, while providing cloud components to software. PAAS is a framework they can build on Develop or customize applications. PAAS makes the development, testing and deployment of applications, Simple, and economical. PAAS providers offer predefined a combination of OS and application servers, such as LAMP Platforms (Linux, Apache, MySql and PHP), restricted J2EE, Ruby etc. Google's App Engine, Force.com, etc. are some popular PaaS examples.

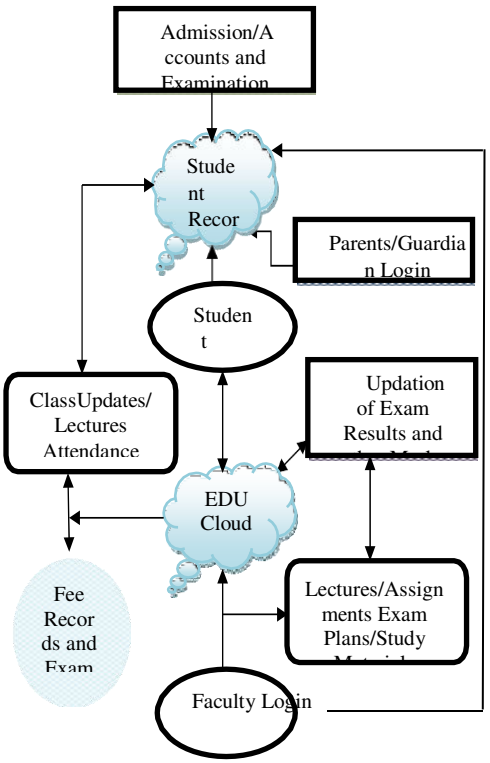
- **Infrastructure as a Service (IAAS):**

In this category, basic storage and computing capacity as a standardized service is provided over the network. They are self-service models accessing, monitoring and managing remote datacenters Infrastructure, such as compute (virtual or bare metal),storage, networking and networking services (eg. firewall). There some common examples are Amazon, GoGrid, 3 Tera, etc

Proposed Model For Education System Using Cloud Computing

This proposed model is for an education system using cloud computing. Cloud computing technology provides reliable resources that can support multiple platforms. Basically in this research paper we are focusing only on education. Major users of Education Cloud include faculty, students, administrative staff, admissions, accounts and examination departments. In this model, we also build a platform for parents so that they can view their children's attendance records, fee vouchers, upcoming quizzes, assignments, exam plans and especially results with the help of login ID and password provided by the cloud service. Providers generally students cannot share their results

and class schedule with their parents so now with the help of this model parents can easily check their children's records and required materials with the help of our proposed model. This model also provides facilities for professors to upload study materials, student attendance and other content to the cloud. Also they can update their class schedule if additional classes are required. In case of any delay due to internet or cloud unavailability, users can only access the last updated data. For the next update, they have to wait for system restore. Our proposed model provides an enabling environment for all stakeholders.



Education System Using Cloud Computing

The proposed model has different phases as shown in Fig.

- ✓ The flow of the model describes that the main stage is the admission/accounting and examination section where all the student data is available. This phase shows the enrollment of students in various programs, strength of students in each program, fee status and examination scheme for students.
- ✓ In the proposed model, we use two clouds namely Student Record Cloud and EDU Cloud which is Education Cloud. Student Record Cloud maintains student profiles, their courses, attendance records and study plans. Students can log in with the login ID and password provided by the cloud service providers so that they can connect to the Education Cloud to get their assignments, quizzes, results and attendance records as well as receive their lectures from the EDU Cloud.
- ✓ On the other side, we also provide facility to faculty so that they can login with their provided login id and password to update their course profile, lectures, study plan, upcoming quiz schedule, assignments, marks and discussion of extra classes. If needed, Instructors may also maintain attendance records of students in case any student does not meet the minimum attendance requirement as per institutional policies to appear for the examination.
- ✓ Another major step of this model is to facilitate the parents as well because most of the students are not able to share their study status and results with their parents. Parent/Guardian phase provides facility to check child profile like marks attendance, class schedule, exam plan and fee status. If parents are not able to check their child's profile, they can take parental help to check their child's status. We wish that this the proposed model helps our society to improve its education system.

Expected result for the Proposed Model

1. With the help of this model the teaching learning process becomes easier.
2. EDU Cloud will be more useful for checking updates and studying content.
3. Users will get their updated data on cloud.
4. Students will get their study material without any delay.
5. With the help of this model, communication between teachers and students will be facilitated.
6. Professors will upload their lectures, study material, lesson plans and maintain attendance records on the cloud.
7. With the help of this model parents/guardians of students will also check their study material, attendance record and upcoming activities.
8. Students will not be able to hide their marks or attendance records from their parents.
9. This model will play an important role in our education system.

Examples of the Application of Cloud Technologies in Education

In this section, we will take a look at some of the most popular cloud-based programs used in education.

1. Google Class

Google Classroom is a cloud-based learning management system that is part of Google Apps for Education. Google Classroom enables students to access the platform from computers, tablets and smartphones.

2. Blackboard

Blackboard provides education, mobile, communications and commerce software and related services to customers including education providers, corporations and government agencies. As of January 2014, its software and services are used by approximately 17,000 schools and institutions in 100 countries.

3. Knowledge Matters

Knowledge Matters is a leading cloud-based virtual business that provides online interactive, game-like business simulations that teach key business lessons to college and high school students

4. Microsoft Education Center

The Microsoft Education Center is designed to help students continue to learn no matter what. They facilitate online education and deliver the best educationdirectly to every student.

Reasons for using the cloud in an educational environment

Some of the advantages of using cloud in education system are mentioned below:

- Significant cost reduction.
- Access to application from anywhere.
- Support for training and learning.
- Exposure to new technologies to students.
- Use on personal devices.
- Exposure to advanced research.
- Ease of sharing learning content.
- Facilitates interaction.

IV. BENEFITS OF CLOUD TECHNOLOGIES IN MODERN EDUCATION

1. Personalized learning

Cloud computing manages open doors for more outstanding student preference of learning. Using internet-related gadgets, students can access a wide range of assets and software applications that suit their learning styles and concerns.

2. Reduced costs

A cloud based system provides cost reduction and faster use of upcoming advancements to meet educational needs.

3. Accessibility

From the user's point of view, utility availability is a key and important step in using the educational cloud. User demands 24/7 accessibility without any delay and latency issues so that information can be accessed from anywhere whenever required.

4. Facilities

This facility is easy to use, straightforward and convenient to operate. Cloud computing provides a user-friendly environment so communication becomes faster.

V. CLOUD COMPUTING LIMITATIONS

Cloud computing offers the possibility of improving efficiency, cost and accessibility for the education sector but it also has some limitations which are mentioned below:

1. Applications on the cloud

Sometimes applications don't run properly on the cloud due to downtime issues.

2. Risk of data protection

Data protection is a major concern when storing personal information or data on the cloud. The major associated risks are unauthorized access or loss of data by the service provider.

3. Institutional support

Usually such solutions are difficult to adopt because the management of any organization does not readily agree to put data on the cloud.

4. Distribution of politics, intellectual property

Data stored on the cloud is rapidly copied to third party interfaces and it is very difficult to monitor the data on the cloud so service providers should focus on disclaimer clauses to address this issue.

5. Protection and security of data

Cloud computing environments need more protection for data to give confidence to people who are reluctant to put their data on the cloud.

6. Development of solutions

Development of solutions to meet the needs is not readily available.

7. Lack of confidence

People lack confidence due to various issues involved in cloud computing.

VI. CONCLUSIONS AND RECOMMENDATIONS

The worldwide lockdown due to the coronavirus pandemic forced us to oversee the e-learning concept. Cloud computing is widely used in businesses, it is also essential in the education process. Cloud technology makes learning a simple and fun experience for participants on both sides of the learning process. Students, professors and teachers are now able to appreciate the ease and accessibility of cloud-based learning.

As we have seen that now a days cloud computing provides fast services through their infrastructure. Cloud computing services are very useful in various fields. It allows us to access our essential information when needed and facilitates sharing certain information with anyone. It also helps us create a convenient environment where you can use the resources available in the cloud computing environment in a very reliable manner. In this paper we present an educational model for teachers, students and parents. The proposed model provides a very user friendly environment for stakeholders. For future work, we will add more attributes and attributes to the hierarchy of this proposed model.

REFERENCES

1. Barhate S. M. and Narale S., "Cloud Based Teaching and Learning Environment for Smart Education," International Journal on Recent and Innovation Trends in Computing and Communication vol. 3, no. 2, pp. 38-41,2015.
2. Saini L., Joyti and Kaur H., "Role of Cloud Computing in Education System," International Journal of Advanced Research in Computer Science vol. 8, no. 4, pp. 345-347,2017.
3. Mathew S., "Implementation of Cloud Computing in Education - A Revolution," International Journal of Computer Theory and Engineering vol. 4, no. 3, pp. 473-475,2012.
4. Rao K. S. and Challa R. K., "Adoption of Cloud Computing in Education and Learning," International Journal of Advanced Research in Computer and Communication Engineering vol. 2, no. 10, pp. 4160-4163,2013.
5. Viswanath D. K., Kusuma S. and Gupta S. K., "Cloud Computing Issues and Benefits Modern Education," Global Journal of Computer Science and Technology vol. 12, no. 10, pp. 15-19, 2012
6. aspect, "The Benefits of Cloud Computing in Education are Huge," 2013. [Online], Available: https://blogs.aspect.com/the- benefits-of-cloud-computing-in-education-are-huge. [Accessed: 30-November-2018].

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023

Paper ID: CS-18

Exploring the Application of Data Analytics in Business Using Python

Mr. Dipak Y. Wani
M.Sc. II (Computer Science), Department of CS & IT, BASPONCCollege, Bhusawal, India
dipakwani2001@gmail.com

Mr. Harshal V. Patil
Research Scholar, Sandip University Nashik, India
harshal4patil@gmail.com

Dr. Vaibhav P. Sonaje
Department of Computer Sciences & Application, Sandip University Nashik, India

ABSTRACT – In recent times, the application of data analytics in business has come progressively popular. This paper introduces Data Analytics with Python and explains why the Python Programming Language is so effective for this field. This research paper explores the various data analytics methodologies that can be executed using Python. The veritably beginning processes of data analysis like cleaning, transforming, modelling of data is compactly explained in this paper and concentrate more on exploratory data analysis. We apply this approach to real- world datasets from colourful diligence, including finance, marketing, and health care, and estimate the performance of the models using colourful criteria. The paper examines various approaches used in data analytics, alike as descriptive, prophetic, and prescriptive analytics. Eventually, we identify coming research directions in data analytics using Python for business.

Keywords: - Data analysis, python, NumPy, pandas, seaborn, exploratory data analysis, business, PyCharm 3.9, Data Pre-processing.

I. INTRODUCTION

Data analytics is the process of analysing and interpreting data to make informed opinions. This research paper explores the use of Python for data analytics in business decision- timber. It is a popular concept, especially in the marketable sector as it allows associations to make data- driven decisions based on the result of Data Analysis. Data judges oversee interpreting data, analysing the results using statistical ways, and producing regular reports. also, to apply Data Analytics with Python, you do not have to learn everything about the programming language. Since you will not be doing the development work,understanding certain libraries and functions offered by Python is sufficient the paper will examine the crucial libraries used for data analytics in Python, similar as NumPy, Pandas, Matplotlib, and Scikit-Learn. The paper will also showcase the various operations of Python in data analytics, including data cleaning, data visualization, statistical analysis, and machine literacy.

The necessity to work in Data Analytics with Python is to have a platform where you can write your law and execute it. So, your first step is to set up a terrain that is`accessible to use and enables you to work in Python.

II. LITERATURE REVIEW

The literature review presents an overview of the different types of data analytics ways. Descriptiveanalytics involves analysing literal data to identify patterns and trends. Predictive analytics involveliteral data to make predictions about unborn events. conventional analytics involves recommending conduct grounded on data perceptivity. The literature also highlights the

challenges of data analytics, including data quality, data privacy, and lack of professed professionals. NumPy is used for numerical computing, Pandas for data manipulation, Matplotlib for data visualization, and Scikit- Learn for machine learning. The literature also highlights the significance of Python in data analytics due to its ease of use, effectiveness, and scalability.

Many Python libraries are available for doing the analysis. For example, NumPy, Pandas, Seaborn, Matplotlib, SK-learn, etc.

NumPy: This Python Library provides strong computation tools that can streamline your Mathematical and Statistical Operations when you are implementing Data Analytics with Python. It stores the data in the form of nd-arrays (n-dimensional arrays).

Pandas: Pandas is mainly used for converting data into tabular form and hence,it supports a Data Structure called Data Frame that is exceptionally good to store data in a tabular format makes the data more structured and easily to read.

Matplotlib: Matplotlib is a data visualization and graphical plotting package for Python and Matplotlib provides you total control over these graphs. You can modify the Colors, Shapes, Axis, Style, Thickness, Range, etc. of your visual plot.

Seaborn: Seaborn is a Python data visualization package based on matplotlib that is tightly connected with pandas' data structures. Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas' data structures. Seaborn helps you explore and understand your data.

Sklearn: Scikit-learn is the most useful library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering, and dimensionality reduction via a consistent interface in Python.

Fig:1.1 Python data libraries

III. DATA ANALYSIS USING PYTHON

A. Main Phases in Data Analysis: --

1) Data requirements:

Data is the most important unit in any study. Data must be handed as inputs to the analysis predicated on the analysis' necessities. It's capable of being done to identify and attain specific population variables (like as height, weight, age, and salary). It doesn't matter whether the data is numerical or categorical.

2) Data Collecting:

Data is gathered from a variety of sources, including relational databases, server databases, and different sources, based on the study needs. Field detectors, similar as marketplace cameras, satellites, covering systems, and so on, can also be operated as data sources.

3) Data processing:

Data that's collected must be reused or organized for analysis. For case, these may involve arranging data into rows and columns in a table format for farther analysis, frequently using spreadsheet or statistical software.

4) Data drawing:

It scans for data inconsistencies, duplicates, and crimes, and removes them. The data cleaning process includes tasks similar as record matching, relating data inaccuracy, data sort, outlier data identification, textual data spell checker, and data quality conservation. thus, it keeps us from having unanticipated products and assists us in delivering high- quality data.

5) Exploratory data analysis:

Once the datasets are cleaned and free of error, it can again be analysed. A variety of ways can be applied similar as exploratory data analysis- understanding the dispatches contained within the attained data and descriptive statistics- chancing normal, standard, etc. Data visualization is also a methodology used, in which the data is described in a pictorial format to gain fresh perceptivity, regarding the information within the data.

6) Modelling and algorithms:

Mathematical formulas or models (known as algorithms) may be applied to the data to identify connections among the variables; for sample, using correlation or occasion.

7) Data product:

A data product is a computer operation that takes data inputs and generates products, feeding them back into the atmosphere. It may be grounded on a model or algorithm.

B. Data Analysis Methods: --

a. ***Textual analysis:*** which can also relate as data booby-trapping it's to arrange the data into large data sets using mining tools. The main end of textual analysis is to collude the data into business data using business intelligence tools.

b. ***Descriptive Analysis*****:**It's to interpret, model and process the former collected data which can be done in
statistical analysis.

c. Diagnostic analysis: These approaches are to research the statistical analysis and find the cause for why it happens.

d. *Predictive analysis*: In this analysis we try to prognosticate what can be by using statistical data. For illustration, in day- to- day life how the person does save on his predictable earning income.

e. *Prescriptive Analysis*: This form analysis is used to unite all the former analysis reports to decide what decision could be taken grounded on current situation.

C. Why data analytics is essential for business?

1. Improved customer service:perfected client service data analytics can also help businesses to perfect their overall standard ofclient service. it provides in- depth perceptivity into what guests want and their preferences and againallowingyour whole clientservicecompany to pierce it can help to guaranty better consistence ofservice quality.

2. Better decision making: More decision-making data analytics allows businesses to edge their decision-making chops.Equipped with a more thorough understanding of their client base and their own performance, they canuse the perceptivity attained via data analytics to make advanced opinions as well as making designoperation more effective.

3. Enhanced efficiency:it's possible for businesses to streamline numerous of their processes,therefore, rendering them more effective while also enabling them to cut costs. it also helps them withfiscal analysis, enabling them to emplace their coffers more efficiently.

4.Competitive advantage:Businesses that use Business Analytics to inform their strategies and opinionsare more likely to gain a competitive advantage over those that don't.

5. Predictive modelling:Business Analytics can help businesses prognosticate unborn trends and issues,assisting in visionary decision- timber and threat guidance.

IV. METHODOLOGY

The research methodology involves a qualitative study of Python libraries and their applications in data analytics in Business. Data will be collected from various sources, including online libraries, repositories, and case studies. The research methodology involves a qualitative study using case studies to explore the application of data analytics in decision making. The data will be analysed using descriptive statistics and content analysis.

V. RUNNING BASIC INFERENTIAL ANALYSES

Python provides numerous libraries for conclusion and statistical analysis similar as Pandas, SciPy, and NumPy. Python is an effective tool for implementing numerous statistical data analysis operations similar as the following.

a) Correlation analysis: -

Correlation analysis is a statistical system used to determine the strength and direction of the relationship between two variables. The result of the analysis is known as the correlation Coefficient, the value between-1 and 1 that shows the strength and direction of the relationship.

A correlation Coefficient of-1 shows that a exactly negative correlation, one variable increases and other one decrements. 1 indicates a perfect positive correlation, one variable increases, the other increases. By understanding the correlation Coefficient and the direction of the correlation, businesses and experimenters can make informed opinions groundedon the data.

Correlation refers to some statistical relationship involving dependence between two data sets, similar as the following graph shows the correlation between the price of a product and its sales volume in PyCharm using python's library functions. Correlation refers to some statistical relationship involving dependence between two data sets, such as the following graph shows the correlation between the price of a product and its sales volume in PyCharm using python's library functions.

```
----
import pandas as pd
import matplotlib.pyplotas plt
import seaborn as sns

data = {
'Price': [10, 12, 20, 25, 30, 35, 40, 50, 60,80,
120, 150,200,220,250],
'Sales':[500, 850, 960 , 700, 650, 1000, 980,
1200, 1150, 1400, 1000, 950, 1500, 1400, 780]
}
df = pd.DataFrame(data)
correlation = df['Price'].corr(df['Sales'])
print('Correlation between price and sales:',
correlation)
sns.pairplot(df,kind="scatter")
plt.show()
```

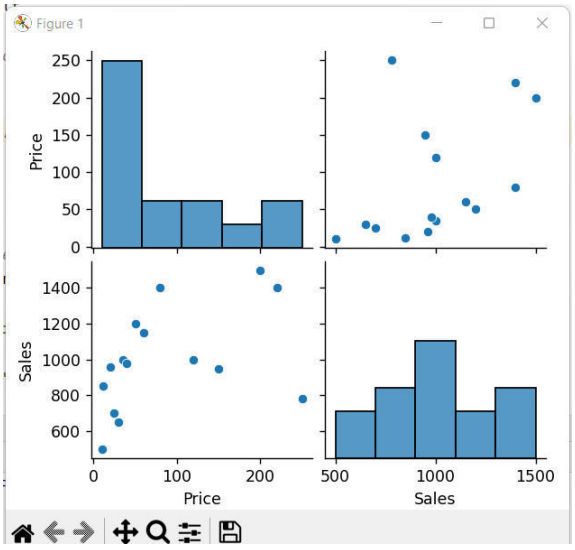


Fig:1.2 Correlation analysis graph

Here,correlation between Correlation between price and sales: 0.45145.

b) Regression analysis: -

Linear regression of two variables initiates a straight line when plotted as a graph, where the power of both variables is 1. A non-linear combination where the power of any variable is not equal to 1 makes a arc shape.

Regression analysis is a statistical system used to analyse the relationship between one dependent variable (also known as the response or outcome variable) and one or further independent variables(alsoknown as predictor or explicatory variables). The thing of regression analysis is to model the relationship between the dependent variable and the independent variables and use that model to make predictions about the dependent variable grounded on the values of the independent variables.

We can see an illustration with relationship between sales and benefit of product and then sales is the independent variable and Benefit is the dependent variable, below graph showing relationship among them using python code.

```
import seaborn as sb
import pandas as pd
from matplotlib import pyplotas plt

data = {
'sales': [140, 160, 200, 240, 250, 300, 320, 360,
400, 450],
'Benifit': [15,19,20,22,24,23,23,28,26,27]
}
df = pd.DataFrame(data)
sb.regplot(x = "sales", y = "Benifit", data = df)
plt.xlabel('sales')
plt.ylabel('Benifit')
plt.show()
```

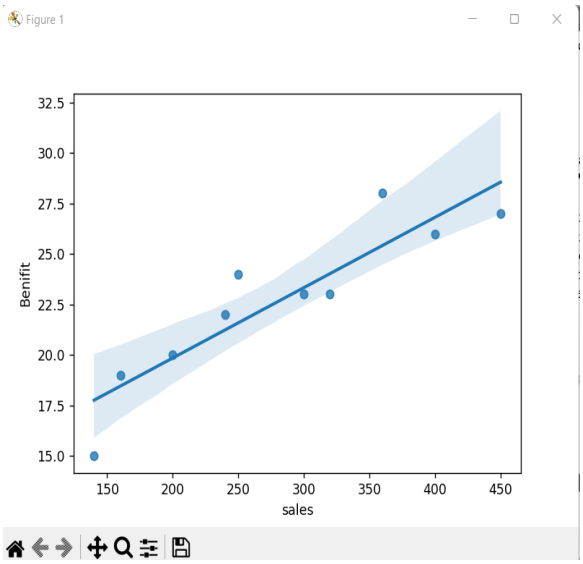


Fig:1.3 Regression Analysis Plotted graph

c) Data Frame: -

Data frame is the data structure which is the main structure for data collection and processing in Pythonprogramming. A Data Frame is a two-dimensional, size-mutable, and tabular data structure in Python's Pandas library and it's extensively used in data analysis. Different attributes can be used to recoup data from a series iloc() and loc() attributes.

Also, Python can automatically retrieve data grounded on the passed value. you can perform a variety of operations on a Data Frame, similar as opting columns, filtering rows, and performing calculations. Then is an illustration of data from which selecting column or row to accessing specific information,

---

```
import pandas as pd
# Create a dictionary with some data
data = {
'name': ['Dipak', 'Dhanshree', 'Tejas', 'Damini'],
'age': [22, 20, 27, 26],
'gender': ['M', 'F', 'M', 'F'],
'city': ['Bhusawal', 'Mumbai', 'Jalgaon', 'Pune']
}
df = pd.DataFrame(data)
print(df)
print(df.iloc[2])
print(df.loc[df['age'] >24])
print(df.iloc[:2])
```

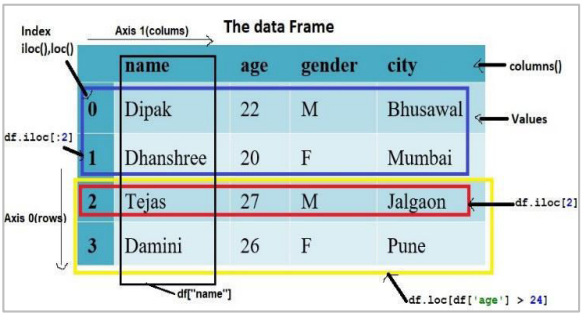


Fig:1.4 The data Frame

VI. FUTURE OF DATA ANALYTICS WITH PYTHON: -

Although it is earlier a well- substantiated programming language in world., Python is enduring to increase at the grade of an over- and- nearing one, growing 151 percent in operation since 2018. The future of data analytics in business using Python looks promising, as it provides businesses with the tools they need to dissect and make sense of their data, automate processes, and make data- driven opinions. As the significance of data- driven decision timber continues to grow, businesses that embrace Python for data analytics will be well deposited to stay ahead of the competition.

"Data analytics will growth at 16% to 18% over the next five years, while core administrator services

are questionable to see important growth due to automation. Python has a large and active community of innovators and users, who continuously develop substitute libraries and tools for data analysis and machine learning. Whether you want to come a data reviewer or make the big bound to data scientist, knowledge and gathering up Python is an arbitrary must**!**

VII. CONCLUSION

In this paper, varied phases of data analysis including data collection, cleaning and analysis are talked over summarily. Explorative data analysis is mainly studied too. For the execution, Python programming language is used. For elaborate examination, PyCharm is used. Different Python libraries and packages are introduced. Using various analysis and visualization styles, many results are wrested.

This article discussed the knowledge process of Data Analytics with Python. It explained the various way that you should follow to succeed in your attempts to apply Data Analytics using Python. also, it pressed the common misapprehensions that you must avoid while starting your knowledge process. All you need to do is follow the way handed in this composition and practice as much as possible.

Python has surfaced as one of the most popularized languages for data analytics in business. Its ease of use, important libraries, and open- source nature make it a glamorous choice for businesses looking to grow insight from their data. The conclusion will brief the judgments of the study and give recommendations for businesses and unions looking to administer Python for data analytics.

VIII. REFERENCES

1. McKinney, Wes. "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython." O'Reilly Media, Inc, 2017.
2. Krishna C V".”Mohana", conference paper,'A Review of Artificial Intelligence Methods for Data Science and Data Analytics', August 2018.
3. "Islam, Mohaiminul "Data Analysis: Types, Process, Methods, Techniques and Tools." International Journal on Data Science and Technology (2020) .
4. "KiranbalaNongthombam"," Deepika Sharma", IJERT, Data Analysis, Issue July 2021.
5. "Jinhua Chen", "Qin Jiang", "Study of data analysis model based on big data technology," *2016 IEEE International Conference on Big Data Analysis (ICBDA)*, 2016.
6. Banerjee, Anindya et al. "Business Analytics Using Python: A Practical Approach." CRC Press, 2020.
7. https://www.researchgate.net/publication/327364915_Data_Science_and_Analytics_with_Python.
8. https://en.wikipedia.org/wiki/Business_analytics
9. https://towardsdatascience.com/exploratory-data-analysis-in-python-a-step-by-step-process-d0dfa6bf94ee
10. https://en.wikipedia.org/wiki/Data_analysis

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023

Paper ID: CS-19

A Study On Robotics

Assist. Prof. Pari U.Talreja
Department of computer science & IT
BAS P.O.Nahata Commerce College, Bhusawal .
talrejapari4@gmail.com

ABSTRACT:

This Research paper comprises of detailed data about the robot's method and system. As one and all know, how artificial intelligence is rising in the open market and the market is getting totally dependent on artificial intelligence for responsibility the multifaceted tasks. Robotics is greatest famous division in the area of industrial knowledge where all engineer is taking deep attention to make a robot which could doe certain task and can give suitable results for the given task. Every engineer is trying to mark a robot through 0% mistake which is quite impossible as the technology is increasing. We can think about it but still 0% doesn't unkind that it will not have any point of error but it means it will give you the correct answer for every question without any doubt. It Show its uses and detailed data how it workings and how it senses working Allis showed in this paper which will be enough for getting and good information about robotics and devices along with the structure of robots. Nervousness is raising that robots and artificial intelligence will substitute many businesses. To remain relevant in this changing occupation landscape, the employee of the future is expectable to be advanced; able to commercial events transform businesses and provide original solutions to come across global tests. To develop such capabilities, work integrated learning (WIL) has performed as an important approach. The resolution of this study is to examine the key aspects powerful creation among WIL scholars. Unlike previous educations that have been mainly qualitative or based on one lone snapshot this computable, longitudinal less tidy actions student capabilities before and after contribution in a WIL assignment at a business. It then assumes positive issue examination to assist and pre- and post-placement abilities.

Keywords: Novelty, Industry4.0, Robots, Artificial intelligence, Employability, Work combined learning.

I. INTRODUCTION

Although everyone knows what a robot is, it is difficult to give an exact definition. "A machine capable of mechanically execution a tough series of jobs, particularly one that can be programmed by a computer."

This definition includes some exciting components:

" Execute actions automatically. "; It is a key component in robotics, but also in many other simple machines called automata. The difference between a robot and a simple automaton like a dishwasher is what a "complex series of actions" is. Is laundry made up of a difficult series of actions or not? Is flying a plane on autopilot a difficult activity? Is baking bread complicated?

For all these tasks there are machines bordering on automata and robots.

"Programmable by a computer" is another key component of a robot, because some automata are planned mechanically and are not very elastic. On the other hand computers are originate universally, so it is hard to use this condition to differentiate a robot from another machine. A critical component of robots that is not stated explicitly in the definition is the use

of devices. Most automata do not have devices and cannot adapt their actions to their environment. Devices are what enable a robot to carry out without difficult tasks. Robot is a human thing which is capable of doing all the work the human can perform in a much less time than a human can take the place of a human but it can help humans for operating much of its task in daily life. Robots are also applications of artificial intelligence and sensors which combine together to form a human machine called robots. There are numerous applications of robots in the world of science and computer application. Scientists and engineers are working on robots to make it almost applicable in every field. It can be semi automatic or fully automatic that is there are many robots which are like human that is they can talk, they can walk without the guidance of a human through programmable language input into them at the time of manufacturing it but there are also semi-automated that is the needle remote for the controllability of its functioning. Robotics is one and only greatest apted and interesting branches in the arena of science and education which is loved by every youth and everyone wants to learn robotics for future use. There are Number of uses in the future where people will be depending on fully automated drama full complex stars as glowing as for everyday works as well as it will decrease manpower in the world because one robot is proficient of doing work of 10 persons. There is world-wide gratitude for the need for invention to transform economies (Atwood et al.2016; Castaño, Méndez, & Galindo, 2016; Jenson, Leith, Doyle, West, & Miles, 2016a; Jenson, Leith, Doyle, West, & Miles, 2016b; Xie & Wang, 2020). With, the advent to industry 4.0 or the fourth manufacturing rebellion characterized by cyber-physical schemes, there is a focus on the advanced application of progressive robotics and AI to bring about digital revolution in productions (Haenein " Emerging Advancement and Challenges in Science, Technology and Management " 23rd & 24th April, 2021 258 Consequently, the research query of this education is 'What are the important factors waying the development of revolution in students from side to side WIL?' It will measure revolution and possible drivers beforehand the after the WIL settlement. Findings of the education are significant in

(1) offering response on career literateness to students on the expansion of
innovation;

(2) attractive WIL program expansion by detection areas of skill insufficiency which can then be castoff to inform remedial action in succeeding WIL groundwork programs and offer additional provision to students;

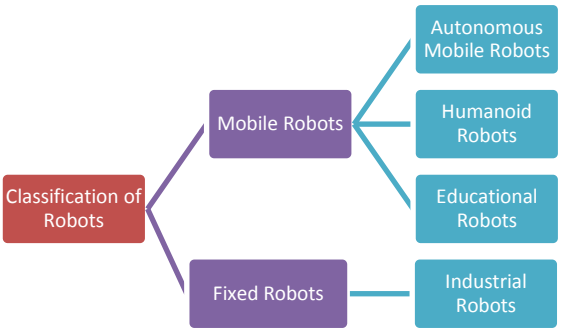
(3) notifying industry appointment labours to WIL hosts over evidence- based communiqué on the competences and welfares of WIL students There is considerable rhetoric and anecdotes on the services needed for the works of the upcoming, but there is little experiential evidence authenticating what are the key factors or causes of innovation. The Foundation of Young Australians formed a report title as 'The new basics: Big data reveals the skills young people need for the New Work Order' where they recommended services that Australian youths essential for jobs of the upcoming, including problematic solving, serious thinking,communication and co-ordination (FYA, 2016). Though, these skills are until now to be empirically tested for their effect on origination. Therefore, the bearing of these skills on modernisation will be observed in this education and discussed more in this segment.

II. CLASSIFICATION OF ROBOTS

Robots can be classified according to the environment in which they work The most common change is between fixed and mobile robots.

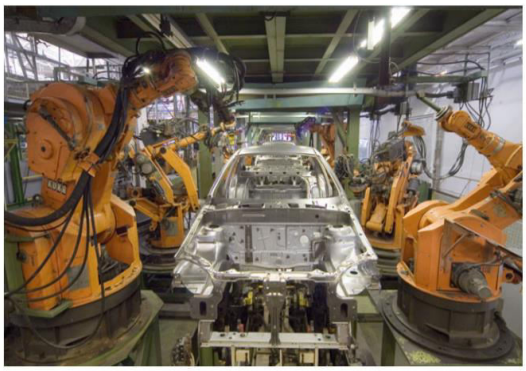
These two types of robots have very changed working environments and therefore require very changed abilities. Fixed robots are mostly industrial robotic manipulators that work in well defined environments adjusted for robots. Industrial robots execute specific repetitive jobs such painting parts in car industrial plants. With the improvement of devices and devices for human-robot interaction, robotic manipulators are progressively used in less controlled environment such as high-precision surgery. Mobile robots are expected to travel around and execute tasks in large, ill-defined and uncertain environments that are not designed specifically for robots. They want to deal with situations that are not precisely known in progress and that change overtime .Such environments can include changeable entities like humans and animals. Instances of mobile robots are robotic vacuum cleaners and self-driving cars. There is no clear dividing line between the tasks carried out by fixed robots and mobile robots—humans may interact with industrial robots and mobile robots can be controlled to move on tracks—but it is suitable to consider the two classes as fundamentally different. In particular, fixed robots are attached to a stable amount on the ground, so they can compute their position based on their internal state while mobile robots need to rely on their observation of the

environment in order to calculate their location. There are three main environments for mobile robots that need significantly altered design principles because they differ in the tool of motion: aquatic (underwater exploration), terrestrial (cars) and aerial (drones). Again, the classification is not harsh, for example, there are amphibious robots that move in both water and on the ground. Robots for these three environments can be more divided into subclasses: global robots can have legs or rolls or ways, and aerial robots can be lighter than air balloons or heavier than aircraft, which are in turn divided into fixed wing and rotary wing (helicopters).



Industrial robots

The first robots were industrial robots that replaced human workers performing simple repetitive tasks. Factory assembly lines can operate without the presence of humans, in well-defined environments where robots have to perform tasks in a specific sequence, working on precisely placed objects in front of it. One could argue that these are truly automata and not robots. However, today's automata often trust on devices to the extent that they can be considered as robots. Yet, their design is simplified since they work in a modified environment which humans are not allowed to access while the robot is working. However, today's robots need more flexibility, for example, the ability to operate objects in different locations or to recognize different objects that need to be wrapped in the right order .The robot can be required to carrying goods to and from warehouses. This brings extra autonomy, but the basic characteristic remains: the environment is more-or-less forced and can be modified to the robot. Added flexibility is required when industrial robots interconnect with humans and this introduces solid safety desires, both for robotic arms and for mobile robots. In specific, the speed of the robot must be reduced and the automated design must ensure that moving parts are not a danger to the user. The benefit of humans working with robots is that each can achieve what they do best: the robots perform boring or risky jobs, while humans perform more difficult steps and define the overall tasks of the robot, since they are quick to identify errors and chances for optimization.



Autonomous Mobile Robots

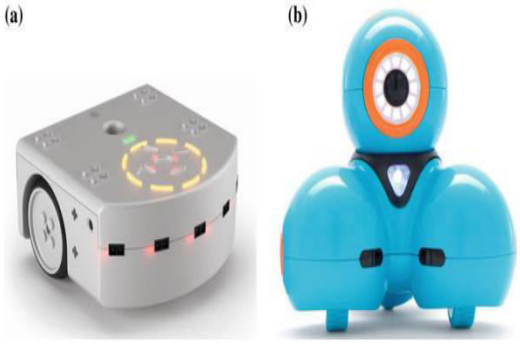
Many mobile robots are remotely controlled, acting tasks such as tube assessment, photography and mine disposal that trust on an operator controlling the device. These robots are not autonomous; they use their devices to give their operator remote access to dangerous, distant or places. Certain of them can be semi-autonomous, executing subtasks automatically. The autopilot of a high stabilizes the journey while the human chooses the flight path. A robot in a tube can control its movement inside the tube while the human examines for defects that need to be renovated. Fully autonomous mobile robots do not trust on an operator, but instead they make decisions on their own and perform tasks, such as carrying material while navigating in uncertain train (walls and doors within buildings, intersections on streets) and in a continuously changing environment (people walking around, cars moving on the streets). The first mobile robots were planned for simple environments, for example robots that cleaned swimming pools or robotic lawn mowers. Currently, robotic vacuum cleaners are broadly available, because it has verified reasonable to build reasonably priced robots that can navigate an internal environment cluttered with difficulties. Many autonomous mobile robots are planned to support professionals working in designed environments such as warehouses. An interesting example is a robot for clearing fields. This environment is partly designed, but advanced sensing is required to perform the tasks of identifying and removing unwanted plant. Even in very structured plants, robot share the environment with humans and therefore their sensing must be

extremely reliable. Possibly the autonomous mobile robot getting the most publicity these days is the self-driving car. These are very difficult to develop because of the highly complex uncertain environment of mechanical traffic and the strict safety requirements. An even more difficult and unsafe environment is space. Much of the research and development in robotics today is focused on making robots more autonomous by successful devices and enabling more intelligent control of the robot. Better devices can observe the details of more complex circumstances, but to handle these circumstances, control of the performance of the robot must be very flexible and adjustable. Visualization is a very active field of investigation because cameras are inexpensive and the information they can obtain is very rich. Efforts are being made to make systems more flexible, so that they can learn from a human to new situations. Another active field of research addresses the communication between humans and robots. This involves both detecting and intelligence, but it must also take into account the psychology and sociology of the interactions.

Autonomous mobile robot weeding a field

Humanoid Robots

Humanoid robots are a form of autonomous mobile robot with an particularly complex mechanical design for moving the arms and for locomotion by the legs .Humanoid robots are used for investigation into the mechanics of walking and into human machine communication. Humanoid robots have been suggested for performing services and maintenance in a house or a space position. They are being considered for providing care to the elderly who might feel concerned in the presence of a machine that did not appear human. On the other hand, robots that look very similar to humans can generate revulsion, a miracle referred to as the creepy valley. Humanoid robots can be very hard to design and control. They are costly to build with multiple joints that can transfer in many different ways. Robots that use tracks are desired for most applications because they are simpler, less expensive and robust.



Educational Robots

Educational robots are used widely in schools, both in classrooms and in additional activities. The large number of educational robots makes it difficult to give a whole impression. Here we give few examples that are classic of robots generally used in education. Pre-Assembled Mobile Robots Many educational robots are scheduled as pre-assembled mobile robots. Fig a shows the Thymio robot from Mobsy and fig b shows the Dash robot from Wonder Studio. These robots are relatively cheap, robust and contain a large number of devices and output components such as lights. An vital benefit of these robots is that you can implement robotic processes "out of the box," without investing hours in mechanical design and manufacture. However, pre-assembled robots cannot be modified, though many do support building additions using, for example, LEGO® components.

Robotics Kits

Robotics Kits consists of standard LEGO® blocks and other building components, together with motors and devices, and a programmable block which contains the computer that controls the components of the robot. The advantage of robotics kits is that they are flexible: you can design and build a robot to perform a specific task, limited onl yby your mind's eye. A robotics set can also be used to teach students powere ddesign. The disadvantages of robotics sets are that they are more expensive than simple pre-assembled robots and that review of robotics algorithms be contingent on the one's skill to positively implement a healthy powered design. A recent trend is to change fixed collections of blocks by parts created by 3Dprinters. An example is the Poppy Ergo Jr robotic arm The use of 3Dprinted parts allows more flexibility in the

creation of the mechanical building and greater roughness, but does require access to a 3D printer.

Robotic Arms

To act on its environment, the robot wants an actuator which is a module of a robot that interrupts the environment. Many robots, in particular robotic arms used in industry, affect the environment through end effectors, The actuators of mobile robots are the motors that cause the robot to change, as well as elements such as the vacuum pump of a vacuum cleaner.
Educational robots are commonly mobile robots whose only actuators are its motors and show devices such as lights, sounds or a screen. End effectors can be constructed with robotics kits or by using extra components with pre-assembled robots, although educational robotic arms do arise Management of objects presents difficulty into the design; however, since the algorithms for end effectors are alike to the algorithms for simple mobile robots, most of the circumstances in the book will accept only that your robot has engines and show procedures.

III. ADVANTAGES

Cost Efficiency

They are very cost effective as they do not take gaps in between as the human body needs a pause while running. So this thing makes it cost effective and it can do the same work continually once a cycle is set in it. There is no danger of RSI. It also decreases the cost of manufacturing with the increase in the amount of construction. The cost that one investment in buying the robot will be simply in a very small period.

Improved Quality Assurance

There are very rare people who like to do their jobs for a certain time and with full attention but after that they drop their interest or attention and start doing it just for money but this is not for robots. There is small risk getting bored or not focused because it is made for doing the work and give the advanced standard of products that are hard to be found by the human race when people are matching their jobs with their currency not with their interest.

Increased Output Robots

Increase the output rate of an industry as humans can do 24/7 work they have aassured time duration but robots can do work without taking pauses and leaves. Single robot can do work of 10 people and it can be used in aindustrial unit for different output easily. You need to focus on the staff for their work but the problem of yours is also not job when a robot is working in your manufacturing industry.

Work in Dangerous

Environments Everyone can't work at a place with the environment but robots can do work in any place without caring about this surrounding. Its production rate is very high. It can work. I do know very high temperatures on a low temperature where people are hard to do work. It gives output for the work and there is no risk with the robot as like with humans. It's also a main advantage of robots.

IV. DISADVANTAGES:

Potential job losses

The largest disadvantage of robots is that good potential people are getting unemployed because robots can do work of a 10 person in a single use so mostly everyone wants to save them money so they buy the robot in its place of paying 10 potential people for their work. Displaythis isa major disadvantage to the human where the unemployment it is more than unemployment and now due to the innovation of robots more peoples are getting jobless day by day.

Initial investment cost

The initial investment is very extraordinary when you are going to buy a robot for your work. Although the cost of the investment is returned in a few months but till one needs to pay much before purchasing it.

Hiring skilled staff

When you have a robot which is not completely automatic then you need to hire skilled staff for doing task of the robots it become very hard to be paid visitor take high salary and placing there salary in your work becomes quite hard so it's superior clue to have a fully automatic robot or pay humans for manpower.
All active advantages and disadvantages are the basics one and the most important one but there are many other disadvantages and advantages for the same.

V. CONCLUSION

This was sufficient detail about robot devices and systems. As the world is getting improved into machinery oriented with robot other top most in demand. All engineers in many corporations work day and night to make robots as fast as possible. High demand and high charge give rise to an economy very fast. So we should keep searching on robots and its other devices which can give us help in making the world full of Technology where manpower is less. This surveys classifies robots. It also requires the general robot and formalisms used to present algorithms

Sensors Robots are more than remotely controlled machines like a television set. They show autonomous behavior based on discovering objects in their environment using devices. This gives an overview of the devices used by robots and describes the concepts of series, resolution and accuracy. It also discusses the nonlinearity of devices and how to deal with it.

When an autonomous robot notices an object in its environment, it responds by changing its behavior. This introduces robotics algorithms where the robot openly changes its behavior based upon input from its devices

We have seen that robots can do every work of humans and it's replacing human power in every field and every side so we want to become trained to that level so that no one can swap you with robots. A robot is a man-made thing and it can't take the place of humans in any aspects. Just keep increasing your skills so a man-made thing can't replace you with your work. Robot manufacturing can't be close so the thing which everyone can do is increase the skill power and get more faster which can help you to remain at your situation without replacement by robot or any other device.

REFERENCES

1. Deen, M.J., &Basu, P.K. (2012). Silicon Photonics: Fundamentals and Devices.

2. Hamblen, J.O., &Bekkum, G.M. (2013). An Embedded Systems Laboratory to Support Rapid Prototyping of Robotics and the Internet of Things. IEEE Transactions on Education, 56, 121-128.

3. Kumar, D.: Learning Computing with Robots. Lulu (2011). Download from http://calicoproject.

4. Chiolerio, A., &Quadrelli, M. (2017). Smart Fluid Systems: The Advent of Autonomous Liquid Robotics. Advanced Science, 4.

5. Krishnan, R.H., &Pugazhenthi, S. (2014). Mobility assistive devices and self-transfer robotic systems for elderly, a review. Intelligent Service Robotics, 7, 37-49.

6. Rahul Reddy Nadikattu, 2014. Content analysis of American & Indian Comics on Instagram using Machine learning", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.2, Issue 3, pp.86-10

7.Aguilar, J., Zhang, T., Qian, F., Kingsbury, M., McInroe, B., Mazouchova, N., Li, C., Maladen, R.D., Gong, C., Travers, M.J., Hatton, R.L., Choset, H., Umbanhowar, P., & Goldman, D.I. (2016). A review on locomotion robophysics: the study of movement at the intersection of robotics, soft matter and dynamical systems. Reports on progress in physics. Physical Society, 79 11, 110001 .

8. Shin, J., Siegwart, R., Magnenat, S.: Visual programming language for Thymio II robot. In:
Proc. of the 2014 Conference on Interaction Design and Children (IDC) (2014)

9. Siegwart, R., Nourbakhsh, I.R., Scaramuzza, D.: Introduction to Autonomous Mobile Robots,
2nd edn. MIT Press, USA (2011)

10. Trobaugh, J.J., Lowe, M.: Winning LEGO M INDSTORMS Programming. Apress (2012)

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-20

"Trends and Challenges in the Indian Stock Market: A Review"

Kishor Gorakh Desale
S.S.V.P.S`s .L.K.Dr.P.R.Ghogrey Science College, Dhule
desalekishor.2008@gmail.com

Dr.Bhagyashri Dinesh Patil
S.S.V.P.S`s B.N.S Patil Art`s & M.F.M.A.Commerce College, Dhule
bhagyashripatil759@gmail.com

ABSTRACT - *The Indian stock market is a vibrant, quickly changing financial environment that is essential to the expansion and advancement of the national economy. This review article provides an overview of the Indian stock market, highlighting its key features, historical evolution, and current state. The article then provides an overview of the challenges and opportunities faced by the Indian stock market, including issues related to market volatility, regulatory reforms, technology disruptions, and investor education. It also discusses the potential of the Indian stock market to grow further and become more resilient in the face of evolving market dynamics and global economic changes. In conclusion, this review article provides a comprehensive overview of the Indian stock market, its historical evolution, current state, and future prospects. The review article is provides some sort of knowledge about Indian stock market for new researchers.*

Keywords*: Indian Stock Market, SEBI, NSE, BSE.*

Introduction:

The Indian stock market is a crucial part of the country's financial system and plays a significant role in driving economic growth and development. It serves as a platform for companies to raise capital, investors to participate in wealth creation, and contributes to overall market liquidity and stability. The Indian stock market has a rich history that dates back to the 19th century, and it has evolved over time through various reforms and regulatory changes.

The origin of the Indian stock market can be traced to the establishment of the Bombay Stock Exchange (BSE) in Mumbai in 1875, making it one of the oldest stock exchanges in Asia. Over the years, additional stock exchanges such as the National Stock Exchange (NSE), Kolkata Stock Exchange (KSE), and others have been established, creating a robust ecosystem for stock trading in India [1].

The Indian stock market has undergone significant changes since its inception. One of the pivotal moments in the history of the Indian stock market was the economic liberalization and financial reforms initiated in the early 1990s. These reforms aimed to open up the Indian economy, promote foreign investment, and modernize the financial sector, including the stock market. The Securities and Exchange Board of India (SEBI) was established in 1992 as the primary regulatory authority for the Indian securities market, with the

mandate to regulate and develop the Indian stock market[2].

Since then, the Indian stock market has witnessed significant growth, both in terms of market capitalization and trading volumes. It has also seen the introduction of various segments, including derivatives, commodities, and currency, which have added depth and diversity to the market [3].

The Indian stock market has also been influenced by technological advancements, with electronic trading and settlement systems becoming prevalent, and increased adoption of mobile trading and online platforms[4].

However, the Indian stock market also faces challenges, including market volatility, regulatory reforms, technological disruptions, and the need for investor education. Ensuring market stability, integrity, and investor protection remain important priorities for the Indian stock market's sustained growth and development. The Indian stock market's future prospects are influenced by factors such as economic growth, policy reforms, global market trends, and technological advancements.

Recent Trends:

Recent trends in the Indian stock market are dynamic and can change over time. As of the knowledge cutoff date of September 2021, some of the recent trends in the Indian stock market include:

Retail Participation: There has been a significant surge in retail investor participation in the Indian stock market. Individual investors, including small traders and first-time investors, have been increasingly participating in stock trading and investment, driven by factors such as easy accessibility to online trading platforms, increased awareness, and availability of low-cost investment options[5].

Digitalization: The growth of digital platforms for stock trading and investment has gained momentum in recent years. Online trading platforms and mobile applications have become popular among investors, providing them with convenient and user-friendly access to the stock market. This trend has been accelerated by the COVID-19 pandemic, which has further emphasized the need for digitalization in the financial markets[6].

Initial Public Offerings (IPOs): The Indian stock market has witnessed a surge in IPO activity in recent years. Many companies, including startups and established firms, have gone public to raise capital, leading to a significant increase in IPO activity. This trend has been driven by favorable market conditions, regulatory reforms, and increased investor appetite for IPOs [7].

Foreign Investor Participation: The Indian stock market has been attracting significant foreign investor participation in recent years. Foreign institutional investors (FIIs) and foreign portfolio investors (FPIs) have been investing in Indian equities, driven by factors such as positive economic growth prospects, favorable policy reforms, and liquidity infusion by global central banks. Foreign investors' participation has a significant impact on the Indian stock market's liquidity and overall performance[8].

Challenges:

The Indian stock market faces several challenges that impact its functioning and performance. Some of the key challenges in the Indian stock market include:

Market Volatility: The Indian stock market is characterized by frequent fluctuations in stock prices, which can lead to high levels of market volatility. Factors such as global economic conditions, geopolitical events, and domestic economic and political developments can impact stock prices in the Indian market, resulting in increased volatility[9].

Liquidity Concerns: The Indian stock market has faced liquidity concerns, particularly in the case of mid and small-cap stocks. Limited liquidity in certain segments of the market can impact price discovery, increase trading costs, and pose challenges for investors seeking to buy or sell securities[10].

Regulatory Compliance: The Indian stock market is governed by strict regulatory frameworks, including regulations by the Securities and Exchange Board of India (SEBI). Compliance with these regulations, such as disclosure requirements, corporate governance norms, and insider trading regulations, can be challenging for market participants, particularly smaller companies[11].

Cybersecurity Risks: The increasing reliance on technology and digital platforms in the Indian stock market has heightened the risks of cyber threats, including hacking, data breaches, and fraud. Cybersecurity challenges can pose threats to market integrity, investor confidence, and operational continuity of stock market entities[12].

Corporate Governance Issues: Corporate governance has been a concern in the Indian stock market, with instances of corporate fraud, accounting irregularities, and related-party transactions. Weak corporate governance practices can impact investor confidence, result in market manipulation, and erode market integrity [13].

Opportunities:

Opportunities in the Indian stock market are diverse and can offer potential benefits for investors and market participants. Some opportunities that can be explored for a review article on the Indian stock market include following things:

Economic Growth and Demographic Dividend: India is one of the fastest-growing major economies in the world, with a large and young population. The country's economic growth potential and demographic dividend can create opportunities for investments in various sectors such as financial services, infrastructure, consumer goods, and technology, among others[14].

Increasing Domestic Savings and Investments: India has witnessed a rise in domestic savings and investments, with a growing middle class and increasing financial inclusion. This trend can result in higher inflows of domestic investments into the stock market, creating opportunities for investors to participate in the growth of the Indian economy[15].

Diversification and Portfolio Allocation: The Indian stock market offers a wide range of investment options, including large-cap, mid-cap, small-cap, and sector-specific stocks. This allows investors to diversify their portfolios and allocate investments based on their risk tolerance, investment objectives, and market outlook, providing opportunities for optimal portfolio allocation strategies[16].

Innovation and Technological Advancements: The Indian stock market has witnessed technological advancements in recent years, with the adoption of online trading platforms, electronic clearing and settlement systems, and other innovative solutions. This has improved the efficiency, transparency, and accessibility of the stock market, creating opportunities for investors and market participants to leverage technology for better trading and investment outcomes[17].

Increasing Institutional Participation: Institutional participation in the Indian stock market, including mutual funds, insurance companies, pension funds, and foreign institutional investors, has been on the rise. This trend can result in higher liquidity, better price discovery, and increased stability in the stock market, offering opportunities for investors to participate in a more mature and institutionalized market[18].

IPO Market: The Indian stock market has witnessed a surge in Initial Public Offerings (IPOs) in recent years, providing opportunities for investors to participate in the growth potential of newly listed companies. This can offer potential gains for investors who can identify and invest in promising IPOs[19].

Conclusion:-

The Indian stock market has an extensive background, and via several reforms and regulatory changes, it has developed through time. It provides a platform for businesses to obtain funds and for investors to take part in wealth creation, and it is crucial for India's economic growth and development. It does, however, also encounter difficulties that call for continued consideration and study. For investors, policymakers, and other stakeholders to make educated decisions and support the expansion and stability of India's financial ecosystem, it is imperative that they comprehend the dynamics of the Indian stock market.

References

1. Hegde, S., Narayan, P. K., & Prabheesh, K. P. (2017). A comprehensive analysis of Indian stock market using extreme value theory. The North American Journal of Economics and Finance, 42, 436-459.
2. Securities and Exchange Board of India (SEBI). (n.d.). About Us. Retrieved from https://www.sebi.gov.in/sebiweb/about_us.jsp.
3. Vijayakumar, N. (2017). Indian stock market: An empirical analysis of efficiency and volatility. IUP Journal of Applied Finance, 23(4), 54-69.
4. Bhattacharya, D., Bhowmik, S., & Saha, S. (2020). Impact of mobile trading on Indian stock market: An empirical analysis. International Journal of Economics, Commerce and Management, 8(8), 9-17.
5. Business Today. (2021). India's retail investors are here to stay. Retrieved from https://www.businesstoday.in/markets/story/indias-retail-investors-are-here-to-stay-305832-2021-01-15
6. Business Standard. (2020). Covid-19 impact: Digital trading platforms see increased adoption in India. Retrieved from https://www.business-standard.com/article/markets/covid-19-impact-

digital-trading-platforms-see-increased-adoption-in-india-120071300130_1.html
7. The Economic Times. (2021). India's IPO boom: What's driving the frenzy. Retrieved from https://economictimes.indiatimes.com/markets/ipos/fpos/indias-ipo-boom-whats-driving-the-frenzy/articleshow/83427802.cms
8. The Times of India. (2021). FIIs invest Rs 14,202 crore in Indian equities in July. Retrieved from https://timesofindia.indiatimes.com/business/india-business/fiis-invest-rs-14202-crore-in-indian-equities-in-july/articleshow/84769732.cms
9. Financial Express. (2020). Market volatility: Why Indian stock market is much more volatile thaother markets. Retrieved from https://www.financialexpress.com/market/market-volatility-why-indian-stock-market-is-much-more-volatile-than-other-markets/2008802.
10. Livemint. (2021). Liquidity squeeze hits mid, small caps. Retrieved fromhttps://www.livemint.com/market/stock-market-news/liquidity-squeeze-hits-mid-small-caps-11615709878686.html.
15.Reserve Bank of India. (2020). Handbook of Statistics on Indian Economy. Retrieved fromhttps://www.rbi.org.in/Scripts/PublicationsView.aspx?id=20041.
16. National Stock Exchange of India Ltd. (2021). NSE Indices. Retrieved from https://www.nseindia.com/indices.
17. National Stock Exchange of India Ltd. (2021). Technology & Innovation. Retrieved from https://www.nseindia.com/about-us/technology-innovation.
11. The Economic Times. (2021). Market experts raise concerns over regulatory compliance in Indian stock market. Retrieved from https://economictimes.indiatimes.com/markets/stocks/news/market-experts-raise-concerns-over-regulatory-compliance-in-indian-stock-market / articleshow/80925377.cms
12. Business Today. (2021). Indian stock market at risk of cyber attacks; Here's what you need to know. Retrieved from https://www.businesstoday.in/markets/stocks/story/indian-stock-market-at-risk-of-cyber-attacks-heres-what-you-need-to-know-300141-2021-01-06 13. The Times of India. (2021). India Inc's corporate governance issues cast shadow on market.Retrieved from https://timesofindia.indiatimes.com/business/india-business/india-incs-corporate-governance-issues-cast-shadow-on-market/articleshow/84335888.cms
14.World Bank. (2021). India. Retrieved from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=IN.
18. Securities and Exchange Board of India. (2020). Annual Report 2019-20. Retrieved from https://www.sebi.gov.in/sebi_data/attachdocs/dec-2020/1608771250924.pdf
19. Economic Times. (2021). IPO market: India witnessed IPO boom in 2020. Retrieved from https://economictimes.indiatimes.com/markets/ipo/ipo-market-india-witnessed-ipo-boom-in-2020/articleshow/81224994.cms.

IJCRT
ISSN 2454-7719
International Journal of Computer Research & Technology
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-21

Performance Evaluation Of Ad Hoc On-Demand Distance Vector (Aodv) Protocol In The Vanet For Urban Scenarios

Ms. Poonam M. Mahajan,
Department of CS and IT, Bhusawal Arts, Science and P. O. Nahata Commerce College, Bhusawal, India
pmm@basponccollege.org

Dr. Gouri M. Patil
Department of CS and IT, Bhusawal Arts, Science and P. O. Nahata Commerce College, Bhusawal, India
gourimpatil@gmail.com

ABSTRACT - Vehicle ad hoc networks (VANETs) have become widely used in intelligent transportation systems over the past few years. The design of VANET incorporates crucial elements including autonomy, dispersed networking, and quickly changing topology. Significant commercial and academic interest has been generated by the properties of VANET and its applications for improving road safety, notably in studies concerning the possible improvement of transport networks. Some reactive protocols like AODV are suitable for such rapidly changing vehicular environment. This paper investigates the performance of AODV protocol in two different mobility models. The effect of accident event is tried to evaluate, the results shows that Packet Loss Ratio (PLR), Packet Delivery Ratio (PDR), Average End to End Delay, and Average Throughput have less value with accident event than without accident event.

Keywords-AODV, OMNeT, Routing Protocols, Simulation, SUMO, VANET.

I. INTRODUCTION

A Due to the high node mobility, a VANET is a type of wireless multi-hop network that must modify its architecture quickly (shown in Figure 1[1]). The number of vehicles with computer technology and wireless communication devices is increasing, and inter-vehicle communication is turning into a viable area for study, standardization, and growth.
Vehicles are very mobile, hence the topology of the VANET is constantly changing. Nowadays, the majority of VANET vehicles have the ability to integrate GPS and other technologies into their own system.

In network communication, routing techniques are crucial. The main objective of every routing protocol is to determine the optimal channel for communication between nodes (vehicles). Although VANET can use the conventional ad hoc routing protocols (AODV, OLSR, and GPSR), they are not adapted to the specific properties of such networks. Several specific protocols for VANETs have been developed to improve performance and suit the requirements of automotive applications. To evaluate the scalability and efficiency of a VANET protocol, some simulations with accurate mobility models are necessary.

In this research paper we have analyzed the performance of AODV routing protocol on CBR connection pattern with and without accident also different network parameters and different measured performance metrics such as Packet Delivery Ratio, Packet Loss, Throughput and End-to-End Delay are compared for varying number of vehicles for their performance analysis.

II. AD-HOC ON DEMAND DISTANCE VECTOR ROUTING PROTOCOL (AODV)

AODV protocol falls under the category of reactive protocols, where the method uses sequence numbers to guarantee change and uniqueness of the route. The protocol is not only scalable, but loop-free. The AODV protocol maintains a two-way path between the source and the destination. As it does not keep routes to nodes that are not communicating, AODV is especially successful in networks that are highly dynamic.

According to the AODV protocol's mechanism, a source node will send a Route Request (RREQ) message when the occasion demands for it. Until a route is found or the destination is reached, intermediate nodes will deliver this Route Request (RREQ) message while keeping track of the asking node's ID. An intermediary node or destination node sends a route back reply using an RREP message if the route is known till the source node while using a temporary route to the requesting node.

III. SIMULATION METHODOLOGY

This research uses an experimental methodology. In this instance, the behaviour of the AODV was examined in two separate real-world VANET environments, namely Mumbai and Mohali, using two distinct simulation tools, SUMO and OMNeT++. For constructing various scenarios, the Manhattan model and the Random Waypoint mobility models both were used.

The simulation visualization for the Mumbai Road and Mohali Road Networks developed in SUMO 1.8.0 can be seen in Figures 1, 2, 3, and 4.

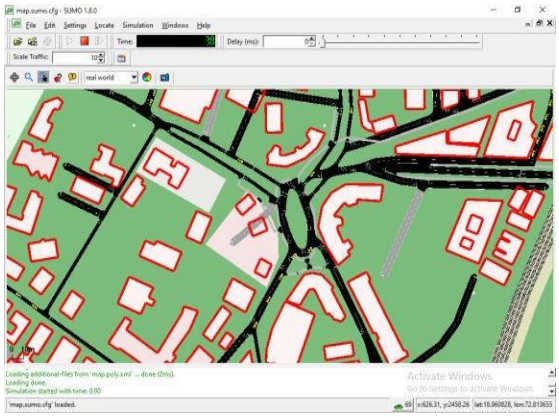


Figure 1. Mumbai Road Implementation in SUMO 1.8.0 (RWP mobility model)

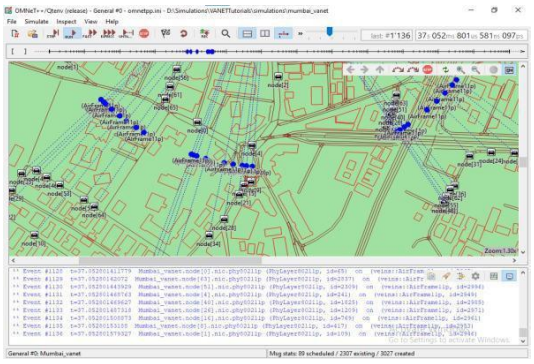


Figure 2. Simulation view in OMNeT++ 5.6.2 of Mumbai Road Network created in SUMO 1.8.0

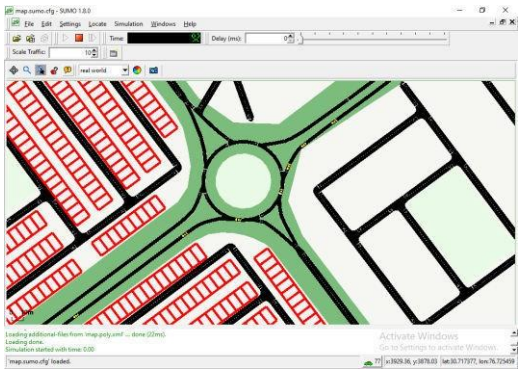


Figure 3. Mohali Road Implementation in SUMO 1.8.0 (Manhattan mobility model)

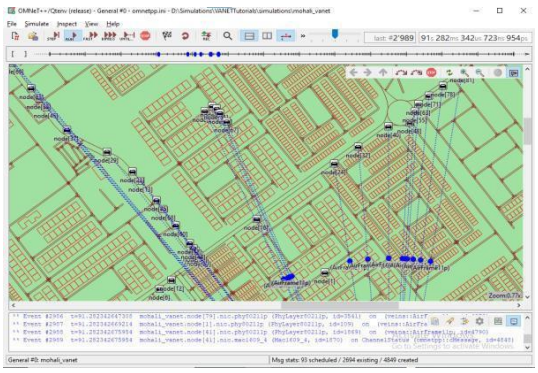


Figure 4. Simulation view in OMNeT++ 5.6.2 of Mohali Road Network created in SUMO 1.8.0

IV. PERFORMANCE METRICS

For the analysis of the AODV routing protocol, several performance measures are taken into account. Here is a list of some of them with definitions:

- **Packet Loss Ratio (PLR)**

Packet Loss Ratio defines the total number of lost packets over total number of transmitted packet by all nodes in the network. It is defined as

$$\text{Packet Loss Ratio (PLR)} = \frac{\text{Total No. Packets Sent} - \text{Total No. of Packets Rcvd}}{\text{Total No. of Packet Sent}}$$

- **Packet Delivery Ratio (PDR)**

The packet delivery ratio can be calculated by dividing the total number of data packets that have reached their destinations by the total number of packets that have been delivered from sources.

It is defined as

$$\text{Packet Delivery Ratio (PDR)} = \frac{\text{Total No. of Packets Received}}{\text{Total No. of Packet Sent}}$$

- **Average End to End Delay**

Average End-to-end delay is the amount of time it takes for a packet to travel from its source to its destination across the network. The mean end-to-end latency of all successfully delivered messages can be used to calculate the average end-to-end delay. As a result, the packet delivery ratio influences end-to-end delay to some extent. The probability of a packet dropping rises with distance between source and destination. The average end-to-end delay incorporates all potential network delays, including as propagation and transmission delays, retransmission delays at the MAC, and buffering and route discovery latency.

Average End-to-end delay is defined as,

Where,

i= Packet Identifier

T_{r_i} = Reception Time

T_{s_i} = Send Time

n = Number of packets successfully delivered

- **Average Throughput**

It represents the mean throughput value. Alternatively, it is expressed in packets per Time Interval Length units.

Mathematically it can be defined as,

$$\text{Avg. Throughput} = \frac{recvdSize}{stopTime - startTime} \times \frac{8}{1000}$$

Where,

recvdSize = Received packet's size

stopTime = Simulation stop time

startTime = Simulation start time

V. SIMULATION EXPERIMENT, RESULT AND ANALYSIS

Here city scenario is taken with maximum 300 vehicles. The reactive routing protocol AODV is used. Both Random Way Point and Manhattan mobility model are used for the experiment. The impact of accident on the performance of network is checked using the accident event. Also the transmission range is varied to check the performance.

Parameter	**Value**
Map	Mumbai (August Kranti Ground Area), Mohali (Mohali Tahsil office Area)
Protocol	AODV
Number of Nodes	10, 20, 30, 40, 50, 100, 150, 200, 250, 300
Simulation Time	2500 seconds
Traffic Type	CBR
Routing protocol	AODV
Transmission Range	2600 m
Mobility Model	Random Waypoint, Manhattan
Simulation area	10000 m*10000 m
Node Speed	max 22.22 m/s (80 km/hr)
Delay Time	00 sec
Interface Type	Queue
Mac Protocol	IEEE 802.11p and IEEE 1609.4 DSRC/WAVE
Packet Size	512 MB
Queue length	50
Radio Propagation Model	Two-Ray Interference model
Obstacle Model	simple obstacle shadowing model
Accident count	0, 1

Table 1. Various parameters used while varying accident event

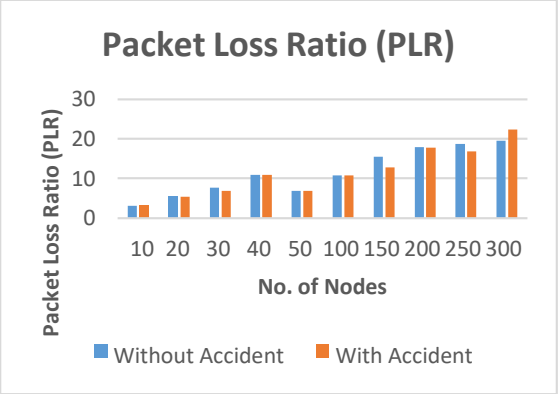
No. of Nodes	Without Accident				With Accident			
	Packet Loss Ratio (PLR)	Packet Delivery Ratio (PDR)	Average End to End Delay (ms)	Average Throughput	Packet Loss Ratio (PLR)	Packet Delivery Ratio (PDR)	Average End to End Delay (ms)	Average Throughput
10	3	4	7156.25	19.284	3	4.000	8156.25	16.92
20	5.286	6.286	3215.909	42.912	5.286	6.286	4181.818	33
30	7.591	8.591	1343.915	102.685	6.8	7.800	1517.949	90.912
40	10.893	11.893	930.931	148.239	10.893	11.893	924.925	149.201
50	6.861	7.861	1169.611	117.988	6.861	7.861	1113.074	123.981
100	10.676	11.676	693.607	198.96	10.676	11.676	746.683	184.817
150	15.423	16.423	576.444	239.399	12.744	13.744	481.058	286.868
200	17.784	18.784	323.558	426.508	17.669	18.669	349.604	394.733
250	18.657	19.657	343.146	402.161	16.72	17.720	328.79	419.721
300	19.471	20.471	319.094	432.475	22.338	23.338	341.988	403.523

Table 2. AODV performance in Mumbai Road Map while varying accident event

No. of Nodes	Without Accident				With Accident			
	Packet Loss Ratio (PLR)	Packet Delivery Ratio (PDR)	Average End to End Delay (ms)	Average Throughput	Packet Loss Ratio (PLR)	Packet Delivery Ratio (PDR)	Average End to End Delay (ms)	Average Throughput
10	0.500	0.000	140500	0.000	0.667	0.000	94333.	0.00
20	7.000	0.000	355000	0.000	3.750	1.875	44625.0	5.80
30	5.222	4.000	39777.78	13.877	3.563	3.125	22562.5	19.11
40	7.765	4.471	21117.65	29.214	3.880	5.560	15600	49.18
50	6.800	7.267	24000	41.783	4.156	7.188	12250	80.97
100	9.833	13.917	6466.67	296.985	8.775	13.056	6253.52	288.12
150	8.992	16.417	3558.33	636.674	8.430	15.124	3925.62	531.66
200	11.767	18.283	2943.40	857.192	10.869	18.769	3600.00	719.47
250	12.495	18.911	2559.41	1019.65	10.554	19.292	2707.92	983.16
300	11.017	21.280	2401.67	1222.77	11.338	21.317	2562.50	1147.98

Table 3. AODV performance in Mohali Road Map while varying accident event

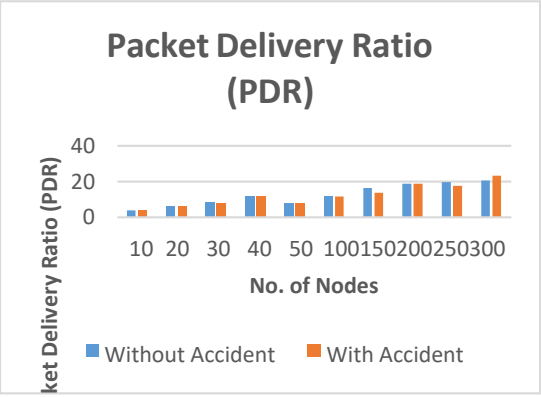
The Packet Loss Ratio (PLR) grows with the number of nodes, as seen in Graph 5. There is no appreciable difference in values between the PLR with accident and the PLR without accident.



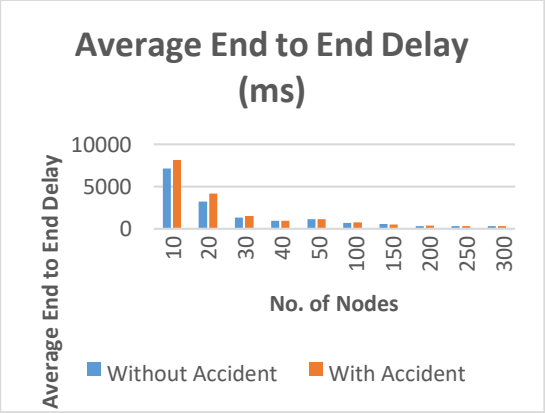
Graph 5. Graph for Packet Loss Ratio (PLR) in Mumbai Road Map while varying accident event

As seen in Graph 6, the Packet Delivery Ratio (PDR) increases as more nodes are added. There is no

apparent difference in values between the PDR with accident and the PDR without accident.

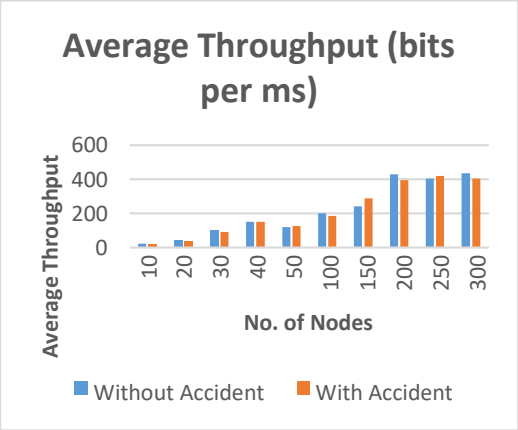


Graph 6. Graph for Packet Delivery Ratio (PDR) in Mumbai Road Map while varying accident event



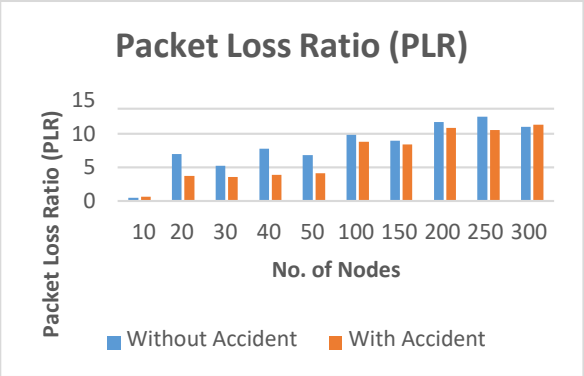
Graph 7. Graph for Average End to End Delay (ms) in Mumbai Road Map while varying accident event

Graph 7 indicates that the Average End-To-End Delay is greater in accident cases than in accident-free cases and its value decreases as number of nodes grow.



Graph 8. Graph for Average Throughput (bits per ms) in Mumbai Road Map while varying accident event

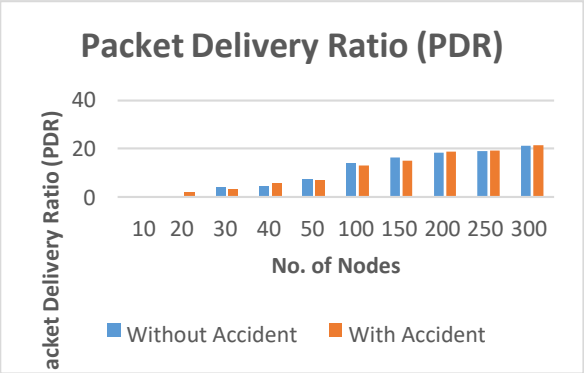
Graph 8 illustrates that the Average Throughput is lower in accident situations than it is in accident-free situations.



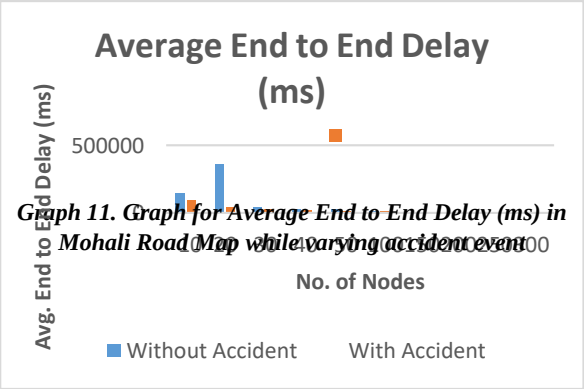
Graph 9. Graph for Packet Loss Ratio (PLR) in Mohali Road Map while varying accident event

In contrast to Packet Loss Ratio (PLR) without accident, Graph 9 demonstrates that PLR with accident has a lower value.

No observable difference between Packet Delivery Ratio (PDR) with accident and PDR without accident can be seen in Graph 10.

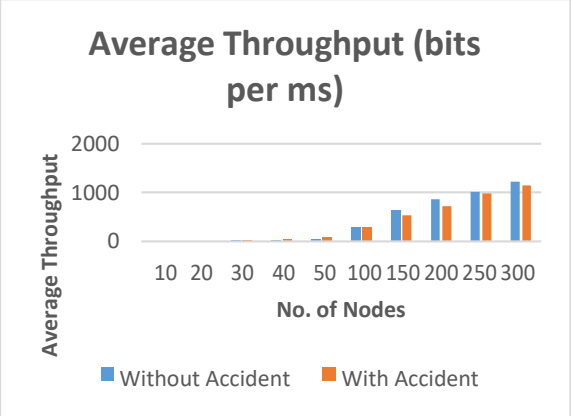


Graph 10. Graph for Packet Delivery Ratio (PDR) in Mohali Road Map while varying accident event



Graph 11. Graph for Average End to End Delay (ms) in Mohali Road Map while varying accident event

According to Graph 11, the Average End-to-End Delay is significantly reduced when there is of an accident compared to the Average End-to-End Delay in the absence of an accident event, and its value continuously drops as the number of nodes rises.



Graph 12. Graph for Average Throughput (bits per ms) in Mohali Road Map while varying accident event

According to Graph 12, the Average Throughput with Accident Event value is lower than the Average Throughput without Accident Event value.

VI. CONCLUSION

This study shows a simulation of a real-world road map. SUMO 1.8.0 and OMNeT++ 5.6.2 were used to analyze the various scenarios. The simulation lasts for 2500 seconds and produces a result file with scalar and vector data for various parameters. The result analysis carried out by observing the graphs.

Packet Loss Ratio (PLR), Packet Delivery Ratio (PDR), Average End-to-End Delay, and Average Throughput have less value with accident event than without accident event, according to the simulation experiment utilising Scenario with and without accident event.

When we compare both mobility model performance, Manhattan model has better performance than RWP model (low PLR, high PDR, low E2E delay and high throughput). Hence the network designed for Mohali city map is efficient than the network designed for Mumbai city map for the VANET environment.

REFERENCES

1. S. Almalki and F. T. Sheldon, "Addressing Polymorphic Attack Strategies with Misbehavior Detection for ITS," 2022.
2. M. Z. Iskandarani, "Vehicular Communication using Balanced Centralized and Decentralized Cluster Heads," *INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING*, vol. 16, pp. 718-723, 2022.
3. P. M. Mahajan and D. G. M. Patil, "A Brief Glance Towards The Simulation In Thevehicular Environment," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 10, no. 6, pp. g383-g393, 2022.
4. D. Yadav and N. K. Chaubey, "A Novel Two Stage Approach To Detect Sybil And Dos Attack In Vehicular Ad Hoc Networks," *Journal of Tianjin University Science and Technology*, vol. 55, no. 1, pp. 235-244, 2022.
5. J. S. Weber, M. Neves and T. Ferreto, "VANET simulators: an updated review," *Journal of the Brazilian Computer Society*, vol. 27, no. 8, p. 31, 2021.
6. I. A. Aljabry and G. A. Al-Suhail, "A Survey on Network Simulators for Vehicular Ad-hoc Networks (VANETS)," *International Journal of Computer Applications(0975 – 8887)*, vol. 174, no. 11, pp. 1-9, 2021.
7. K. D, "VANET: Framework, Challenges and Applications," *Indian Journal of Data Communication and Networking (IJDCN)*, vol. 1, no. 2, pp. 10-15, 2021.
8. J. Hu, J. Zhang and H. Nian, "Intimacy and Influence Based Scheduling Algorithm for Vehicle-to-Grid Network," *Journal of Software*, vol. 17, no. 2, pp. 48-57, 2021.
9. M. M. A. Elgazzar and A. Alshareef, "VANET Simulator: Full Design Architecture," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 3, pp. 2024-2034, 2020.
10. M. Ushakova, Y. Ushakov, I. Bolodurina, D. Parfenov, L. Legashev and A. Shukhman, "Research of productivity of software configurable infrastructure in vanet networks on the basis of models of hybrid data transmission devices," in *International Scientific and Technical Conference Modern Computer Network Technologies (MoNeTeC)*, 2020.
11. M. Amoozadeh, B. Ching, C.-N. Chuah, D. Ghosal and H. M. Zhang, "VENTOS: Vehicular Network Open Simulator with Hardware-in-the-Loop Support," *Procedia Computer Science*, vol. 151, pp. 61-68, 2019.
12. D. A. N. Upadhyaya and D. J. S. Shah, "AODV ROUTING PROTOCOL IMPLEMENTATION IN VANET," *International Journal of Advanced Research in Engineering and Technology (IJARET)*, vol. 10, no. 2, pp. 585-595, 2019.
13. A. Yanez, S. Cespedes and J. Rubio-Loyola, "A baseline for context-aware system for safety messages dissemination in VANETs," *Revista Facultad de Ingeniería*, vol. 93, pp. 9-18, 2019.
14. G. M. Patil, A. D. Shaligram and A. Kumar, "Performance Comparison of MANET Routing Protocols (OLSR, AODV, DSR, GRP and TORA) Considering Different Network Area Size," *International Journal of Engineering and Management Research*, vol. 6, no. 3, pp. 475-484, 6 2016.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023

Paper ID: CS-22

A Basic Study: Data Science and Artificial Intelligence

Prof.Archana Prakash Bhalerao,
Assistant Professor,
Bhusawal Art's science and P.O.Nahata commerce college Bhusawal
bhalerao.archana22@gmail.com

ABSTRACT - *Data Science, Data Engineering and Artificial Intelligence are hot topics in the current digital age. These technologies have changed the way humans describe a problem. These technologies work on data, but utilize it for different outcomes. Data Science and Artificial Intelligence are technologies that compare with each other in many ways. Artificial Intelligence in Data Science as a function has taken over technological Mechanization but requires Data Engineering in work to function properly.*

Technologies are interrelated in more ways than one. Data Engineering deals with the gathering and preparation of data so that it can be used by Artificial Intelligence in Data Science applications. Data Science utilizes this data and predictively and analyses it to gain insights. Artificial intelligence is concerned with working with data using tools to develop smart systems. Data science and artificial intelligence operate on data to produce similar results related to analysis.

Keywords - *Data Science, Artificial Intelligence.*

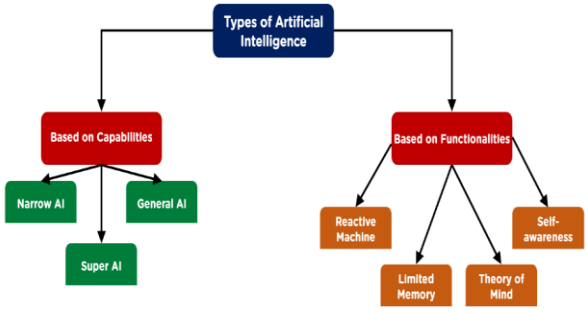
I. INTRODUCTION

Artificial Intelligence

The ability of digital computers to perform tasks that are normally performed by humans is called artificial intelligence. AI (artificial intelligence) attempts to mimic the human mind by incorporating problem-solving, decision-making, and reasoning abilities into machines.

Development in the field of artificial intelligence began after the development of computers in the 1940s. The field of artificial intelligence research was established in 1956 as an academic discipline. Since then the development in the field of data science and artificial intelligence has been rapid. Since then there have been great improvements in how well machines perform complex tasks. Yet, despite continuing these advances, computers have not been able to match the flexibility of the human mind.

Types of Artificial Intelligence



There are 3 types of Artificial Intelligence-based on capabilities -

- Narrow AI
- General AI
- Super AI

And 4 types of Artificial Intelligence-based on functionalities.

- Reactive Machines
- Limited Theory
- Theory of Mind
- Self-awareness

Based on Competencies:

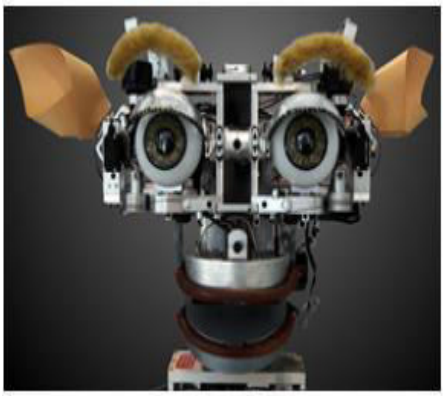
Artificial Narrow Intelligence (ANI): This is the most basic form of artificial intelligence. These systems are designed to efficiently solve a single problem. They have narrow capabilities meaning they can excel at a specific task but in a highly controlled environment and with limited parameters.

1. **Artificial General Intelligence (AGI):** This is the theoretical concept of artificial intelligence. Its main objective is a machine with human-level intelligence in various parameters such as language processing, image processing and computational capabilities. Multiple ANIs need to work in unison for an AGI to function. With the most advanced computing pieces, such as Fujitsu's K and IBM Watson, it took about 40 minutes to simulate one second of the human brain's neuro-communications. This shows that our computing power is not enough and therefore AGI is still theoretical in nature.
2. **Artificial Super Intelligence (ASI):** This is the most advanced theory made for artificial intelligence. This theory states that artificial intelligence will surpass human thinking ability by constantly adapting and being able to perform multiple tasks simultaneously. As computing capabilities have not yet reached the threshold needed to mimic human intelligence, AGI is still a theory. As ASI is an advanced version of AGI, its reality may not be seen in near future.

Based On Functionalities:

1. **Reactive Machines:** Reactive machines perform basic operations. This level of A.I. is the easiest. These types reply to some input with some output. No such education takes place. This is first stage of any A.I. A machine learning system is a simple, reactive machine that takes a human face as input and outputs a box around the face to identify as a face. A model does not store any input, it does not do any learning.
2. **Limited memory**: Limited memory artificial intelligence trains from past data to make decisions. Such systems have a short-term memory. They can use this past data for a certain period of time, but they cannot add it to their library of experiences. This type of technology is used in self-driving vehicles.
   - Limited memory AI monitors how other vehicles are moving around them, both currently and as time passes.
   - This ongoing, collected data is added to the AI machine's static data, such as lane markers and traffic lights.
   - They are included when a vehicle decides when to change lanes, avoid cutting another driver, or collide with an adjacent vehicle.
3. **Theory of Mind:** Theory of Mind represents an advanced class of Artificial intelligence technology & exists only as a concept. This type of AI requires a deep understanding that people and things in the environment can change emotions and behavior. People's emotions, feelings, thoughts should be understood. Although there have been many improvements in this area, this type of AI is still not fully developed.

A real-world example of the AI theory of mind is Kismet. Kismet is a robot head developed in the late 90s by researchers at the Massachusetts Institute of Technology. Kismet can recognize human emotions by mimicking them. Both abilities are major advances in theory of mind AI, but Kismet cannot stare or pay attention to humans.



Sophia from Hansen Robotics is another example where the AI theory of mind was applied. Cameras in Sophia's eyes, combined with computer algorithms, allow her to see. She can make eye contact, recognize people and follow faces.

4. **Self-awareness:** A self-awareness AI exists only hypothetically. Such systems understand their internal characteristics, states and conditions and sense human emotions. These machines are smarter than human mind. This type of AI is not only able to understand and evoke the emotions of those it interacts with, but also has its own emotions, needs and beliefs.

The purpose of artificial intelligence

The main purpose of artificial intelligence is to assist human abilities and predict far-reaching consequences that the human brain cannot process. Artificial intelligence has the potential to reduce the drudgery of human labor and provide a way for the human species to thrive in a beneficial way. Artificial intelligence in data science has similar objectives.

Applications of artificial intelligence

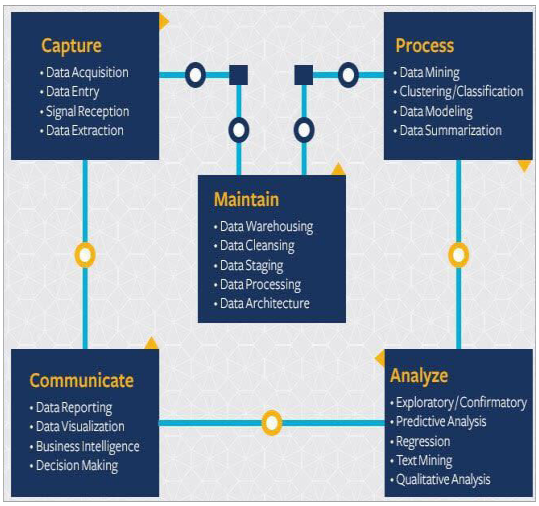
While the field is still under development, the scope of applications increases with each iteration. AI in data science is often used in real life. The applications of Artificial Intelligence are:

- **Personalized online shopping:** A user's search trends and search history are tracked and, based on the data, specific product ads are shown that may meet the user's needs and expectations.
- **Enhanced imaging and surveillance:** Images are enhanced using computer vision, which is used by apps like Snapchat and Instagram. Image enhancement is also used for surveillance by security and military services.
- **Video Games:** Computer games have bots that are controlled by the system. These characters are customizable meaning they change the difficulty level based on the actual player. It works on the adaptive capabilities of artificial intelligence.
- **Healthcare**: This is the sector that has made the most use of AI. Automated systems that facilitate drug development have helped find cures for more diseases. Applying artificial intelligence to historical data has helped predict the impact of bacteria and viruses.
- **Chatbots**: It has become necessary to include alternative chatbots in websites and online stores. These provide maximum information in the most humane way possible. Artificial intelligence in data science-based chatbots works efficiently.

Data Science

Data is a boon to organizations only if processed efficiently. Data science involves the study of data, its origins, its value and transformations to gain valuable insights. Today's businesses run on large amounts of data, and standard business intelligence tools fall short when processing large amounts of data simultaneously.Data science has more advanced features, which can process such large amounts of unstructured data. It can process data from sources such as financial logs, multimedia files, marketing forms, sensors, instrumental values and text files.

The following image shows the process data to perform data science efficiently. This is also known as data science life cycle.



- **Capture data**: Data capture is the collection of raw data. Data is obtained from various

sources like data entry, signal reception and data extraction process is followed.

- **Maintain the Data:** Data is Stored in Data Warehouse after Data is Cleaned and processed.
- **Process the Stored Data:** The Data from the warehouse is Processed, Mined clustered, and summarized.
- **Analyze data:** Exploratory analysis by regression, predictive analysis and qualitative analysis.
- **Communicate results**: Visualize results using data reporting and business intelligence tools.

Purpose of Data Science

The main purpose of data science is to find patterns in data. It is used to analyze and gain insights using various statistical techniques. Current data and historical data are used to predict future results. These valuable forecasts and insights give businesses the opportunity to thrive and adapt based on market trends.

Applications of Data Science

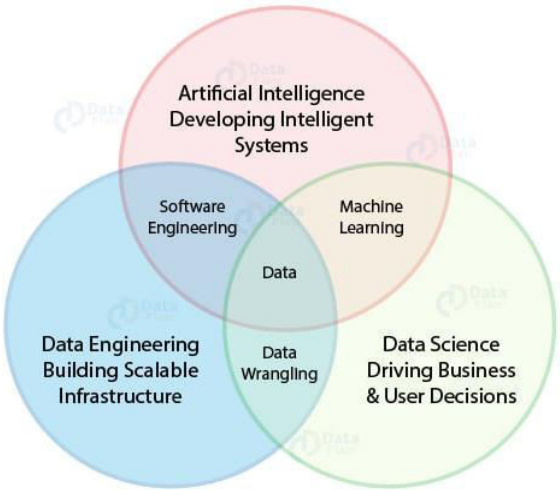
Data science works on data and as the amount of data is growing at a very fast pace, its benefits are also growing at a good pace. Artificial Intelligence works prominently on this growing amount of data in data science methodologies.

- **Banking:** Data science allows banks to use data-driven resources efficiently. Data science allows risk management and risk modeling based on customer data. Also predict customer churn and fraud detection using data.
- **Production:** Data science allows for optimizing production, reducing costs and increasing profits. Also, the inclusion of data from sensors allows detection of potential problems in the system. Also, data allows optimizing quality and production capacity.
- **Transportation:** Data science helps build systems for self-driving cars using sensory data. Data science enables extensive analysis of fuel consumption patterns, driver monitoring and path selection to help optimize the industry.
- **Healthcare:** Data science helps in predictive analysis of prognosis, drug discovery based on disease data and medical image analysis to predict diseases from images.
- **E-commerce:** Data science helps in finding potential customers. It helps in optimizing the customer base and clusters them based on trends. It is also used to roughly analyze goods and services for maximum coverage. Using customer data, companies use sentiment analysis to find feedback based on reviews.

Artificial Intelligence in Data Science: Understanding the Relationship

Data science and artificial intelligence are closely related. All have data as a common element as shown in the image below. Since data engineering deals with data extraction, transformation and storage, this is the most important step. Artificial intelligence in data science applications operates on processed data so raw data can only work after being engineered.

The image below depicts the products generated by the interplay of the 3 concepts. Machine learning is created by data science and artificial intelligence. Software engineering is shaped by artificial intelligence and data engineering.



Role of Artificial Intelligence in Data Science

Artificial intelligence plays an important role in enabling data science. The following points illustrate the role of AI in the field of data science:

Machine learning is a supervised version developed in conjunction with data science and artificial intelligence, where a limited amount of data is fed into a system to predict probabilities.

- Data science and artificial intelligence are terms that are used interchangeably because of their work, but artificial intelligence is a tool of data science. Artificial intelligence is not fully represented by data science because data science is only concerned with predictive analytics and uses machine learning tools for that. Machine learning is only a subset of artificial intelligence, and AI can provide many complex tools for analysis.

Comparing Data Science and Artificial Intelligence

Data science was created with the aim of finding hidden trends in large amounts of data. This discipline is useful for extracting raw data, processing it and analyzing the data for better understanding. On the other hand, we can deploy artificial intelligence to manage data autonomously. This means, we can remove human dependency from our work and fully automate it. Compares Data Science & Artificial Intelligence using the following 3 factors:

- Goals
- Fundamental Technologies
- Use Cases

Goals

The main objective of data science is to finalize an appropriate problem statement, document business requirements & use data analytics & machine learning models to develop viable solutions. Furthermore, data scientists also perform data visualization to present the insights generated from their proposed solutions.

The main goal of artificial intelligence is to simulate human intelligence using computers so that machines can make smart decisions in complex situations. To achieve this goal, AI professionals work to develop new algorithms, optimize existing neural networks, & perform data automation to process large volumes of data.

Fundamental Technologies

Data science leverages many statistical techniques to process and transform large datasets. This domain deploys machine learning models on source data to identify actionable insights. To achieve this goals, data scientists deploy tools such as Tableau, the Python programming language, MATLAB, Natural Language Processing (NLP), & more.

Artificial intelligence mainly relies on machine learning-driven algorithms that are designed for distinct purposes. AI professionals use a variety of tools to teach computers how to make decisions. learning process. All the work done in the field of AI revolves around tools like Keras, Spark, Tensor Flow, Scala, Skit Learn etc.

Use Cases

An important factor in comparing data science & artificial intelligence is their use cases. The following use cases are beneficial for applying data science methodologies:

- Identifying market patterns & popular trends.
- To perform exploratory data analysis (EDA) for business.
- Require high-speed mathematical processing.
- Work related to predictive analytics.

Future of Data Science

The future of data science is believed to witness some of the biggest innovations seen in the last decade, from the data explosion to the rise of the Internet of Things (IoT) and social media. Experts predict that over the next decade, the rise of machines will increase the use and usefulness of computer systems and mobile devices.

Moreover, experts also claim that the use of social media will increase exponentially and users will consume huge amounts of data online. Consumers will use social media for entertainment, transactions, surveillance, etc. Some experts predict that machine learning algorithms will also see experts.

Future of Artificial Intelligence

Artificial intelligence is an emerging industry that is becoming a proxy for the human brain. It performs various business functions without human intervention like customer interaction, building brand awareness on social media etc. Many experts believe that AI can beat humans in almost every cognitive task.

Artificial intelligence applications are transforming the healthcare, insurance, finance and marketing sectors by automating various administrative processes such as employee or patient record management, conducting market research and communicating with potential customers.

II. CONCLUSION

This review paper provides a brief overview of popular technologies such as data science & artificial intelligence. It also covered the relationship between data science & AI. Furthermore, it discusses the roles of artificial intelligence in data science & provides a comparison between the 2 technologies.

Artificial intelligence plays a vital role in data science by providing advanced tools for proper predictive analysis and appropriate parameters to be applied to data engineering software as well. As we learn about the future of data science and AI systems, we will look at how to develop a successful career in the respective fields.

REFERENCES

1. Brodie, M.L. "Understanding Data Science: An Emerging Discipline for Data-Intensive Discovery", in Shannon Cutt (ed.), Getting Data Right: Tackling the Challenges of Big Data Volume and Variety, O'Reilly Media, Sebastopol, CA, USA, June 2015
2. "A Review of Artificial Intelligence Methods for Data Science and Data Analytics: Applications and Research Challenges." Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2018) IEEE Xplore Part Number:CFP18OZV-ART; ISBN:978-1-5386-1442-6
3. "Application In Artificial Intelligence" Contemporary Research In India (ISSN 2231-2137): Special Issue : April, 2021

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-23

Natural Language Processing: Objectives and Scope

1.Vrunda M. Chaudhari
Smt. P. K. Kotecha Mahila Mahavidyalaya, Bhusawal

2.Manisha S. Deshmukh
G.H.Raisoni Institute of Engineering and Business Management, Jalgaon

3.Gajendra.R.Wani
Bhusawal Arts, Science and P. O. Nahata Commerce College, Bhusawal

ABSTRACT - *Natural Language Processing refers to the branch ofComputer Science, Machine Learning and Artificial Intelligence technologies, to infer linguistic analysis. Natural language processing concerned with the conversation between machine and human language. In simple word, it allows computers to understand the text. Around the globe many languages are spoken and messages are share virtually every single day. By analyzing messages, computer can recognize correlations, attitudes, entities and other useful information. This is improving, because now machines can understand beyond 0's and 1's or simply put machine language. Apart from, natural language processing understanding the simple messages, concepts, themes, sentiments and emotions. In this paper, we have also seen its applications in education field to understand the objectives and scope.*

Keywords*: Natural Language Processing,Natural Language Understanding*

I. Introduction

Human communicates with each other mostly via text or speech. The human way of communicating is known as **natural language**. Around the globe many languages are spoken and messages are share virtually every single day.All these conversations, feedback and messages are data and themselves. This data are extremely valuable at is can give us customize information and insight into human sentiment. However, this data is not useful to computer at it is in the form of that can understood by machine. Machine can communicate using 1's and 0's and not via words. They cannot understood French, Spanish, English, only binary. Hence Natural Language processing comes in the picture.

Natural Language Processing refers to the branch of Artificial Intelligence that allow machine to understand human language. Using the natural language, the ultimate the objective of NLP is to read, decidable, understand and make sense up of a human language. In a manner that is valuable and to build the system that make sense up text and perform task like translation, grammar checking and a topic classification. NLP basically is the task of processing written form of languages and making a computer understand them. Companies are increasingly using NLP equipped tools to gain insight from data and to automate routine task.Apart from, natural language processing understanding the simple messages, concepts, themes, sentiments and emotions.We have also seen its applications in education field to understand the objectives and scope.[1]

In India's context, however, NLP is in the preliminary stage.This has particularly to do with the various spoken languages and dialects in India. Even though English is our official language, most of the Indians prefer speaking in regional languages. Given that Natural Language Processing is based on parsing text into computer language or data and then use to recognize the pattern, making NLP mainstream in India would require a deep learning of regional languages.

This survey is structured as follows. In Section 2, we define a general model of NLP Pipeline and its components. We use a well-known text parsing

components introduced by N.Peng et.al.[2]as an illustration and instance of the general NLP Pipeline. In Section 3, we elaborate on the uses of attention in various NLP tasks. Section 4 supports NLP libraries in some of the languages of the Indian constituent. Section 5, discuss recent technology of NLP pre-trained models. Section 6 is devoted to use-cases and key benefits of NLP. Section 7 concludes this article.

II. NLP Pipeline

A pipeline is a set of data processing element which are connected in series whether the output of one element with the input of next one. It represents various steps which are taken in computation and the order which they occur [2].

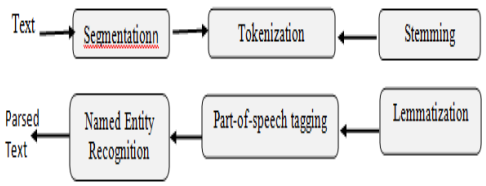


Figure 1: Components of NLP Pipeline

Segmentation

The process of dividing a sentence into its component sentences, usually along punctuation marks.

Example: Consider the following paragraph

The lemonade quenched her thirst,but nor her longing

Sentence segmentation produces the following result.

a. The lemonade quenched her thirst.

b. But nor her longing.

- Tokenization :

Tokenization is the process of splitting sentences into their constituent words.

Example:

Natural Language Processing

Word tokenization generates the following result

"Natural","Language","Processing"

- Stemming:

The process obtaining the words Stem of a word. Word Stem give new words upon adding affixes to them.

Example:

Jumping, Jumped and Jumps, all these words are originated with a single root word "Jump".

- Lemmatization :

The process of obtaining the Root Stem of a word. Root Stem give new base form of word.

Example:

In lemmatization, the words did, doing, done and does has a root word do, which has a meaning.

- Part of speech tagging (POS) :

Identifies which part of speech a word belongs to. It tags a word as a verb, noun, pronoun etc.

Example:

The lemonade quenched her thirst.

In the above example, **The** used as determine, **lemonade** used as noun, **quenched** is used as verb, **her** used as pronoun and **thirst** is used as noun.

- Named Entity Recognition:

Named entity recognition determines words that can be categorized into subgroups.

Example:

Consider the following sentence:

"Smith was born in Bhusawal on February 1, 1998"

In this sentence, "Smith" is a person, "Bhusawal" is a location, and "February 1, 1998" is a date. We can extract valuable information about Smith's birthplace and birthdate by identifying and classifying these named entities.

1. Components of NLP

Natural Language is that the language refers to the human way of communicating through text or speech. Natural language processing permits machines to interpret meaning from the word and phrases that person useand acknowledge in a similar way once presenting information back to them.

Generally, there are 2 components of NLP, the first one is Natural Language Understanding and the second is Natural Language Generation[3].Thesetwo components further divide into various sub-components, as shown in following Figure 2.

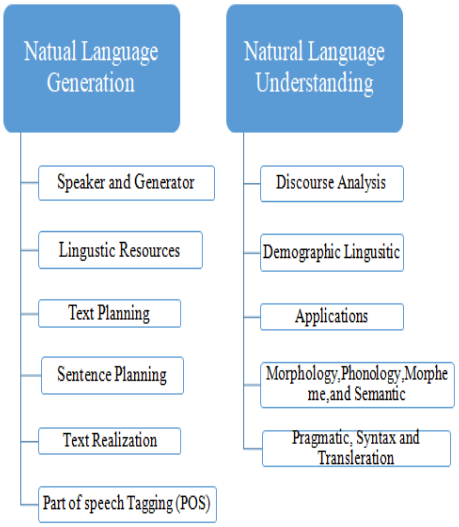


Figure 2. Components of NLP

a) Natural Language Generation (NLG)

Natural Language Generation acts as a translator that converts the computerized data into natural language representation. According to Khurana et al[4], the NLG components consist of speaker and generator; components and levels of representation (sentence planning, text planning, linguistic resources, and text realization); and application or speaker.

- Speaker and generator:

To parse a text, we need to have a speaker and a generator or a machine that articulate phrases significant to the situation.

- Sentence Planning :

Sentence Planning includes selecting required words, constituting meaningful phrases, forming tone of the sentence.

- Text Planning:

Text Planning includes extracting the specific content from knowledge base.

- Text Realization:

Realization of the chosen and organized resources must be realized as an actual text or voice output.

- Linguistic Resources :

Linguistic resources [5], to assist the realization of information, linguistic resources must be select. These resources will come down to choices ofspecific words, idioms, syntactic structures.

- Part of speech tagging(POS) :

Identifies which part of speech a word belongs to. It tags a word as a verb, noun, pronoun etc.

a. Natural Language Understanding (NLU)

Natural Language Understanding helps the machine to understand and analyse human language by extracting the metadata from content.

- Demographic Linguistic :

Demographic linguistic is the realization of speech and text to implement better understanding of application in speech recognition, artificial intelligent web browsing and machine translation.

- Discourse Analysis :

Discourse Analysis deals with how the immediately preceding sentence can affect the interpretation of the next sentence.

- Applications:

Some of these application areas include Question Answering, Spam detection, Sentiment analysis, Chatbot, Speech Recognition, Machine translation, spelling correction and Information extraction

- Morphology,Phonology, Morpheme and Semantic:

These are 4 distinct, yet related terms used in NLP [6].

Morphology: Study of construction of words from primitive meaningful units.

Phonology: Organizing sound systematically. .

Morpheme: Primitive unit of meaning in a language.

Semantic: Concerned with the meaning of words and how to combine words into meaningful phrases and sentences.

- Pragmatic, Syntax and Transliteration :

Pragmatics: Pragmaticdeals with using and understanding sentences in different situations and how the interpretation of the sentence is affected [7].

Syntax: Refers to arranging words to make a sentence.

Transliteration: Refers the process of transferring a word from the alphabet of one language to another. Transliteration only gives you an idea of how the word is pronounced, by putting it in a familiar alphabet

2. NLP in Indian Context

Putting NLP in the context of India becomes few challenges though. Only 10% of Indians speak and understand English language. Remaining 90% speak and understand the regional languages. This poses language barrier for the technology such as NLP.

There are 3 libraries supported in some of the languages of the Indian subcontinent [8].

- iNLTK: Hindi, Sanskrit, Gujarati, Kannada, Malayalam, Nepali, Odia, Marathi, Bengali, Tamil, Urdu
- Indic NLP Library: Assamese, Sindhi, Sinhala, Sanskrit, Kannada, Telgu.
- StandfordNLP: Many of the above languages.

3. Technology innovation

Broadly recent development has been GPT-3 pre-trained language model. It is the large scale transformer- based language model. GPT-3 was trained with 175 billion parameters,10 times more than any non-sparse language model.

Following timeline highlights the updates in the NLP field, beginning from 2018 shown in following Figure 3.

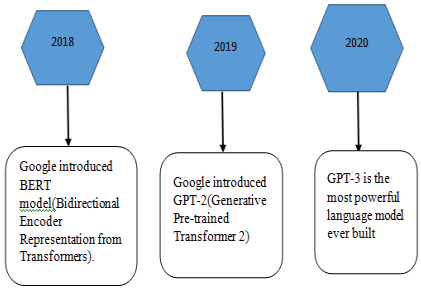


Figure 3. Developments in NLP

4. Use-cases and key benefits of NLP

Natural Language processing is just beginning to impact business management beyond different industries [9].

- Financial services sector:

NLP offers ultimate use cases for financial services sectors, particularly around sentiment analysis and content enrichment that can help the financial sector in making an informed decision about investments and risk management. NLP, when applied on the unprocessed data, can help in pulling insights from underused data points.

Manufactures can use NLP transportation-related information to streamline processes and improve automation.

- Insurance :

Insurance companies can use NLP to recognize and dismiss fraudulent claims. Insurance agent can use machine learning and artificial intelligence to analyse customer communication to recognize indicators of fraud and mark out these claims for deeper analysis.

- Healthcare:

NLP can interpret patient communication through chatbot, from emails, patient's helpline and with the help of medical professionals prioritize patients on their needs, improving patient's diagnosis and treatment.

III. Conclusion

For processing some natural language, it is essential to "understand" them. The text parsing results can be vague without syntax and semantic analysis. It is subpart of linguistic language that studies meaning. This article discuss various natural language generation and natural language understanding techniques by different researchers on natural language processing. Recently, the area has led to the incorporation of contextual terms based on language pre-trained models. This review studies the different pre-trained language models such as BERT, GPT-2 and GPT-3 to mention a few. Broadly, recent development has been GPT-3 pre-trained model.

IV. Reference

[1] Nadkarni Prakash, Ohno- Machado Lucila, and Chapman Wendy. Natural language processing: An Introduction. Journal of American Medical Information Association, vol. 18, pp.544-551, 2011.

[2] N.Peng, F. Ferrao, M.Yu, N. Andrews, A Concrete Chinese NLP Pipeline, Proceedings of NAACL- HLT 2015, pages 86–90, Denver, Colorado, May 31 – June 5, 2015.
Vol.3, No.1, February 2013

[3]A.Reshmwala, D.Mishra, P.Pawar,Review of Natural Language Processing, IRACST – Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498,

[4] Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. arXiv preprint arXiv:1708.05148 2017.

[5] S. Pais , J. Cordeiro and M. Luqman Jamil, NLP-based platform as a service: a brief review, Pais et al. Journal of Big Data (2022)

[6] Liddy ED., Natural language processing, 2001.

[7] Friedman C, Johnson SB. Natural language and text processing in biomedicine. In: Springer (ed.) Biomedical Informatics, 2006;312–343.

[8] G.Arora, iNLTK: Natural Language toolkit for indic languages, arXiv 2009.12534v2, Oct-10-2020

[9] M.Bahja, Natural Language Processing applications in business, E-Business-Higher Education and Intellience 2020

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID:CS-24

Leveraging influencing attributes to predict mental health using machine learning algorithms

Mr.Anand J. Maheshwari,
Dr. Assistant professor, R.C.Patel Arts, Commerce, and Science. College, Shirpur, K.B.C.N.M.U., Jalgaon, India,
a.j.maheshwari39@gmail.com

Dr. G. R. Wani
Professor and Head,DOCS, BASPONC College, Bhusawal, K.B.C.N.M.U., Jalgaon, India
gajuwani03@gmail.com

ABSTRACT-*Youths develop physically, sociologically, and psychologically during their years at college or univ_compus. Youngsters are susceptible to acquiring psychological disorders. Issues during their studies due to the variety of pressure and expectations they must negotiate with. Age, relationship status, panic disorder, nervousness, and depression are all crucial factors in determining a student's present state. Because of its rapid development, machine learning is now widely employed to address a wide range of problems in fields such as biology and medicine. Laying depression at the center, in this paper we analyze data and apply machine learning algorithms to predict whether the student needs to consult a specialist for treatment and perform analytics on their accuracy.To predict whether or not a student requires treatment, this Paper uses Random forest, naive Bayesianalgorithm, and k-NN algorithm and predictsoutcomes. Though naive Bayes appears to perform well, random forest and k-NN predicts with greater accuracy.*

Keywords-*student mental health, machine learning, algorithm, accuracy, analytics*

I. INTRODUCTION

The significance of mental health in college students cannot be overstated. Students may experience psychological and emotional strain as they leave that entire familiar behind and gain entry to university. Students with different courses, age range, and decades may suffer from a mental breakdown at some point in their lives. A few examples anxiety are: feelings of awkwardness, impatience, or stress that are out of proportion. An unfounded sense of imminent disaster, threat, or distress, convulsing, perspiring or quivering, vulnerability and tiredness, somnolence or trouble resting, metabolism problems (not eating enough or binge eating), headaches or dizziness.A number of contributing factors in a university student's lack of good mental state. Academic stress, athletics, relationships, and irresponsible judgement are just a few instances.Another largest source to a collegiate student's emotional stability is substance abuse. Liquor, e-cigarettes, medicine pain relievers, and sometimes even cell phones have a significant effect on the way people lives. [1]College learners are especially susceptible during their growth because the most of career psychiatric conditions manifest themselves in between ages of 18 and 24.[2]Multiple genetic, biological, and social and economic factors most impact the likelihood that an individual grows a mental disorder.Depression has an influence on a person's thoughts, behavior, inspiration, emotions, and perception of well-being. Students' interpersonal, academic, and All of these issues can have a sizable effect on their psychological wellness.[3]A sense of belonging and a strong support network were strongly linked to mental wellbeing and happiness, whereas interpersonal and private isolation were heavily linked to poor mental health.[4]

Machine learning techniques have the ability to provide exciting opportunities for gaining knowledge

human patterns of behavior, as well as acknowledging signs of psychological illness and comorbidities, establishing increased infection forecasts, and types of works and boosting therapeutic approaches. Regardless of the potential advantages of using ML in psychological disorders, this is a new field of study, and creating an effective ML-enabled app is fraught with an array of complicated, interrelated obstacles.To be completely incorporated and beneficial in practical psychological settings, ML solutions should also take into account the conceivably much further intimate, cultural, and practical impacts of ML initiatives. [5]

For the past twenty years, machine learning methods have been regularly used in the growth of forecasting analytics to aid in excellent decisions.Accuracy assessment, recall, precision, and thus the space beneath the ROC will serve as the basis to gauge the efficacy of such techniques. Data analysis has recent times become an established minimal tool for uncovering hidden correlations in datasets. [6]

This research study is divided into the following sections: Section 2 is about machine learning methods, section 3 is about data preprocessing, Section 4 is about a broad view of experimental analysis fifth section discusses the results and discussion, while sixth tries to address the conclusion and future scope.

II. MACHINE LEARNING METHODOLOGIES

In ML techniques, the learning methods are frequently broken down into two groups: unsupervised and supervised learning. Modeling the correlation between estimated current attributes and maybe some label related to it is what Learning Algorithm is all about; once proven, this prototype can be used to assign probabilities to fresh, unlabeled data. This is split further into assignments for regression and classification.The labels in categorization are discrete categories, whereas the labels in correlation are continuous quantities.It anticipates distinct or hypothetical unambiguous class labels and categorises data (assembles a model) on the basis of the training set and the values (class labels) in a categorization ascribe, which it then uses to identify new information. [9]

Unlabeled data models the characteristics of a dataset without using labels. These features serve various tasks such as clustering and dimensionality reduction.Data is automatically classified into discrete groups.Clustering techniques range of specialties data collections, whereas features extraction algorithms look for more summarized data models. [14]

Among the most promising technologies is classification, which extracts data from data sets in the form of instances and designates class labels to them. It enables the retrieval of models.

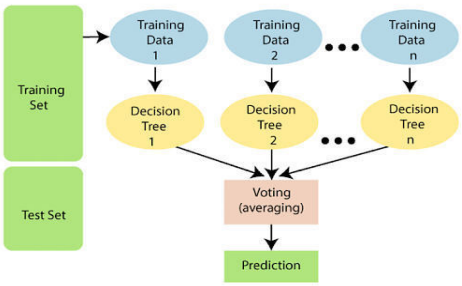
Random forest:

Decision Tree algorithm is a supervised learning method that averages a collection of decision trees on subsets of a given data to continue improving its forecast performance. A RF is a conceptual appraiser that aims to optimize predictive ability and overrules over-fitting by adapting a sum of boosted decision trees on different sets of the data source. The RF approach is rational in nature, with each version encompassing the choices of one sample group of size N from the data set with substitute and also another large selection from of the indicators without substitute. After that, the data is divided into parts. The data from the bag is then fallen, and the intervention trials are repeated lots of times based on the amount of shrubs needed. Finally, the trees that categories the assertions into one of two categories are counted. Just after that, instances are categorized using a decision tree and a clear majority. RF must adhere to the following requirements in order to operate well:

There must be some actual indication in our attributes for models built with either of those features to outperform arbitrarily defined making assumptions.. The specific trees' forecasts (and hence errors) must have weak relationship between them.It attempts to create an uncorrelated forest of trees by using bagging and feature unpredictability when constructing each specific forest.

Overfitting is a major issue with decision tree classifiers, and random forest helps to avoid it.

Multiple trees are created using different sample sizes and feature sets to create a random forest.



Figure#1: Working of random forest Algorithm[Source:java point]

Naïve BayesianAlgorithm:

A learning algorithm based on Naive Bayeswith estimator classes.It is a categorization technique that is based on the Bayes Rule and the estimator freedom insinuation. Simply put, a naive Bayes algorithm implies that the presence of one feature in a class is incidental to the emergence of any further feature. Naive Bayes methodologies are a type of algorithmic auto encoder that relates Bayes' theorem with the "naive" assertion of volitionally between any two features based on a class variable value. Bayes' theorem states the following relationship x1 through xn given a class variable and a relying feature vector:

$$P(y \mid x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n \mid y)}{P(x_1, \dots, x_n)}$$

Using the naive transition probability supposition,

$$P(x_i|y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i|y),$$

This association is refined for all i's:

$$P(y \mid x_1, \dots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1,\dots,x_n)}$$

Since $P(x_1, \dots, x_n)$Provided that is a constant value, we can apply the following classifier:

$$P(y \mid x_1, \dots, x_n) \propto P(y)\prod_{i=1}^{n} P(x_i \mid y)$$

$$\Downarrow$$

$$\hat{y} = \arg\max_y P(y)\prod_{i=1}^{n} P(x_i \mid y),$$

And we can forecast using Max Posteriori (MAP) approximation$P(y)$ and $P(x_i \mid y)$

Despite their seemingly overly simplistic hypotheses, naive Bayes classifiers have exceeded all expectations in a wide range of practical applications. To approximate the necessary parameters, they need a relatively small amount of data to be trained.

K-NN:

This algorithm is classified as instantiation learning phase or lazy learning.Lazy learning is the term used to indicate that the classifier doesn't really build a model until a forecasting is needed; k-NN is exhorting since it accepts no records beyond the distance metric which may be approximated among both two instances.

Assume that m represents the total amount of learning data samples. Let p represent an unnamed point.

a. Save the training samples in a sample points array arr[]. This implies that each element of such a array reflects a tuple (x, y).
b. from 0 to m:Determine the Distance measure d (arr[i], p).
c. Generate a set S of K least distances procured. Each of those distances is equivalent to an earlier classified data point.
d. Reinstate the conventional label.

Objectives and significance of the study:

1. The current work will investigate novel approaches to categorizing students' mental health.
2. The task is to train classifiers and evaluate algorithm efficiency and effectiveness.
3. To assess classifier performance measurement.

III. RELATED WORK

[1]Khaldoun M. AldiabatNasir A. Matani , Carole-Lynne Le Navenec, titled ” Mental Health among Undergraduate University Students: A Background Paper for Administrators, Educators and Healthcare Providers”, demonstrates some statistics about the incidence, comorbidities, and implications of psychological disorders among students, and also the motives why these educators do not receive help.

[2] MersihaĆeranić, SamedJukić, titled,” Depression and Anxiety Analysis and Prediction using Decision Tree and Logistic Regression”, They carried out studies to find which factors contributed to the greatest extent to the occurrence of symptoms of anxiety and depression even during COVID-19 lockdown, and they also determined who was most highly susceptible to the beginnings of psychological illnesses in regard to age, identity, academic achievement, and residence location.

[3] MengjunLuo,” Research on Students’ Mental Health Based on Data Mining Algorithms”, they proposed a mental health intelligence evaluation system based on a joint optimization algorithm is

proposed to address the problems of high misjudgment rate and low work efficiency in the current mental health intelligence evaluation process.

[4] A. Velinov and C. MartinovskaBande, titled," Classification with ID3 and SMO using Weka", They use ID3 and SMO algorithm to classify dataset on wine quality. SMO has more classified instances than ID3, but ID3 has fewer misclassified instances, based on the results incurred by them.

[5] Rita Rezaee, Reza Akbari, MiladNasiri, FarzanehForoughinia, NasrinShokrpour, titled", An evaluation of classification algorithms for prediction of drug interactions: Identification of the best algorithm", individuals were interested in discovering drug - related problems in biological sources as well as categories and identify the most effective algorithm. They used a two-stage categorization to tackle the problem of unseen data scattering in adverse drug events categories..

[6] Satish Kumar David, Amr T.M. Saeb, Khalid Al Rubeaan, Titled," Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics", In their study, they offer additional comprehensive details regarding data mining algorithms, with such a concentrate on classifiers as one crucial method for supervised learning.

Another major step of this model is to facilitate the parents as well because most of the students are not able to share their study status and results with their parents. Parent/Guardian phase provides facility to check child profile like marks attendance, class schedule, exam plan and fee status. If parents are not able to check their child's profile, they can take parental help to check their child's status. We wish that this the proposed model helps our society to improve its education system.

Problem statement:

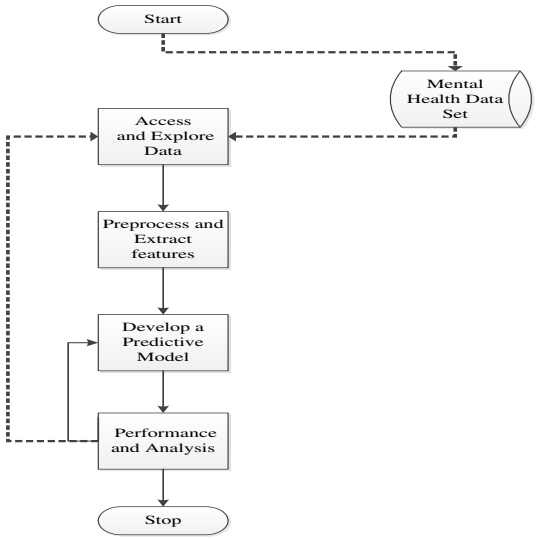
Using Weka to perform data analytics and predict whether a student should see a doctor for mental health issues based on instances of available datasets.

Methodology:

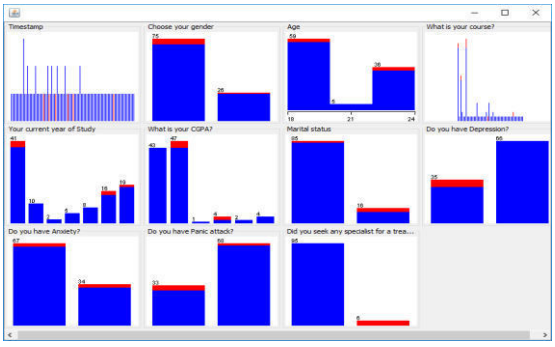
Many techniques have been proposed by research, but picking the right approach is crucial in classing the instances.

3.0 Explore and Preprocess data:

The data set for this study is mentalhealth.csv, which is open source in nature (CC01) and can be downloaded from kaggle.com.This dataset looks into students' empathy, psychological disorders, and exhaustion in Switzerland. It assembles crucial demographic details, along with inner statistics and psychological lab results, to offer a detailed picture of students' psychological disorders.Throughout this studies, we hope to improve our comprehension of how being a student affects health and well-being.The data set has 11 attributes, one of which is age which is numeric,and the rest are nominal or categorical.



Figure#2: Methodology of proposed work

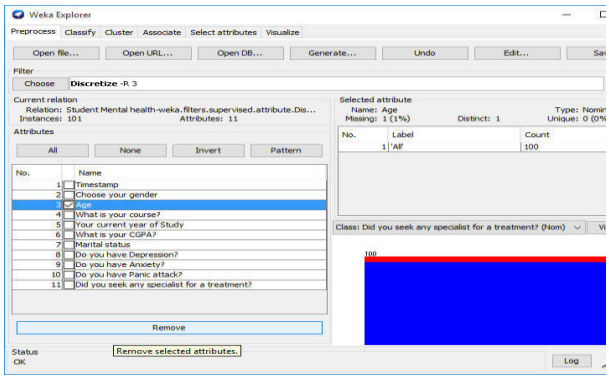


Figure#3: Visualization of Attributes

3.1 Preprocess and Extract features:

The data set recognized student mental health. The csv data set has been loaded into Weka, and 9 features have been assessed from 101 extracts.While loading data into Weka, a summary of all data values is collected from the data source.We have far more data on females than on males. As a result, this data will be skewed towards women.We can see that student from the first year participated the most, with 43 students from the first year, 26 students from the second year, 24 students from the third year, and 8 students from the fourth year. The data we have is skewed towards the first year of college. [11]

To preprocess data we use Weka→ filter→ supervised→ attribute → discretize, we choose an age attribute as the attribute index and apply supervised instance filter to discretize a range of numerical values into nominals or categorical.



Figure#4: Preprocessing of data

Feature Extraction:To extract features, we use the best first method, in which no attribute is initially selected, the search direction is forward, only five nodes are expanded, and a total of 58 subsets are evaluated.Ascribes of features extracted:What is your grade point average?Do you have a marital status? Do you suffer from depression?

According to the extracted data, nearly half of the students are depressed.Female students seem to be depressed much as male students.There are 16 wedded students, and they're all depressed.

The vast bulk of depressive disorders are characterized by a mix of depression and anxiety. As per studies, 60-70% of individuals with depression also feel anxious. Moreover, quarter of those who have persistent nervousness have clinically relevant depression.

IV. EXPERIMENTAL ANALYSIS

4.1.Training and testing set:e use a proportion of total split wherein 70% of the data serves as training data and 30% of the records is assessed as test or unfamiliar records.

4.2.Experimental Observations:

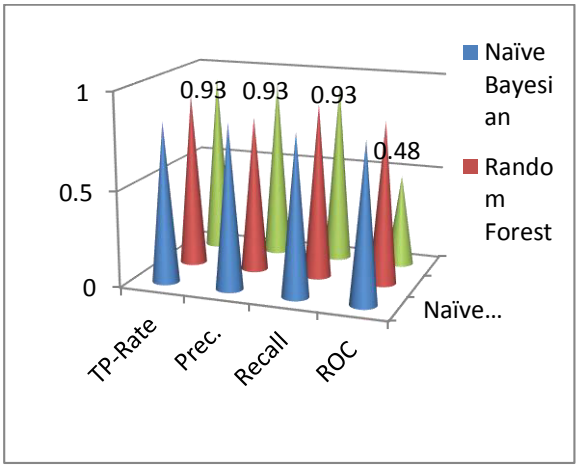
The classifier model is built to develop a predictive model, and various metrics such as true positive rate, precision, recall, and accuracy are employed to properly assess it. ROC analysis was also carried out. [12]

Following Table Shows details:

Algorithm	TP-Rate	Prec.	Recall	ROC
Naïve Bayesian	0.83	0.85	0.83	0.82
Random Forest	0.9	0.81	0.9	0.84
k-NN	0.93	0.93	0.93	0.48

Table#1: Performance on various metrics

We can see from the table above that there are differences among each of the performance measures.

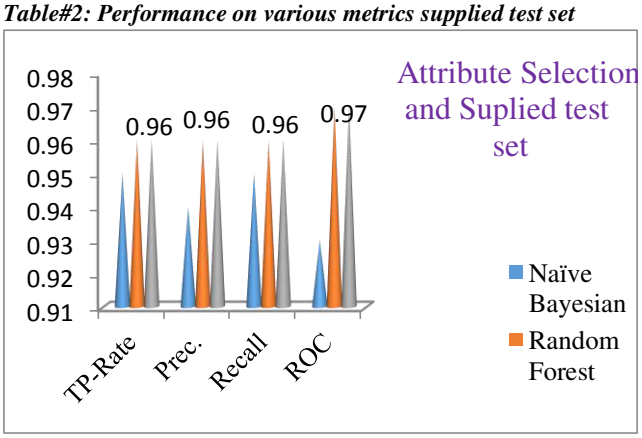


Figure#5: Graphical representation of performance measures

The experimental observations demonstrate what follows after selecting attributes and performing experiments as supplied test set as entire reduced set:

Algorithm	TP-Rate	Prec.	Recall	ROC
Naïve Bayesian	0.95	0.94	0.95	0.93
RF Algorithm	0.96	0.96	0.96	0.97
k-Nearest Neighbor	0.96	0.96	0.96	0.97

Table#2: Performance on various metrics supplied test set



Figure#6: Graphical view of performance measures on supplied test set

The following observations result from attribute selection and the use of a smaller data set for model evaluation.

1. Panic attacks affecting an estimated 37.5% of engineering students.
2. Panic attacks reach nearly 62.5% of IT students.
3. Panic attacks affect nearly 18% of BCS students.
4. Other than those over the age of 24, neither year 4 student academic experience depression, anxiety, or panic attacks.
5. Marital status is strongly linked to depression.
6. There is a significant relation between depression, anxiety, and panic attacks.[15]
7. There is a limiting condition between medical assistant and marital status.

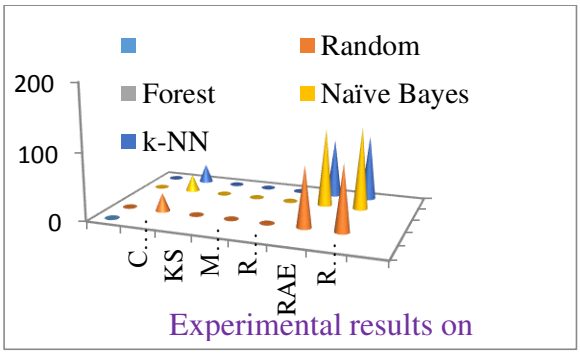
V. EXPERIMENTAL RESULTS AND DISCUSSION:

5.1 Results:

In this investigation, we explore mental health data and attempt to figure out whether or not a student requires medical assistance based on a variety of factors. [16] Then we construct three models based on it. Their outcomes are shown in the tables below:

Classifier	Model Building Time	C.C. instances	I.C.C. Instances	KS	MAE	RMSE	RAE	RRAE
Random Forest	0.19	27	3	0	0.12	0.29	90.03	98.12
Naïve Bayes	0.02	25	3	0.19	0.16	0.37	117.76	123.86
k-NN	0	28	2	0.47	0.12	0.29	87.4	98.14

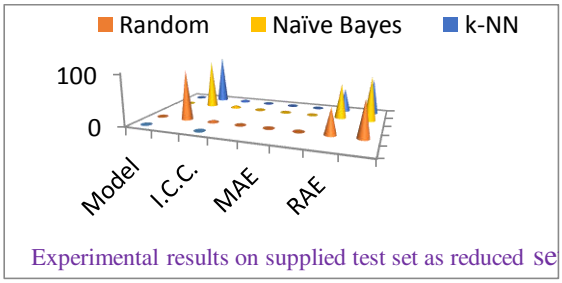
Table#3: Model evaluation based on test set



Figure#7 Graphical view of experimental results on test set.

Classifier	Model Building time	C.C. instances	I.C.C. Instances	KS	MAE	RMSE	RAE	RRAE
	e							
Random Forest	0.11	97	4	0.48	0.06	0.16	50.47	70.04
Naïve Bayes	0	96	5	0.42	0.8	0.2	67.96	86.32
k-NN	0	97	4	0.48	0.05	0.16	45.09	68.7

Table#4: Model evaluation based on supplied test as reduced set.

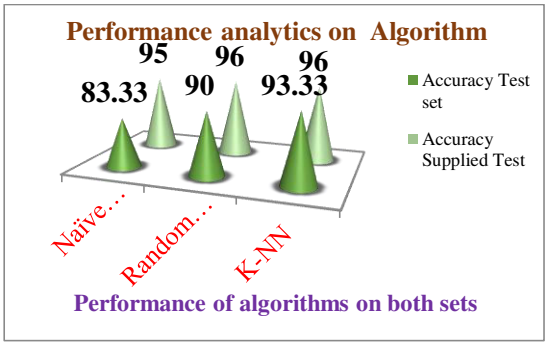


Figure#8: Graphical view of experimental results on supplied test set.

The following table illustrates the performance analytics of each classification across both the test and reduced sets:

Algorithm	Accuracy Test set	Accuracy Supplied Test
Naïve Bayes	83.33	95.00
RF Algorithm	90.00	96.00
K-Nearest Neighbor	93.33	96.00

Table#5: Performance analysis of algorithms



Figure#9: Graphical View of performance analysis

VI. DISCUSSION

Each algorithm evaluates a total of 30 instances. Although the Naive Bayes algorithm is robust, the random forest and k-NN algorithms outperform it. They have a low MAE and RMSE.

In the supplied test set, we consider all instances of the original data set while focusing on only a few attributes.

There has been a slight improvement in performance error rates, with only minor variations in performance metrics.

VII. CONCLUSION

The investigation of mental health conditions is entirely based on various characteristics. The primary goal of this study is to probe these conditions. According to the findings, age, depression, anxiety, and marital status all have a serious influence on their exam grades. The findings from experiments indicate that the k-NN algorithm outperforms the other two algorithms on both datasets. When the data is sufficient and complex, the random RF algorithm could also be a viable solution.In the future, a broad extent is given to information gathering, removing features that are irrelevant, adding characteristics that contribute to our cause, and carefully choosing data so that large amounts of data can be explored and analyzed.

REFERENCES

1. Barhate S. M. and Narale S., "Cloud Based Teaching and Learning Environment for Smart Education," International Journal on Recent and Innovation Trends in Computing and Communication vol. 3, no. 2, pp. 38-41,2015.
2. Saini L., Joyti and Kaur H., "Role of Cloud Computing in Education System," International Journal of Advanced Research in Computer Science vol. 8, no. 4, pp. 345-347,2017.
3. Mathew S., "Implementation of Cloud Computing in Education - A Revolution," International Journal of Computer Theory and Engineering vol. 4, no. 3, pp. 473-475,2012.
4. Rao K. S. and Challa R. K., "Adoption of Cloud Computing in Education and Learning," International Journal of Advanced Research in Computer and Communication Engineering vol. 2, no. 10, pp. 4160-4163,2013.
5. Viswanath D. K., Kusuma S. and Gupta S. K., "Cloud Computing Issues and Benefits Modern Education," Global Journal of Computer Science and Technology vol. 12, no. 10, pp. 15-19, 2012
6. aspect, "The Benefits of Cloud Computing in Education are Huge," 2013. [Online], Available: https://blogs.aspect.com/the- benefits-of-cloud-computing-in-education-are-huge. [Accessed: 30-November-2018].

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-25

Smart Homes Using IOT

Prof. Megha R. Chaudhari
Bhusawal Art's science and P.O.Nahata commerce college Bhusawal
meghac1912@gmail.com

Prof. Tejal K.Warake
Bhusawal Art's science and P.O.Nahata commerce college Bhusawal
tejalwarake@gmail.com

ABSTRACT - *The paper "Smart Homes Using IoT" explores the concept of using Internet of Things (IoT) technologies to create smart homes that are more energy-efficient, secure, and comfortable. The paper first provides an overview of IoT and its applications in various domains. Then, it discusses the key components of a smart home system, including sensors, actuators, and controllers, and how they work together to automate various aspects of home management. The paper also examines the benefits of smart homes, such as reducing energy consumption and enhancing security, as well as the potential challenges and risks associated with IoT-based systems. Finally, the paper concludes by outlining some future directions for research in the area of smart homes and IoT.*

Keywords - *Smart homes, Internet of Things (IoT), Home automation, Energy Efficiency, Security, Sensors*

I. INTRODUCTION

Smart homes using IoT technology is an emerging concept that refers to the use of Internet of Things devices, sensors, and systems to automate various aspects of home management. The goal of smart homes is to create a living environment that is more energy-efficient, secure, and convenient for residents. IoT technology enables devices to communicate with each other and with the cloud, enabling data collection, analysis, and control.

Smart homes can include a wide range of devices, such as smart thermostats, lighting systems, security cameras, and appliances. These devices can be controlled remotely through a smartphone, tablet, or computer, providing residents with greater flexibility and convenience. IoT sensors can also be used to monitor environmental conditions, such as temperature and air quality, to improve energy efficiency and promote better health.

The importance of smart homes using IoT technology is growing as people become more aware of the benefits of automation and the potential for energy savings. In addition, the aging population is looking for ways to live independently and safely in their own homes, and smart homes can provide the necessary assistance and monitoring to enable this. However, there are also concerns around privacy and security, as IoT devices can be vulnerable to cyber-attacks.

Overall, smart homes using IoT technology have the potential to revolutionize the way we live by creating a more efficient, secure, and convenient living environment. This paper aims to explore the key components of smart homes, the benefits and challenges associated with IoT-based systems, and the future directions for research in this area.

A. Definiton of Smart Homes and IoT

A smart home refers to a residence that uses Internet of Things (IoT) technologies to automate various aspects of home management, such as lighting, temperature, security, and entertainment. It involves the integration of devices, sensors, and systems to create a network that can be controlled and monitored remotely through a smartphone, tablet, or computer.

IoT refers to a network of physical devices, vehicles, buildings, and other objects that are embedded with sensors, software, and connectivity to exchange data over the internet. IoT enables devices to communicate with each other and with the cloud, enabling data collection, analysis, and control. In the context of smart homes, IoT technology allows for the integration of various devices and systems to create a more efficient and convenient living environment.

B. Importance of Smart Homes Using IoT:

There are several important benefits of smart homes using IoT technology:

1. Energy efficiency: Smart homes can significantly reduce energy consumption by automating the control of heating, cooling, lighting, and appliances. IoT sensors and data analytics can optimize energy usage and reduce waste, leading to lower utility bills and a smaller carbon footprint.

2. Enhanced security: Smart homes can be equipped with security systems that use IoT sensors and cameras to monitor and alert homeowners of any unusual activity. This can help prevent break-ins and theft, as well as provide peace of mind for homeowners.

3. Improved convenience: Smart homes can make life easier and more convenient by automating tasks such as turning lights on and off, adjusting the thermostat, and locking doors. This can save time and reduce the workload of homeowners.

4. Better health and well-being: Smart homes can include IoT-enabled health monitoring devices that track vital signs and activity levels. This can help individuals manage chronic conditions and maintain overall health.

C. Objective of the Research Paper

The objectives of a research paper on "Smart Homes Using IoT" may include:

1. To investigate the current state-of-the-art in IoT technologies and their applications in smart homes.

2. To analyze the benefits and challenges associated with IoT-based smart homes, including energy efficiency, security, convenience, health and well-being, and aging in place.

3. To identify the key components of smart homes, including sensors, devices, and systems, and their integration using IoT technology.

4. To propose solutions and recommendations for addressing the challenges and concerns associated with IoT-based smart homes.

5. To identify the future directions and opportunities for research and development in the field of smart homes using IoT technology

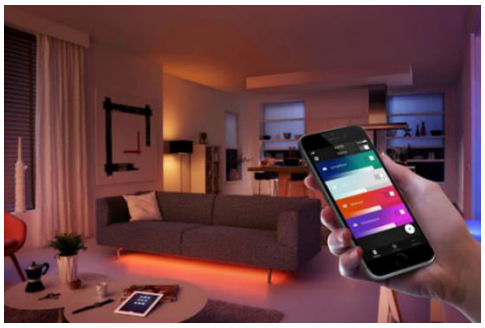
II. OVERVIEW OF IOT IN SMART HOMES

A. IoT Devices in Smart Homes

There are a wide variety of IoT devices that can be used in smart homes, ranging from basic gadgets to advanced systems. Here are some examples:

1. Smart thermostats: These devices can monitor the temperature in your home and adjust it based on your preferences. They can also learn your habits and adjust the temperature accordingly to save energy.

2. Smart lighting: These devices can be controlled remotely and can also be set to turn on or off automatically based on your preferences. Some systems can also be set to dim or change colors based on your mood or time of day.



3. Smart locks: These devices can be controlled remotely and can provide greater security than traditional locks. Some systems also allow you to give temporary access to visitors, such as housekeepers or repair personnel.

4. Smart cameras: These devices can provide surveillance and security, allowing you to monitor your home remotely. Some systems can also be set to send alerts if motion is detected.

5. Smart appliances: These devices can be controlled remotely and can also provide energy usage data, helping you to save money on your utility bills. Some systems can also be set to start and stop automatically based on your habits.

6. Smart speakers: These devices can be used to control other smart devices in your home through voice commands. They can also be used to play music or provide information.

B. IoT Technologies Used in Smart Homes

There are several IoT technologies used in smart homes that enable devices to communicate with each other and with the internet. Here are some of the main technologies:

1. **Wi-Fi:** Wi-Fi is the most common technology used to connect smart devices to the internet. It provides a high-speed connection and is widely available in most homes.
2. **Bluetooth:** Bluetooth is a short-range wireless technology that is commonly used to connect smart devices to each other. It can be used to create a mesh network of devices, allowing them to communicate with each other and with the internet.
3. **Zigbee:** Zigbee is a low-power wireless technology that is designed for use in IoT devices. It is commonly used in smart home systems to create a mesh network of devices that can communicate with each other.
4. **Z-Wave:** Z-Wave is another wireless technology that is commonly used in smart home systems. It is a low-power technology that operates on a different frequency band than Wi-Fi and Bluetooth.
5. **Cloud computing:** Cloud computing is used in many smart home systems to store and process data from IoT devices. This allows for remote access to devices and the ability to analyze data to improve device performance.

C. Benefits of IoT in Smart Homes

1. **Convenience:** One of the most significant benefits of IoT in smart homes is convenience. IoT devices can be controlled remotely, allowing you to adjust settings or monitor your home even when you are away.
2. **Energy efficiency:** IoT devices can help to make your home more energy-efficient by monitoring energy usage and adjusting settings accordingly. This can induce to lower utility bills and a reduced carbon footprint.
3. **Security:**IoT devices can provide greater security than traditional systems, allowing you to monitor your home remotely and receive alerts if anything is amiss.
4. **Comfort:**IoT devices can help to make your home more comfortable by adjusting settings like temperature and lighting based on your preferences.
5. **Accessibility:**IoT devices can make homes more accessible for people with disabilities by allowing them to control devices using voice commands or other methods.

III. COMPONENTS OF SMART HOMES USING IOT

A. Sensors And Actuators

Sensors and actuators are essential components of IoT systems in smart homes. Sensors are devices that detect

changes in the environment, such as temperature, humidity, light, motion, or sound.

Actuators, on the other hand, are devices that control or change the environment, such as opening or closing a door or turning on or off a light. Here are some examples of sensors and actuators commonly used in smart homes:

1. **Temperature sensors:** Temperature sensors can be used to monitor the temperature in different rooms and adjust heating or cooling systems accordingly.
2. **Motion sensors:** Motion sensors can be used to detect movement in a room and turn on or off lights or other devices.
3. **Light sensors:** Light sensors can be used to detect the level of natural light in a room and adjust artificial lighting accordingly.
4. **Humidity sensors:** Humidity sensors can be used to monitor the humidity level in a room and adjust air conditioning or dehumidifiers accordingly.
5. **Door sensors:** Door sensors can be used to detect whether a door is open or closed and control access to different areas of the house.
6. **Window sensors:** Window sensors can be used to detect whether a window is open or closed and adjust heating or cooling systems accordingly.
7. **Smart plugs:** Smart plugs can be used to control the power supply to devices plugged into them, allowing you to turn them on or off remotely.
8. **Smart locks:** Smart locks can be controlled remotely and can provide greater security than traditional locks.
9. **Smart blinds:** Smart blinds can be controlled remotely and can adjust the level of natural light in a room.
10. **Smart sprinklers**: Smart sprinklers can be controlled remotely and can adjust the amount of water used to irrigate plants based on weather conditions.

B. Communication Protocols

Communication protocols are a set of rules and standards that enable different devices and systems to communicate with each other in a smart home system that uses the Internet of Things (IoT). These protocols define the way data is transmitted, received, and processed between devices, ensuring that they can communicate with each other in a seamless and efficient manner.

For example, a temperature sensor may use a communication protocol to transmit data to a smart thermostat, which in turn adjusts the temperature of a room based on the sensor's readings.

There are several different communication protocols used in smart homes using IoT, including Wi-Fi, Bluetooth, Zigbee, Z-Wave, Thread. Each protocol has its own advantages and disadvantages, and different protocols may be used for different purposes within a smart home system.

By using communication protocols, smart home systems can achieve greater interoperability and flexibility, allowing different devices and systems to work together seamlessly and providing users with greater control over their homes.

C. Cloud Computing

Cloud computing in smart homes using IoT refers to the use of remote servers and software platforms to store and process data generated by various smart devices and sensors in the home. These cloud-based platforms can provide a range of services, such as data analytics, machine learning, and artificial intelligence, which can be used to analyze the data generated by smart home systems and provide insights that can help optimize energy use, improve security, and enhance the overall functionality of the home.

For example, homeowners can use a mobile app or web interface to remotely control their smart devices and sensors, view real-time data, and receive alerts and notifications about important events in the home.

One of the main benefits of using cloud computing in smart homes using IoT is scalability. Cloud-based platforms can easily scale up or down to meet the changing needs of a smart home system, providing greater flexibility and cost savings compared to traditional on-premises solutions.

Overall, cloud computing plays a crucial role in enabling smart home systems using IoT to function efficiently and effectively, providing homeowners with greater convenience, comfort, and control over their homes.

IV. APPLICATIONS OF SMART HOMES USING IOT

There are many applications of smart homes using IoT, including:

1. **Home Security:** Smart home systems can include security cameras, door and window sensors, and motion detectors that can be monitored and controlled remotely. These systems can provide real-time alerts and notifications to homeowners and help deter intruders.

2. **Energy Efficiency:** Smart home systems can include smart thermostats, smart lighting, and smart appliances that can be controlled remotely and optimized for energy efficiency. This can help homeowners save money on their energy bills and reduce their carbon footprint.
3. **Home Automation:** Smart home systems can automate various tasks such as turning on and off lights, adjusting the thermostat, and opening and closing blinds. This can provide greater convenience and comfort to homeowners.
4. **Health and Wellness:** Smart home systems can include health monitoring devices such as fitness trackers and smart scales, as well as devices that can help elderly or disabled individuals with daily tasks.
5. **Entertainment:** Smart home systems can include smart TVs, audio systems, and streaming devices that can be controlled remotely and integrated with other devices in the home.

V. CHALLENGES OF SMART HOMES USING IOT

While smart homes using IoT offer many potential benefits, there are also several challenges associated with their implementation. Some of the key challenges include:

1. **Security and Privacy:** Smart home systems are vulnerable to cyber attacks, which can compromise the security and privacy of the homeowners. This is especially concerning as smart home systems may include sensitive information such as video and audio recordings. It is important to implement strong security measures to protect against such attacks.
2. **Interoperability:** There are many different smart devices and platforms available on the market, and they may not always be interoperable with each other. This can lead to compatibility issues and difficulty in integrating different devices into a cohesive system.
3. **Complexity:** Smart home systems can be complex and difficult to set up and manage. They may require technical expertise and specialized knowledge, which can be a barrier for some homeowners.
4. **Cost:** Smart home systems can be expensive, especially if they involve the installation of new devices or appliances. This can be a barrier for some homeowners who may not be able to afford the upfront costs of implementing a smart home system.
5. **Energy Consumption:** While smart home systems can be designed to optimize energy consumption, the devices themselves may consume energy even when not in use. This can lead to increased energy consumption and may reduce the overall energy savings that can be achieved.

VI. FUTURE OF SMART HOMES USING IOT

A. Trends in Smart Homes Using IoT

There are several key trends in smart homes using IoT that are shaping the way these systems are designed, implemented, and used. Some of these trends include:

1. **Voice Control:** The use of voice assistants such as Amazon's Alexa and Google Assistant is becoming increasingly popular, as they provide a convenient and intuitive way for homeowners to control their smart home devices using voice commands.
2. **Artificial Intelligence:** The integration of artificial intelligence (AI) into smart home systems is expected to become more prevalent, allowing for more personalized and intelligent automation of various tasks. This will enable smart home systems to learn from the behavior of the homeowners and adjust accordingly.
3. **Energy Management:** Energy management will continue to be an important application of smart homes using IoT, as consumers become more conscious of their energy consumption and seek to reduce their carbon footprint. Smart home systems will be designed to optimize energy consumption and integrate with renewable energy sources.
4. **Integration with Wearable Devices:** Smart home systems are being designed to integrate with wearable devices such as smartwatches and fitness trackers, allowing for greater automation and control based on the health and wellness of the homeowner.

B. Future Innovations in Smart Homes Using IoT

The future of smart homes using IoT is filled with possibilities, and there are several innovations that are likely to shape the way these systems are designed and implemented in the years to come. Some of these innovations include:

1. **Advanced Sensors:** The development of advanced sensors, including sensors that can detect things like motion, temperature, and humidity, will enable smart home systems to become even more responsive to the needs of homeowners.
2. **Predictive Analytics:** The integration of predictive analytics into smart home systems will enable these systems to anticipate the needs of homeowners and automate various tasks based on historical and real-time data.
3. **Robotics:** The use of robotics in smart homes is likely to increase, enabling these systems to perform physical tasks such as cleaning and maintenance.
4. **5G Technology:** The rollout of 5G technology is expected to enable faster and more reliable connections between smart home devices, enabling even more sophisticated and responsive systems.
5. **Augmented Reality:** The use of augmented reality (AR) technology in smart homes is likely to increase, allowing homeowners to interact with their smart home systems in new and innovative ways.

VII. CONCLUSION

IOT brings a new era for IT technologies and can change our life and job to a more intelligent and modern stage. Internet of Things has numerous applications in different areas. It has been expanded for Smart Homes System. There are some issues found in IoT and Smart Homes. New technologies could help to reduce some of them. This paper presents the problems and challenges that could come. New technologies and methodologies which are already used to improve applications of IoT have been discussed in this paper.

REFERENCES

1. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., &Ayyash, M. (2015). Internet of Things: A survey on enabling technologies, protocols, and applications. IEEE Communications Surveys & Tutorials, 17(4), 2347-2376.
2. Banerjee, N., &Chakraborty, C. (2019). A review of the state-of-the-art IoT in smart home. Computers & Electrical Engineering, 78, 20-39.
3. Han, S., Kim, D., & Kim, S. (2017). Smart home energy management system including renewable energy based on internet of things. Energies, 10(2), 219.
4. Hossain, M. S., Muhammad, G., &Hossain, M. A. (2019). Internet of things (IoT) for smart homes: A review of technology, applications and challenges. Journal of Sensor and Actuator Networks, 8(1), 1-35.
5. Li, X., Li, X., Liang, X., Lu, R., Lin, X., &Shen, X. S. (2019). A survey on internet of things for smart home: Recent advances and future directions. IEEE Internet of Things Journal, 6(4), 6596-6613.
6. Ma, Y., Gao, W., Wang, Y., & Wang, Q. (2019). Smart home energy management based on internet of things and big data analytics. Journal of Ambient Intelligence and Humanized Computing, 10(11), 4499-4511.
7. Singh, J., &Pasquier, T. (2018). A survey of security in the deployment of internet of things for smart homes. ACM Computing Surveys (CSUR), 51(5), 1-36.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-26

Artificial Intelligence and Machine Learning In Robotics

Prof. Tejal K.Warake
DOCS and IT, BAS P.O.Nahata College Bhusawal,
tejalwarake@gmail.com

Prof. Megha R. Chaudhari
DOCS and IT, BAS P.O.Nahata College Bhusawal,
meghac1912@gmail.com

ABSTRACT - *Information and Communications technology (ICT) has been widely applied in digital library services to improve the knowledge and the public interest. Researchers can easily obtain the information of interest through internet surfing. Digital libraries seek to provide prospective readership with alternative materials for reading. A digital library, also called an online library, an internet library, a digital repository, or a digital collection is an online database of digital objects that can include text, still images, audio, video, digital documents, or other digital media formats or a library accessible through the internet. Objects can consist of digitized content like print or photographs, as well as originally produced digital content like word processor files or social media posts. In addition to storing content, digital libraries provide means for organizing, searching, and retrieving the content contained in the collection. As far as usage of Digital Libraries are concerned, following factors are taken into consideration viz, 24 x 7 availability, access speed, Simplicity in finding relevant information, minimal storage space, unavailability of the required reference at the time of need etc. Also in case of research students are concerned, they had taken under consideration stream wise such as Arts, Science, Commerce and Management, Engineering, Other etc*

Keywords - *Artificial intelligence, robotics, intellectual.*

I. INTRODUCTION

Artificial Intelligence and Robotics have a ordinary root and a long history of communication and scientific conversation. The beginning of Artificial intelligence and Robotics takes place in the same period, and at first there was no clear distinction between the two disciplines. The field of robotics is intimately linked to that of AI, although definitional issues abound. Giving AI motor capability" seems a reasonable definition, but most people would not regard a cruise bullet as a robot even though the steering and control techniques draw heavily on robotics research[2]. AI and robotics are possible to continue to creep into our lives without us really noticing. Unfortunately, many of the applications appear to be taking place among agencies, mainly the military that do not readily respond to public concern, however well-articulated or thought.

II. AI IN ROBOTICS

AI is uncertainly the most exciting field in robotics. It's certainly the most controversial: everyone agrees that a robot can work in a meeting line, but there's no consensus on whether a robot can ever be intelligent. Similar to the word "robot" itself, artificial intelligence is hard to define. Ultimate AI would be an activity of the human thought process a human-made machine with our intellectual abilities. This would contain the ability to learn just about anything, the ability to reason, the ability to use language and the ability to formulate original ideas. Roboticists are nowhere close to attaining this level of artificial intelligence, but they have made a lot of progress with more limited AI. Now a days AI machines can replicate some specific elements of intellectual ability.

Computers can previously solve problems in limited realms. The vital idea of AI problem-solving is simple, though its execution is complicated. First, the

AI robot or computer collects facts about a situation through sensors or human input. The computer links this info to stored data and decides what the information signifies. The computer runs through numerous possible actions and predicts which action will be most successful based on the collected information. For the greatest part, the computer can only solve problems it's programmed to solve.it doesn't have any generalized analytical ability. Chess computers are one example of this sort of machine.AI in robotics helps robots make the critical tasks with a human-like vision to detector recognize the various objects. A vast amount of datasets is used to train the computer vision model, so that robotics can know the various objects and carry out the actions accordingly with right results.

III. INTERACTION WITH OTHER AI FIELDS

As currently mentioned, the research on AI Robotics interconnects a number of subfields of AI. Indeed, the robotic mediator can be seen as a main target for the grand goal of Artificial Intelligence, and thus for all the aspects of AI slightly related to Robotics. Below, we address the main networks with the other AI research topics included in this collection.

Machine Learning

Learning methods are being applied to many problems rising in the design of robots. According to the assembly adopted above, both action and perception can be supported by learning approaches. Moreover, some approaches that include a training step are pursued ranging from machine learning approaches to genetic programming, and neural networks. Since the standpoint of action, learning approaches can be used for the basic action skills, specifically locomotion, but also learning cooperative behaviours, version to the environment, and learning opponents' behavior, among others. Perceptibly, the learning process must face the challenges of the experiments with real robots. Nevertheless, in several investigational situations learning and adaptation of the basic skill, such as walking, vision calibration, have shown to be much more effective than parameter tuning by hand.

Edutainment

Toy robots are very auspicious to be used both for research purposes and for education, because of low costs and high attraction for students. Even though, at this miniature, the available educational kits seem to provide too limited capabilities, toy robots are certainly an interesting commercial market. Accordingly, the design of smart toy robots is a motivating opportunity for AI researchers [3]. The information with Aibo robots shows this prospective: they have been successfully used by many research groups in the world not only in the RoboCup competitions (Four-Legged League), but also for representative other AI and Robotics research issues.

Multi agent systems

A multi-robot structure (MRS) can be stately as a multi-agent system (MAS), but the techniques for achieving coordination and cooperation in MAS are often not well suited to deal with the ambiguity and model incompleteness that are typical of Robotics. Multiple robots may attain more robust and more effective behavior by accomplishing coordinated tasks that are not possible for single robots. Groups of similar and heterogeneous robots have a great potential for application in complex domains that may require the intelligent use and merge of diverse capabilities. The strategy, implementation, and assessment of robots organized as teams pose a variety of scientific and technical challenges.

Natural Language Processing

It is an obvious state of home and service robotics the capability to relate with people in natural language; therefore, natural language processing methods find an interesting application domain on robots.

Logics for AI and Automated Reasoning

The construction to the Logics for AI and Automated Reasoning is central to the work on Cognitive

Robotics, but we do not auxiliary expand it here, as it is discussed in the previous section.

Evolutionary Computation and Genetic Programming

Evolutionary Robotics is a new method that looks at robots as autonomous artificial organisms that develop their own skills in close interaction with the situation without human intervention. Evolutionary robotics thus applies methods coming from evolutionary computation.

IV. INTERACTION WITH OTHER DISCIPLINES

Robotics is a multidisciplinary field: to create an operational robot, some contributions from many corrections are needed: physics, electrical engineering, electronic manufacturing, mechanical engineering, computer science, AI, and so on. It is therefore difficult also to have a common background of terms, codes and methodologies. In this logic, the energies to define a common ontology of terms for robotics science are noteworthy [4]. In specific, AI Robotics interacts with numerous research disciplines outside AI.

Industrial Robotics

Many connection points may be found amongst AI, Robotics and Engineering Robotics. In early days there were not clear and cut distinctions between the two fields, as already mentioned. Today, study in Industrial Robotics is oriented towards the safe and intelligent control of industrial manipulators and in the field of service robotics. The methodologies in Industrial Robotics are grounded in Programmed Control Theory[5] . The connection between the robot and the environment is generally modeled by means of several types of feedback systems. Moreover, organizations are typically based on numerical methods and optimization theory.

Computer Vision

Robot Vision is exact with respect to computer vision, because Robot Vision is basically active, in the sense that the robot may actively find its information sources and it can also influence the best view position to maximize the visual information. Moreover, Robot Vision must be achieved in real-time, because the robot must immediately react to visual stimuli. In general, the robot cannot procedure for a long time the same image because the environmental conditions may vary, so the robot has to deal with estimated, but just in time information. Numerous research topics and debates in this field have strong correlations with AI and Robotics, for example, if a Computer Vision system may be based on inner representation of the atmosphere or it should be purely reactive.

Mechatronics

Mechatronics includes competencies from electrical engineering, electronic engineering, and mechanical engineering. All of these capabilities are strictly connected to AI and Robotics: the research field of electrical engineering concerns motors and actuators, while electronic engineering mainly concerns boards for robot control, for data procurement and in general for the hardware that marks the robot operational. From this fact of opinion, Mechatronics, AI and Robotics have tight relations: Mechatronics generally focuses on the robot hardware at all stages, while AI and Robotics take care of the software that sorts the robot operative and autonomous.

Embedded Systems

The robot software system necessities to work in real time in order to assurance that the robot correctly copes with the varying environment; it must be failsafe with elegant degradation in order to guarantee that the robot may operate also in case of damages; the hardware system of the robot must be low power aimed to optimize the batteries, and so on. From this point of view, numerous of the typical assessments of embedded systems are also challenges for robotics systems.

V. APPLICATIONS

Rescue Robotics

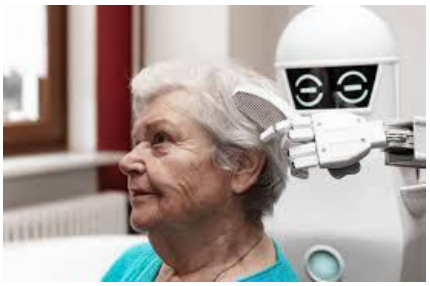
Further soccer, RoboCup promotes other leagues, aiming at the transport of the research results into within society and technologically relevant contexts. Particularly, RoboCup Rescue [6] aims at the design of systems to search and rescue for large scale disasters. Here we focus on the saving robot league,that aims at the strategy of robots searching victims in an unknown environment representing a disaster scenario. This type of function brings in scientific challenges, related to the ambiguity about the environment, that are not present in the soccer leagues.

Space Robotics

The plan of the project An Intelligent System for the Supervision of Autonomous Robots in Space, funded by the Italian Space Agency (ASI) during years 1997-2000, is the application of AI techniques to the design and realization of space robotics systems for planetary exploration missions that need an increasing autonomy. The scientific objective of the project is the design and development of an intelligent system able to supervise self-directed robots in space. The system is based on a multi agent planning in which each block a software agent is interfaced with the rest of the system.

Robotics for Elderly and Impaired People

The matter of elder care is one of supply and demand. As the elder inhabitants increases in numbers, the number of caregivers is not correspondingly increasing. Japan foresees a shortage of 1 million caregivers by the year 2025. And the United States predicts that the percentage of people aged 65 and older is expected to increase to roughly 26% of the population by 2050. The virtual lack of people able to care for the elderly is what keeps the costs of elderly care so high and creates a weigh down on family and caregivers.



There are many ways that medical robotics can help the elderly:

- Robots can act upon small tasks like fetching food and water.
- Some senior care robots knob social and emotional needs by providing entertainment through games, helping remind them of proceedings and actions, and providing social engagement [8].
- Other eldercare robots take a extra direct, muscular move toward and use powerful hydraulics to help provide mobility and shipping support to seniors.

VI. CONCLUSION

This paper begin by stressing the required to provide background information on AI. In doing so, it was hoped that the prospects of these emerging technologies to affect quality of life in the awaiting decades could be realistically assessed.

Robotic skill will definitely grow and include a huge range of surgical procedures, particularly in urology [7]. India should not disregard the robotic revolution and should wholeheartedly imbibe the future making of technological expertise. Robotic surgical procedure has come to stay and will make slow but steady inroads into Indian healthcare delivery system. Fitness is a human right, which has also been accepted in the Constitution. Its convenience and affordability has to be insured.

REFERENCES

1. Kumar,Yuvraj, Artificial Intelligence & Robotics – Synthetic Brain in Action (September 14, 2018).
2. https://www.automate.org/blogs/the-future-of-elder-care-is-service-robots
3. M. Veloso, W. Uther, M. Fujita, M. Asada, and H. Kitano. Playing soccer with legged robots. In Proceedings of IROS-98, Intelligent Robots and Systems Conference, Victoria, Canada, October 1998.
4. J. Hallam and H. Bruyninckx. An ontology of robotics science. In H.I. Christensen, editor, European Robotics Symposi2006, pages 1–14, Berlin, Heidelberg, 2006. Springer-Verlag.
5. L. Sciavicco and B. Siciliano. Modelling and Control of Robot Manipulators, 2nd ed. Springer-Verlag, Berlin Heidelberg, 2000.
6. H. Kitano et al. Robocup-rescue: Search and rescue for large scale disasters as a domain for multiagent research. In Proceedings of IEEE Conference on Man, Systems, and Cybernetics(SMC-99), 1999.
7. https://www.researchgate.net/publication/220672741_Artificial_Intelligence_and_Robotics
8. https://medium.com/vsinghbisen/ai-in-robotics-use-of-artificial-intelligence-in-robotics-726a4e9ade18

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-27

"Real time Face Mask Detection and Mental Stress Detection using Facial Expression"

Mr. Dipak Y. Wani
M.Sc. II (Computer Science),
Dept. of CS & IT, BASPONC College, Bhusawal
dipakwani2001@gmail.com

Prof. Poonam M. Mahajan
Asst. Professor,
Dept. of CS & IT, BASPONC College, Bhusawal
amritamah@gmail.com

ABSTRACT - *In this paper, we developed a system that hold down the growth of COVID- 19 by finding out people who aren't wearing any facial mask in a smart town network where all the public sites are covered with CCTV cameras and mental stress finding using facial expressions can also be salutary in relating person should have internal stress or not. A face mask discovery is a fashion to find out whether someone is wearing a mask or not. Real time recognition system that traces the veritably mood of the human. Human expresses their mood and occasionally what they need through their expression. It can be a smiling face, or it can be the face full of angriness. Occasionally words aren't that important as our expressions.*

It consists of models made through colorful algorithms of machine as well as deep literacy. It also uses some of the authentically strong packages in python to produce an operation software that recognizes the expression of human in real time. Some of the usual python libraries are TensorFlow, Keres, OpenCV, Face Net, Matplotlib, imutils.

Keywords - *Face Net, Haar Cascade Classifier, Keras, Modal, NumPy, OpenCV, TensorFlow.*

I. INTRODUCTION

The COVID-19 pandemic has made it mandatory for people to wear masks to prevent the spread of the virus. The objective of this research paper is to develop a real-time face mask detection and mental stress detection system using facial expressions. The system will use machine learning algorithms to detect whether individuals are wearing masks or not and will also identify facial expressions that indicate mental stress.

❖ **Real Time Face Mask Detection:**

People wear face masks once they step out of their homes and authorities strictly ensure that people are wearing face masks while they are in groups and public places. Face mask detection involves detecting the location of the face in the real world and then determining whether it has a mask on it or not and informing to keeping touch. This system is based on Machine Learning (ML) packages such as TensorFlow, Keras, OpenCV. Real time system to detect whether the person on the webcam is wearing a mask or not. Real time face masks can and are being used in all hospitals and industries. Haar cascade mainly works with face detection. The algorithm requires a lot of training datasets the image which contain faces as positive dataset and images without faces as negative datasets.

Fig 1.1 Real time face mask detection

❖ **Real Time Mental Stress Detection:**

I am proposing an application that is capable of determining human emotion. By emotion detection or

facial expression, we can find out whether he has mental stress or not. Explanation: If he/ she is happy, they have no mental stress and if he/ she is sad then they have mental stress. Can use for those who take more tension & can use for hospitals & use in home also. Haar cascade classifier will be used for face detection & for facial Expression or Emotion detection and detect person have mental stress or not by their facial expression. The system can be used to identify individuals who may require assistance due to mental stress.

Fig 1.2 Real time mental stress detection

II. LITERATURE SURVEY

In recent years and with the rapid growth of technology, object detection, and image processing are manipulating and detecting using different machine learning and deep learning techniques. COVID-19 Face mask detection model discussed here, the help of machine learning and image processing in order to identify the people wearing a mask or not wearing a mask. This work's experimental result is done using Open CV to detecting live stream objects.

Open-Source Computer Vision System (OpenCV) image processing techniques with the help of Haar classifier Cascade algorithm for face detection A high-speed image preprocessing method was proposed in to build a security system in sensitive areas. It is very complicated to detect the emotions and distinguish among them. Before a decade or two emotions started to become a concern as an important addition towards the modern technology world. With the special kind of calculation then that machine could predict the mental stress by their mood and by which mankind could avoid serious circumstances and lot more.

III. REQUIREMENT ANALYSES

Here we have need to some pretrained python libraries for face mask and facial emotion detection they are,

1. OpenCV: -

OpenCV is the huge open-source library for the computer vision, machine learning, and image processing To Identify image pattern and its various features we use vector space and perform mathematical operations on these features. OpenCV provides a powerful set of tools for image and video processing and detect live stream objects. OpenCV Python library that allows you to perform image processing and computer vision tasks. .

2. TensorFlow:

Tensor Flow It is intended to ease the use and broadly relevant to both numeric and neural system issues just as different spaces, Tensor Flow is a low-level tool for doing entangled math and it targets specialists who recognize what they're doing to construct exploratory learning structures.

3. Keras:

Keras is a python library which is widely used for training deep learning models. In this article, we will

see the list of popular datasets which are already incorporated in the keras datasets module.

4. FaceNet Model:

FaceNet is a face recognition method created by Google researchers and the open-source Python library that implements it. FaceNet takes an image of the person's face as input and outputs a vector of 128 numbers which represent the most important features of a face. FaceNet is a face recognition pipeline that learns mapping from faces to a position in a multidimensional space where the distance between points directly correspond to a measure of face similarity.

5. Haar Cascade Classifier:

Haar feature is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. It detects input image and apply classification on it.

IV. PROPOSED METHODOLOGY

This particular research is achieved in two phases, where phase 1 would be to train the face mask detector and phase 2 would be to apply or deploy the face mask detection Software. In phase 1 of the project, our main focus would be to create a training model using datasets, where datasets consist of a huge number of images of people with face masks and

people without a face mask. The system will be developed using Python and the OpenCV library. The dataset used for training and testing the system will comprise of facial images of individuals wearing and not wearing masks, as well as facial expressions indicating mental stress. To detect face masks, the system will use the Haar Cascade Classifier, a machine learning-based approach for object detection in images and videos. The Haar Cascade Classifier works by detecting specific features of an object, such as the edges and corners of the object, in an image or video.

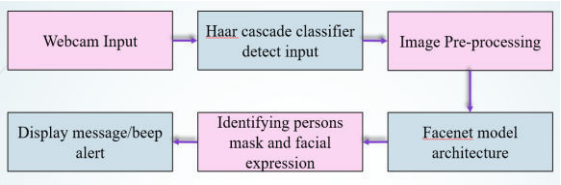


Fig 1.3 proposed working Methodology

To detect mental stress, the system will use the Face Net facial expression recognition module. Face Net is a pre-trained CNN which embeds the input image into a 128-dimensional vector encoding. It is trained on several images of the face of different people. Face Net: A Unified Embedding for Face Recognition and Clustering.

V. ARCHITECTURE OF FACE MASK AND MENTAL STRESS DETECTION

Well-established we apply the face mask CLASSIFIER that is the Facenet Model on the newly taken face images to determine whether it is having a mask or no mask. After Face Detection Completed then will start the process of mental stress detection. The distance and angle between these points calculates the category of emotion and gives the percentage (ranges from 0 to 100) of facial expression on the user's face.

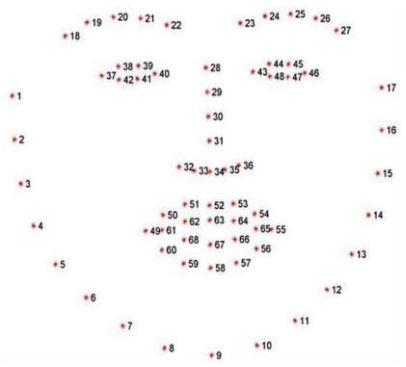


Fig 1.4 Valance Point Detection

The valance points on the user's face correspond to the array of locations on the face. The distance and angle between these points calculates the category of emotion and gives the percentage (ranges from 0 to 100) of facial expression on the user's face.

VI. IMPLEMENTATION

Step 1: Data Pre-processing

The dataset we are using consists of images with different colours, different sizes, and different orientations. Therefore, we need to convert all the images into grayscale because we need to be sure that colour should not be a critical point for detecting mask.

Step 2: Apply Face Net Model

Face Net is a pre-trained CNN which embeds the input image into a 128-dimensional vector encoding. It is trained on several images of the face of different people. Face Net: A Unified Embedding for Face Recognition and Clustering.

Step 3: Feature Extraction:

Facial feature extraction is the method of extracting facial features such as eyes, nose mouth, The extraction of a facial feature is very important in the implementation of processing techniques

Step 4: Emotion Detection

Emotion Detection Emotions and expressions are calculated using valance points which will be placed around the face. The distance and angle between these points calculate the category of emotion and gives the percentage (ranges from 0 to 100) official expression on the user's face.

VII. APPLICATION

- This implementation can be used at various places and platforms. The very first example can be feedback through moods at any restaurants and hotels about their services and foods. It can be much impactful in the field of military.
- Face mask detection refers to detect whether a person is wearing a mask or not. This implementation can be used in various locations and platforms.
- Improved security: Face detection improves surveillance efforts and helps track down criminals and terrorists. It can be very effective in the field of military.

- Facial recognition is a technology that can benefit society, including increasing safety and security, preventing crimes, and reducing human interaction.
- Its very usage can be helpful for recognizing the people's behaviour at the border areas to find out the suspects between them.

VIII. FUTURE ENHANCEMENT

The future of real-time face mask detection and mental stress detection using facial expressions looks promising, as technology continues to advance and improve. Here are some potential developments that could shape the future of these technologies:

- **Improved accuracy:** Machine learning algorithms and deep learning models could be used to improve the accuracy of real-time face mask detection systems.
- **Increased accessibility:** As the technology becomes more advanced, it is likely that real-time face mask detection and mental stress detection systems will become more widely available and affordable.
- **Integration with other technologies:** Real-time face mask detection technology could be integrated with other technologies, such as facial recognition or temperature scanning, to provide a more comprehensive assessment of an individual's health status.

IX. CONCLUSION

As the technology is growing with surging trends, we have a novel face mask detector which can possibly contribute to the public health care department. The proposed system is expected to achieve high accuracy in real-time face mask detection and mental stress detection using facial expressions. OpenCV provides a powerful set of tools for image and video processing, which can be leveraged to develop real-time face mask detection and mental stress detection systems. Further research can be done to improve the accuracy of the system and to develop more advanced features, such as detecting social distancing violations he/she has mental stress or not he can find out by his emotion. It can be very effective in the field of Military and for security purpose about terrorism in public area.

The work opens interesting future indications for researchers. Firstly, the proposed technique can be integrated into any high-resolution video observation devices and not limited to mask detection only. Secondly, the model can be expanded to detect facial landmarks with a facemask for biometric purposes. For the future, this app can be developed so that it is compatible on multiple platforms. The emotion can be used as an input in order to achieve a variety of outputs based on the requirements.

REFERENCES

1. B. Qin and D. LI, "Identifying Facemask wearing Condition Using Image Super resolution with Classification Network to Prevent COVID-19", Research Square platform LLC, May 2020.
2. S. Gupta, "Facial emotion recognition in real time and static images," Jan. 2018.
3. C. Li, Y. Li, J. Chen, and J. Lu, "Facial expression recognition based on deep learning," in Proceedings of the 9th International Conference on Intelligent Computing and Applications (ICICA), 2020, pp. 22-30.
4. https://pyimagesearch.com/2020/05/04/covid-19-face-mask-detector-with-opencvkerastensorflow-and-deep-learning/
5. Cheng, Y., Chen, H., & Zhou, X. (2021). Real-time mask detection using deep learning for COVID-19 prevention. IEEE Access, 9, 33482-33489.
6. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 815-823.
7. https://opencv.org/course-deep-learningwith tensor flow-and-keras/
8. https://en.wikipedia.org/wiki/Facial_recognition_system

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-28

An Overview Of Storage Area Network

Manasi R.Ingale,
Bhusawal Art's science and P.O.Nahata commerce college Bhusawal
manasiingale22@gmail.com

ABSTRACT - *As a network administrator we all be asked to connect many different kinds of storage system to network .One of these systems might be a Storage Area Network or a SAN To the users, a SAN appears like a local drive so it acts-it works very similar to hard drive or an SSD that you might have in our system.*

A SAN generally operates as a block level access .This means if we have a very large files, and we need to change one small piece inside of that file, we don't have to overwrite the entire file to the SAN instead, we would only change the small block inside of that file to be able to make that change That means it's very efficient when we are reading or writing information from SAN. But of course, unlike the local hard drive or SSD that may be in our local computer, we have to access this SAN across the network .This mean there will be additional network bandwidth required and it's not unusual to see a SAN that would have its own dedicated network, so that we have the best possible throughout to the SAN.

Keywords - *Storage area network (SAN), Communication, Fibre channel, Host, RAID, iSCSI, Communication, Security.*

I. INTRODUCTION

An storage area is an interconnection of devices like disks,tapes and servers etc that are connected for data transfer and common communication .As a network administrator we all be asked to connect many different kinds of storage system to storage .Storage Area Network(SAN) is a specialized ,high speed , high availability network that uses fiber channel technology to connect servers to storage disks . We can Say that SAN is a type of method by using which we can deliver the storage service to the clients or customers. In cloud computing ,SAN provides a centralized storage pool that can be accessed by multiple servers ,allowing for more efficient use of storage resources .With a SAN, storage can be allocated and reallocated to server as needed, and data can be replicated and backend up more easily .A SAN is a network of storage devices that can be accessed by multiple servers or computers, providing a shared pool of storage space. SAN is a dedicated network for storage that's the reason naming convention says everything storage area network so it's completely a different isolated network exclusively for storage .Each computer on a network can access storage on the SAN as though they were local disks connected directly to the computer.

II. ARCHITECTURE OF SAN

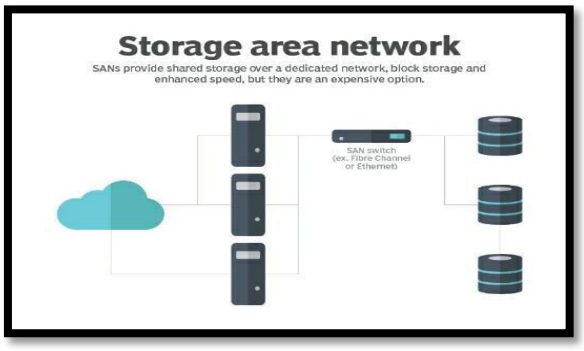


Figure 1. Architecture of SAN

A SAN or Storage area network is a special high-speed network that stores and provides access to large amounts of data .So basically it's a dedicated network that's used for data storage and this network consists of multiple disks arrays, switches and servers. And because it has more than one of these devices, a SAN is fault tolerant and the data is also shared among several disk arrays So if a switch or disk array or if a

server goes down the data can still be accessed. When a server accesses the data on a SAN, it accesses the data as if it was a local hard drive because that's how operating system recognize a SAN it's recognized as a local attached hard drive rather than a shared network drive like in a NASA and SAN's are also highly scalable because adding more storage space can easily be done without an interruption on the network.

Now as We stated a SAN is high speed network and that's because in a SAN all the devices are interconnected to each other and they are interconnected using fibre channel ,Which is a standard for SAN .Fibre channel is simply a fibre optic and its has speeds between 2 gigabits per second i.e 128 gigabyes per second so fibre channel is extremely fast and it's also very expensive and most of the SAN's today use fibre channel but also an alternative to fibre channel some SAN's use iSCSI (internet Small Computer System Interface)instead which is a cheaper alternative to fibre channel but it's not as fast as fibre channel. SAN are highly scalable and very redundant.

III. COMPONENTS OF SAN

SAN components is divide into 3 layers, Host layer, Fabric layer/components and storage layer. SAN is formed when these three layers are connect with each other. Each layer has its own function. Representation of SAN components are given below in the figure.

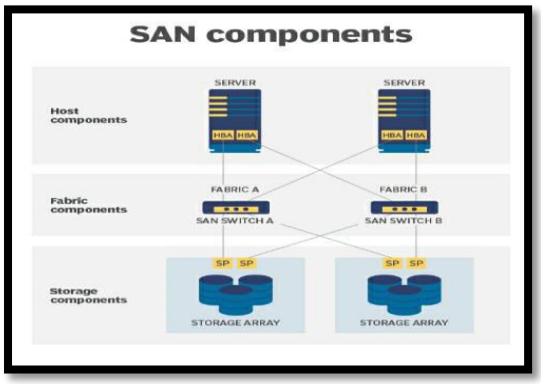


Figure 2. SAN Components

3.1. Host Layer

The host layer of SAN storage has servers or hosts .Hosts are simply an hardware having resources like CPU ,memory ,Disk Storage And Operating system running on it. It also has dual HBA (Host Bus Adapter) cards which connects to the fabric layer of SAN network

3.2. Fabric Layer

Fabric Layer or fabric of SAN Consists of one or more SAN switch, which provides connections points in a SAN architecture. All the hosts and storage array are connects to the SAN switch .Thus it acts as a mediator between Host and Storage.

3.3. Storage Layer/Components

In storage Layer, Storage array consists of a physical disks .These disks are grouped logically using a special technology called RAID (Redundant Array of Independent Drive) .Most of the storage arrays have Dual storage processors (SPs).This SP has a front end IO module that has Fibre Channel (FC) ports. These FC ports connect to SAN switch via FC cable

IV. HOW CAN WE USE STORAGE AREA NETWORK?

Storage Area Network (SAN) might bring a highly data dependent business infrastructure that are abbreviated as follows,

4.1. Infrastructure simplification

There are 4 important ways through which infrastructure simplification can be fulfil they are,

a. **Consolidation**-Consolidation refers the process of uniting thus, a method of centralized data storage among multiple servers which helps to increase availability and scalability.
b. **Virtualization-** Storage virtualization is the process of presenting a logical view of the physical storage resources. Storage virtualization helps in making complexity nearly transparent. Virtualization is used to make infrastructure responsive.
c. **Automation-**selecting storage components with autonomic capabilities can improve availability and responsiveness, also help to protect data as storage needs to grow.
d. **Integration-**Storage integrated environment simply manage tasks and improve security.

4.2. Information life cycle management

Information lifecycle management (ILM) is a process for managing information through its lifecycle, from conception until intentional disposal. The ILM process manages this information in a manner that optimizes storage and maintains a high level of access at the lowest cost.

4.3. Business Continuity

Business Continuity is a framework which is used to detect the organization internal and external issues. Businesses must also be increasingly sensitive to issues of client privacy and data security so that vital information assets are not compromised. It must include disaster recovery, business recovery, crisis management and contingency planning.

V. SECURITY FOR SAN

Storage area network security holds all the measures company takes to protect the data contained within storage devices on their network. Storage are network security protects a company's network of

HDD's(Hard disk drive), SSd's (solid state drives) and RAID(redundant array of independent disk),Storage and servers. They must contains the component like access, encryption, controls etc.

SAN security can be applies for the both IP and Fibre channel networks:,

- The IP-based protocol iSCSI used TCP/IP for transmission of data, which gives flexibility to run on various networks but also it is unsafe protocol.
- If a FC network connects to an IP network, it is considered to be IP vulnerabilities as well.
- FC network should still have stringent access controls.

5.1. Types of SAN Security:-

Nowadays, various organization uses a multiple types of security to secure both the physical and Digital part of the networks. In Physical security consists the physical elements of SAN such as routers, network switches, Storage devices and another kinds of hardware components and also to protect all the types of data from theft, damage and hacking digital security is must.

Some of the key components for SAN security as as follows,

- **Access Control:** The primary Goal of access control is to minimize the security risk of unauthorized access. By applying these unauthorized access, credential theft hacking can be controlled.
- **Network Protocols:-**Protocols like Secure socket Layer(SSL) helps companies to protect SAN's That's are connected to the internet
- **Digital certificates:-**SAN which are connected with switches should be authenticated through a digital certificate before being permitted on the network. Due to this the chances of switch spoofing attack is reduced
- **Backup: -** Backup is used to make copies of data that is actively in use. All the copies of backed-up data should be encrypted and at least two should be stored in two different physical location unless one is damaged

VI. CONCLUSION

- SAN is significantly different from its predecessors such as Direct Attached Storage (DAS) and Redundant Array of Independent Disks (RAID), partly because it creates a shared pool of storage devices that are networked together to create a single view of storage.
- Sharing storage usually simplifies storage administration and adds flexibility since cables and storage devices do not have to be physically moved to shift storage from one server to another.
- SAN helps companies consolidate their storage management by allowing their storage management by allowing them to apply the same policies for multiple devices and arrays in different locations. But for that storage must be secure.
- As SANs are scaled to become campus or metropolitan area networks they become more vulnerable to security breaches.
- SAN is expensive and complex to manage and set up because it is tightly coupled network structure.

REFERENCES

1. Storage Area Networks: Volume 14/Article 27/Manjari Mehta, University of Houston, Manjari.Mehta@mail.uh.edu
2. Storage Area Networks and the High Performance Storage System
    a. This article was submitted to Nineteenth IEEE Symposium on Mass Storage Systems April 15-18, 2002/UCRL-JC-146951
3. Introduction to Storage Area Networks and System Networking. **Fifth Edition (November 2012) ,**This edition applies to the products in the IBM System Networking portfolio https://www.redbooks.ibm.com
4. Storage area network fundamentals-Meeta Gupta,C Anita Sastry,Cisco Press,2002
5. https://www.vmware.com/topics/glossary/content/storage-area-network-san.html
6. https://www.cisco.com/c/en/us/solutions/data-center/data-center-networking/what-is-storage-area-networking.html
7. Introduction to Storage Area Network, SAN Ravi Kumar Khattar, Mark S Murphy, Giulio John Tarella, Kjell E Nystrom IBM Corporation, International Technical Support Organization, 1999
8. Figures:- https://www.techtarget.com/searchstorage/definition/storage-area-network-SAN

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-29

MAC Protocol Finding Division of Chromosome's in using Genetic Algorithm and SVM

Dr. Mrs. Tanuja Fegade.
KCE's IMR, Jalgaon
Email:tanujamahajan18@imr.com

Ms. Bhavana Jawale.
KCE's IMR, Jalgaon
Email:bhavanajawale@imr.com

***Abstract*—**Cognitive wireless network, *SVM Clustering using Genetic Algorithm* MAC sub layer protocol for supporting the usage of multiple heterogeneous network in Ad hoc wireless network have been proposed. SVM based clustering method by using SVM model in separable in structural risk minimization to observe the division of chromosomes. An analyze the genetic algorithm in SVM clustering techniques to use as mac protocol choose of qualified kernel and separated Transmitted in traffic & exchange information of chromosomes.

Index Terms*—Ad hoc wireless networks, Multiple access control (MAC) protocol, connectivity, cooperation, heterogeneous wireless networks, selective multicast, complete multicast, Genetic algorithm*

I. INTRODUCTION

Cognitive radio ad hoc networking has emerged as a promising technology to enhance spectrum utilization by sensing the spectrum and opportunistically accessing the spectrum of primary (licensed) systems [1], [2]. cognitive networks was introduced to increase the efficiency of spectrum utilization. The basically cognitive networks is to permit other users to utilize the spectrum allocated to licensed users (*primary users*) when it is not being used by them. These other users who are opportunistic users of the spectrum are called *secondary users[3]*.

Multi-channel MAC protocols for ad-hoc wireless networks have represented a first step in the development of MAC protocols for cognitive radio in unlicensed scenarios. The number of channels available at each user is fixed in a multi-channel network, while it varies with time and space in a cognitive network[4]. we study the capacity scaling laws of the cognitive network consisting of the primary ad hoc network (PaN) and the secondary ad hoc network (SaN). We directly derive the multicast capacity that unifies results on unicast and broadcast capacity [5], [15]. An important constraint for a cognitive network model is that the primary network does not alter its protocol due to the secondary network anyway. Such control information is multicast to its neighbors in a traditional network. Since in a cognitive network, each node has a set of channels available, a node receives a message only if the message was sent in the channels on which the node was listening to. So, to ensure that a message is successfully sent to all neighbors of a node, it has to be multicast in every other channel[6]. This is called *complete multicasting* of information. In a cognitive environment, the number of channels is potentially large. As a result multicasting in every channel causes a large delay in transmitting the control information, it does not contain any traffic[7].

II. BACKGROUND

A. Percolation Analysis:

The concept of percolation is originated from statistical physics and random networks [9], [17]. In our study, as the transmit power of a node increases, the mean degree increases. If we draw links between each node and its neighbors, the resulting network graph percolates (i.e., an infinite connected component appears in the network) when is above a percolation

threshold .On the other hand, when below the percolation threshold, the network does not percolate and consists of many isolated finite components. The percolation threshold can be computed by finding the critical point at which the mean component size blows up, indicating the formation of an infinite connected component[8].

B. Selective Multicasting:

In a MHCRN, each node has a set of channels available when it enters a network. In order to become a part of the network and start communicating with other nodes, it has to first know its neighbors and their channel information. Also, it has to let other nodes know its presence and its available channel information. So it multicasts such information over all channels to make sure that all neighbors receive the messages. Similarly, when a node wants to start a communication it
should exchange certain control information useful, for example, in route discovery. However, a cognitive network environment is dynamic due to the primary user's traffic [6]. The number of available channels at each node keeps changing with time and location. To keep all nodes updated, the information change has to be transmitted over all channels as quickly as possible. So, for effective and efficient coordination, fast dissemination of control traffic between neighboring users is required. So, minimal delay is a critical factor in promptly disseminating control information. Hence, the goal is to reduce the multicast delay of each node.
Multicasting is used to send a message to a specific group of nodes in a particular channel. In a multichannel environment where the nodes are listening to different channels[11].
Now, consider that a node has M available channels groups. Let Tt be the minimum time required to multicast a control message. Then, total multicast delay = M *Tt.
So, in order to have lower multicast delay we need to reduce M . The value of Tt is dictated by the particular hardware used and hence is fixed. M can be reduced by finding the minimum number of channels groups, M ' to multicast, but still making sure that all group nodes receive the message. Thus, multicasting over carefully selected M ' channels instead of blindly multicasting over M (available) channels on groups are called *Selective Multicasting*

III. NETWORK MODEL

A. Support-vector machine

Support-vector machines (SVMs, also support-vector networks[1]) are supervised learning models with associated learning algorithms analyze data for classification and regression analysis. Developed at AT&T Bell Laboratories by Vladimir Vapnik with colleagues (Boser et al., 1992, Guyon et al., 1993, Cortes and Vapnik, 1995,[2] Vapnik et al., 1997).SVMs are one of the most robust prediction methods are being based on either statistical learning frameworks or VC theory which is proposed by Vapnik (1982, 1995) and Chervonenkis (1974).The set of training examples, each marked as belonging to one of two categories.SVM training algorithm builds a model that assigns new examples to one categories. The other making it a non-probabilistic binary linear classifier .SVM maps training to points in space maximize the width of the gap between the two categories. The same space and predicted to a category based on which side of the gap them fall. The linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces[14].
When data is unlabeled, supervised learning don't possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The support-vector clustering [3] algorithm, created by the Siegel Mann and Vladimir Vapnik, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data.
Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection.

The advantages of support vector machines are:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

The disadvantages of support vector machines include:

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
- SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation (see Scores and probabilities, below).

B. Genetic Algorithm:

They are commonly used to generate high-quality solutions for optimization problems and search problems.

Genetic algorithms simulate the process of natural selection which means those species who can adapt to changes in their environment are able to survive and reproduce and go to next generation. In simple words, they simulate "survival of the fittest" among individual of consecutive generation for solving a problem. Each generation consist of a population of individuals and each individual represents a point in search space and possible solution. Each individual is represented as a string of character/integer/float/bits. This string is analogous to the Chromosome

1) Generality and Versatility: GAs are apply in a wide variety of settings and easily molded to particular problems. GAs constitute a very general meta-heuristic technique which thought of as the sledgehammer of the craft of algorithms, much like the technique of artificial neural networks (ANNs). GAs readily invoked in areas don't yield readily to standard approaches, or when more specialized techniques fail. GAs are capable of solving extremely large problems that are large search spaces. GAs are very good at navigating through huge search spaces to heuristically find near optimal solutions in quick time. GAs are work even when the objective function is not Exactly known since GAs rely only on an objective function's evaluation (without necessarily knowing the objective function explicitly). Although GAs guarantee optimality generally useful in practice. Bandar et al. [8] showed that although GAs don't guarantee convergence to an optimal solution, GAs avoid local optimal with a high probability through the use of mutation and crossover operators.

1) Addictiveness and Online Problem Solving: To understand adaptively in changing conditions, there are initially develop the idea of a 'landscape' both as a metaphor and as a mathematical object. Building upon the metaphor of fitness landscapes, and the insight of GAs as stochastic search algorithms, two landscape models of relevance to optimization through GAs are proposed [9]. Rugged landscapes are landscapes in which they are many peaks, valleys, and troughs (and not a single peak). Rugged landscapes fitness levels do not change; in evolutionary systems, the fitness function is also dependent on context, and on the behavior of other competitors. They can be captured by the metaphor of dancing landscapes—which are adapted from rugged landscapes. Local peaks may change, making a solution that was earlier optimal no longer a peak when a landscape 'moves'. Traditional techniques from optimization theory assume static and well-known topologies and does not suited to dynamic environments[13]. In many problems of practical interest, this focus is more on satisficing (i.e., finding a sufficiently good solution that satisfies one's purposes) rather than on optimizing (i.e., finding the best possible solution). A key benefit of GAs is that they are well suited to the optimization task in the dancing landscapes that characterize wireless networks. GAs are well suited for building met heuristic adaptive algorithms that can provide satisfactory performance in changing network conditions.

Another important aspect of GAs are an online adaptive algorithm that can operate in unknown environments in an online fashion. Such abilities are crucial in wireless networks in various Control settings by which decisions have to make automatically in run-time to cater to dynamic channel parameters. In dynamic channel conditions that typically exist in most wireless networking configurations, the optimal solution keeps changing as the conditions changes when optimization is done by an elusive moving target. In online adaptive optimization algorithms, they are fundamental tradeoff between exploration—which involves looking for potentially better previously unexplored solutions—and exploitation that implies the use of previously known good solutions. Exploration in effect is an attempt to find good adaptive building blocks and exploitation is the use and propagation of adaptations known to perform well. They can envision mutation and recombination of genes as analogous to exploration, whereas natural selection can be envisioned as a form of exploitation [9]. It is important to always keep on exploring when the dynamic environment changes , such as wireless networks. Holland's original GA work used schema analysis to show that GAs can achieve near-optimal exploration and exploitation tradeoff, under qualifying assumptions

2). GA Terminology: Since the field of GA is inspired from the biological genetic evolution, the field uses a number of biological metaphors in its terminology. The purpose of this short subsection is biological metaphors in the context of applications of GAs in wireless networking.

The terminology through the following list.

_ Organism: It represents the entity (e.g., a radio parameter, or a wireless resource) being optimized.

_ Population: simulated genetic evolution is a collection of organisms.

_ Chromosome: It encodes a particular solution to the problem under study. (Biologically, a chromosome contains an organism's genetic makeup.)

_ Fitness: It represents the utility of the current chromosome.

_ Gene: It is the basic building block of the chromosome defining a particular feature of the simulated organism. (Each chromosome can contain a number of genes.)

_ Allele: Each gene can take several alternative forms, each of which is called an allele.

_ Locus: It is the position on the chromosome containing a particular gene of interest. Support Vector Machines (SVMs) provide a powerful method for supervised learning classification. Use of SVMs for clustering (unsupervised learning) is now being considered in a number of different way

Supervised cluster evaluation

Support Vector Machine uses Structural Risk Minimization to compare various separation models and to eventually choose the model with the largest margin of separation.

Support Vector Machine uses Structural Risk Minimization to compare various separation models and to eventually choose the model with the largest margin of separation.

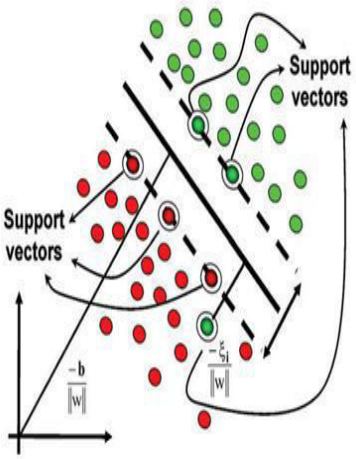


Figure1: Support Vector Machine uses Structural Risk Minimization to compare various separation models and to eventually choose the model with the largest margin of separation.

The performance of unsupervised algorithms, such as clustering, don't measured with the same certitude as for the classification problems. The clustering is measured using the externally derived class labels for the patterns. Subsequently, they are use some of the classification-oriented measures to evaluate their results. These measures evaluate the extent to which a cluster contains patterns of a single class ([4]).

Applications:

SVMs are used to solve various real-world problems:

- SVMs are helpful in text and hypertext categorization, as our application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.[8] Some methods for shallow semantic parsing are based on support vector machines.[9]
- Classification of images is also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback. There is also true for image segmentation systems, including those using a modified version SVM that uses the privileged approach as suggested by Vapnik.[10][11]
- Classification of satellite data like SAR data using supervised SVM.[12]
- Hand-written characters can be recognized using SVM.[13][14]
- The SVM algorithm is widely applied in the biological and other sciences. They are used to classify proteins with up to 90% of the compounds classified correctly. Permutation tests based on SVM weights are suggested as a mechanism for interpretation of SVM models.[15][16] Support-vector machine weights have also used to interpret SVM models in the past.[17] Postdoc interpretation of support-vector machine models in order to identify features used by the model to make predictions is a relatively new area of research with special significance in the biological sciences.

A. Wireless Ad Hoc Networks:

Wireless ad hoc networks are multi-hop systems in which nodes assist each other in transmitting and receiving packets across the network. In contrast to infrastructure-based networks, in an ad hoc network nodes can discover and exchange information with each other directly without involving central access points. A node may join or leave the network at any time. In ad hoc networks, all nodes have equal right to access the medium. To be able to establish communication with each other, each node needs to be able to see the others. If a node wishes to communicate outside its range, another node within the same range operates as a gateway and forwards the contact in a multi-hop fashion. The recent work related to ad hoc networks focus on many issues such as network architecture and network capacity.

B. Medium access control Protocols:

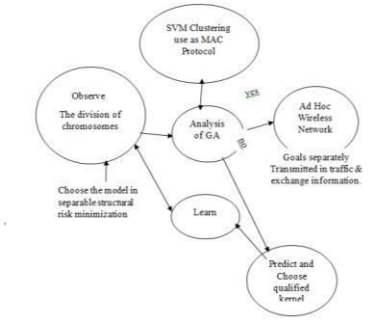
Multiple access protocols are a set of protocols operating in the Medium Access Control sub layer (MAC sub layer) of the Open Systems Interconnection (OSI) model. These protocols allow a number of nodes or users to access a shared network channel. Several data streams originating from several nodes are transferred through the multi-point transmission channel.

The objectives of multiple access protocols are optimization of transmission time, minimization of collisions and avoidance of crosstalk's.

Proposed Work:

In proposed system, SVM based clustering method by using SVM model in separable in structural risk

minimization to observe the division of chromosomes. An analyze the genetic algorithm in SVM clustering techniques to use as mac protocol. The ad hoc wireless network.to specify the learn to predict and choose of qualified kernel and separated Transmitted in traffic & exchange information of chromosome.



Conclusion: SVM clustering techniques using mac protocol analyze the genetic algorithm in SVM to remove producing traffic and exchange information without any chromosomes of ad hoc wireless network to specify the learn to predict and choose of qualified kernel and separated in chromosomes.

References:

[1] F. Hillier and G. Lieberman, "Introduction to operations research," McGraw Hill, New York, 2001.

[2] M. Ridley, Evolution. 3rd Edition. Blackwell, 2004.

[3] M. Mitchell, An introduction to genetic algorithms. MIT press, 1998.

[4] J. Holland, "Genetic algorithms, computer programs that evolve in ways that even their creators do not fully understand," Scientific American, pp. 66–72, 1975.

[5] A. M. Turing, "Computing machinery and intelligence," Mind, pp. 433– 460, 1950.

[6] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," Machine learning, vol. 3, no. 2, pp. 95–99, 1988.

[7] R. High field and M. Nowak, Super Cooperators: Evolution, Altruism and Human Behavior (or why we need each other to succeed). Text Publishing, 2011.

[8] D. Bhandari, C. Murthy, and S. K. Pal, "Genetic algorithm with elitist model and its convergence," International Journal of Pattern Recognition and Artificial Intelligence, vol. 10, no. 06, pp. 731–747, 1996.

[9] S. E. Page, Diversity and complexity. Princeton University Press, 2010.

[10] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," Reliability Engineering & System Safety, vol. 91, no. 9, pp. 992–1007, 2006.

[11] L. Davis et al., Handbook of genetic algorithms, vol. 115. Van Nostrand Reinhold New York, 1991.

[12] W. Pedrycz and A. Vasilakos, Computational intelligence in telecommunications networks. CRC press, 2010.

[13] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey," Computer, vol. 27, no. 6, pp. 17–26, 1994.

[14] K.-F. Man, K.-S. Tang, and S. Kwong, "Genetic algorithms: concepts and applications," IEEE Transactions on Industrial Electronics, vol. 43, no. 5, pp. 519–534, 1996.

[15] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty, "NeXt generation/ dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, Sept. 2006.

[16] S. Haykin, "Cognitive radio: brain-empowered wireless communications, *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[17] A. Goldsmith, S. A. Jafar, I. Mari´c, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009. [16] W. Ren, Q. Zhao, and A. Swami, "Connectivity of heterogeneous wireless networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4315– 4332, July 2011.

[18] K. Bian and J.-M. Park, "MAC-layer misbehaviors in multi-hop cognitive radio networks," 2006 US – Korea Conference on Science, Technology, and Entrepreneurship (UKC2006), Aug. 2006.

[19] J. Zhao, H. Zheng, G.-H. Yang, "Distributed coordination in dynamic spectrum allocation networks", in: Proc. IEEE DySPAN 2005, pp. 259- 268, November 2005.

[20] Weng Chon Ao, Shin-Ming Cheng, *Member, IEEE,* and Kwang-Cheng Chen, *Fellow,* "Connectivity of Multiple Cooperative Cognitive Radio Ad Hoc Networks", IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 30, NO. 2, FEBRUARY 2012.

[21] R. Muthukkumar1, Lalit Garg2, K. Maharajan3, M. Jayalakshmi3, Nz Jhanjhi4, S. Parthiban5 and G. Saritha6" A genetic algorithm-based energy-aware multi-hop clustering scheme for heterogeneous wireless sensor networks" Published 15 August 2022.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-30

Next Wave of Computing-Quantum Computing

Swati Borse,
Associate Professor,
Department of Computer Science,
SSVPS's L.K.Dr. P.R. Ghogrey Science College, Dhule, Maharashtra, India.
borseswati28@gmail.com

Amodkumar Suryawanshi

Scientist, National Informatics Centre, Mumbai, Maharashtra, India.
kahonaamod@gmail.com

Abstract - *The upcoming era of Computer Science is Quantum Computing. This paper provides an overview of quantum computing, classical computing and its limitations, building blocks of Quantum Computing. We explained the fundamental principles of quantum mechanics that enable quantum computers to perform certain tasks exponentially faster than classical computers. We discussed potential applications of Quantum computing in various fields, including cryptography, optimization, machine learning, and simulation. We also highlight some of the current challenges in developing and scaling quantum computing technology, such as noise and error correction, hardware limitations, and the need for new algorithms and software. Finally, we explore some of the ethical and societal implications of quantum computing, including its potential impact on cybersecurity, privacy, and economic competitiveness.*

Keywords - Quantum, Computing, cryptography, optimization

I. INTRODUCTION

Quantum computing is an emerging field that aims to develop computers based on the principles of quantum mechanics. Classical computers rely on binary digits, or bits, which can only represent two possible states: 0 or 1. In contrast, quantum computers use quantum bits, or qubits, which can represent both 0 and 1 at the same time through a phenomenon known as superposition. This allows quantum computers to perform certain tasks exponentially faster than classical computers, such as factoring large numbers and solving complex optimization problems. The principles of quantum mechanics also enable a second key feature of quantum computing, known as entanglement. Entanglement allows two or more qubits to be linked together in such a way that the state of one qubit can affect the state of the other qubits, even if they are physically separated by large distances. This property has potential applications in cryptography, where entangled qubits can be used to create secure communication channels that are resistant to eavesdropping and hacking.

Despite the potential advantages of quantum computing, there are many challenges that must be overcome in order to develop practical quantum computers. These challenges include developing hardware that can reliably store and manipulate qubits, as well as developing new algorithms and software that can take advantage of

the unique properties of quantum mechanics. Nonetheless, the field of quantum computing is rapidly advancing, and many experts believe that quantum computers will play an

increasingly important role in fields ranging from drug discovery to financial modeling in the coming years.

II. REVIEW OF RELEVANT LITERATURE

Quantum is the smallest amount of any physical entity involved in an interaction. During a lecture in 1982, Richard Feynman, an American physicist, suggested the utilisation of quantum

mechanical principles to carry out computations that would be difficult or unfeasible to achieve through classical computers [1]. This idea was later developed into the field of quantum computing. Quantum mechanics emerged as a branch of physics in the early 1900s to explain nature on the scale of atoms. Quantum comes from Latin, meaning "an amount" or "how much?" If something is quantifiable, then it can be measured. Although the notion of combining quantum mechanics and information theory emerged in the 1970s, it received scant consideration in 1982 when Richard Feynman, an American physicist, suggested the utilization of quantum mechanical principles to carry out computations that would be difficult or unfeasible to achieve through classical computers. The year 1994 witnessed a significant surge in the interest in quantum computing, following the development of a quantum algorithm by mathematician Peter Shor. The algorithm was capable of determining the prime factors of large numbers in polynomial time on an efficient quantum computer.

E. Farhi, J. Goldstone, S. Gutmann, and M Sipser in their research paper introduced the concept of quantum annealing, a method of solving optimization problems using a quantum computer that is designed to minimize the energy of a physical system[11]. I. Chuang and M. Nielsen provide a comprehensive introduction to quantum computing, covering topics ranging from basic quantum mechanics to quantum algorithms and quantum error correction[12]. J. Martinis, S. Boixo, S. Isakov, et al. proposed a method for demonstrating quantum supremacy, the point at which a quantum computer can solve a problem that is intractable for a classical computer, using superconducting qubits[13]. These papers have contributed significantly to the development of quantum computing, laying the foundations for many of the current approaches to building practical quantum computers and exploring their potential applications.

A. Ambainis, J. Kempe, and A. Rivosh explore in their paper the use of quantum walks on graphs, which could have applications in graph theory, network analysis, and optimization[14]. H. Spedalieri, and I. Chuang proposes a quantum annealing algorithm for clustering data, which could have applications in machine learning and data analysis[15]. E. Farhi and J. Goldstone explores the potential applications of quantum computing in finance, including portfolio optimization, option pricing, and risk analysis[16]. T. Wong, A. Ekert, and J. O'Brien explores the potential applications of quantum computing in image processing, including image compression, pattern recognition, and machine vision[17]. J. Biamonte, P. Wittek and et. al. provides an overview of quantum machine learning, which involves using quantum computers to speed up classical machine learning algorithms and develop new algorithms based on quantum principles[18]. A. D. King, J. Carrasquilla and et. al. report on the largest-scale quantum annealing experiment to date, using a D-Wave quantum annealer with more than 100 qubits to solve optimization problems[19].

These papers demonstrate the wide range of potential applications of quantum computing, spanning fields from finance to machine learning and image processing.

III. CLASSICAL COMPUTING AND IT'S LIMITATIONS

The number of transistors in an integrated circuit (IC) doubles about every two years, this observation is named after Gordon Moore (co-founder of Intel), which is known as Moore's law [2]. Another tenet of Moore's Law says that the growth of microprocessors is exponential. If Moore's Law is extrapolated naively to the future, it is learned that sooner or later, each bit of information should be encoded by a physical system of subatomic size [3]. Classical computers are built with microprocessors having limitations for expansion and rely purely on classical computing principles and properties.

A. Invent of Quantum Computing

Although the notion of combining quantum mechanics and information theory emerged in the 1970s, it received scant consideration in 1982 when Richard Feynman, an American physicist, suggested the utilization of quantum mechanical principles to carry out computations that would be difficult or unfeasible to achieve through classical computers.

B. Quantum Computing

Quantum computing is a multidisciplinary field comprising computer science, physics, and mathematics that utilizes quantum mechanics (superposition, entanglement, and interference) to solve complex problems faster than classical computers.

1) Superposition: Superposition is the ability of a quantum system to be in multiple states at the same time until it is measured. [4] This idea is often illustrated by the famous thought experiment of Schrödinger's cat [5], where a hypothetical cat in a box can be simultaneously alive and dead until the box is opened and the cat's state is observed. In another double-slit

experiment, a single particle can pass through two slits at the same time and create an interference pattern on a screen.

2) Entanglement: Quantum entanglement is the relation between two subatomic particles that can be intimately linked to each other even if separated by a large distance [6]. This strange behaviour, which was famously referred to by Albert Einstein as "Spooky action at a distance," has been demonstrated experimentally many times and is now a well-established phenomenon in quantum mechanics. Entangled qubits always correlate with each other to form a single system. Consider two particles, A and B, that are entangled. The particles are created together in a way that their properties become correlated. For example, they could be created in such a way that their spins are always opposite. This means that if particle A is measured to have "spin up," then particle B is guaranteed to have "spin down," and vice versa. Entanglement allows the state of one qubit to affect the state of another qubit, even if they are separated by large distances.

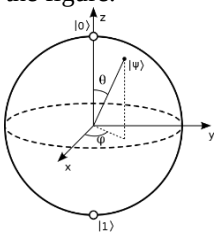
3) Quantum interference : Quantum interference in quantum computing is used to affect probability amplitude. It can influence the probability of the outcomes when the quantum state is measured. Every possible outcome has some probability of occurring. Quantum interference, along with quantum entanglement, is essential to the operation of quantum computers.

4) Quantum Supremacy : Quantum Supremacy was termed by John Preskill in 2012. The quantum device can solve a problem that no classical computer can solve in any feasible amount of time[8]. It can be named quantum primacy or quantum advantage

Overall Quantum Computing works by using unique properties of quantum mechanics to process information in different ways than classical computers , allowing certain types of computing much faster.

IV. BUILDING BLOCKS OF QUANTUM COMPUTING

1) Qubits: The fundamental unit of Information in quantum computing is Qubit (Quantum Bit). The binary bit is either 0 or 1 used in a classical computer, whereas a qubit exists in the superposition state i.e. in multiple states of 0 and 1 at the same time until it is measured. Qubit is labelled as |0> and |1>. Bloch sphere representation of a qubit is shown in the figure.



Qubits can be made from trapped ions, photons, artificial or real atoms, or quasiparticles.

2) Quantum Gates: Quantum gates are operations that act on qubits and transform their state. They are the basic building blocks for quantum algorithms, and they enable complex computations to be performed on qubits.

3) Quantum Registers: A quantum register is a collection of qubits that are treated as a single unit. Registers are used to store and manipulate quantum information.

4) Quantum Circuits: A quantum circuit is a sequence of quantum gates applied to one or more qubits. Quantum circuits are used to implement quantum algorithms and perform quantum computations.

5) Measurement: Measurement is the process of reading out the state of a quantum system. In quantum computing, measurements are used to extract classical information from a quantum state.

6) Quantum Error Correction: Due to the fragile nature of quantum states, quantum computers are susceptible to errors caused by environmental noise and other factors. Quantum error correction techniques are used to detect and correct errors in quantum computations.

7) Quantum Algorithms: Quantum algorithms are designed to take advantage of the unique properties of quantum mechanics to solve problems faster than classical algorithms. They can use techniques like quantum parallelism and quantum interference to speed up certain types of calculations.

Basically, the building blocks of Quantum computing include qubits, quantum gates, quantum registers, quantum circuits, measurements, quantum error correction, and quantum algorithms. These building blocks work together to enable the unique capabilities of quantum computing.

V. QUANTUM COMPUTING APPLICATIONS

Quantum Computing applications have become a reality now. Quantum computing has the potential to revolutionize many areas of science and technology. At Quantum Technology & Application Consortium (QUTAC), important German companies come together to put our country and Europe at the forefront of the development of this crucial future technology [28]. Some potential applications of Quantum computing include :

1) Artificial Intelligence & Machine Learning: Quantum computers could be used to speed up certain machine learning algorithms, such as those used for clustering and classification. The advancements in time and space complexity of quantum computation are making it a highly promising tool for machine learning and data analysis applications. Quantum machine learning is a field of study that explores the potential of utilizing phenomena from the quantum realm to address machine learning problems.[20]

2) Computational Chemistry: Quantum computing will assist in solving classical chemistry problems such as high-temperature superconductivity, solid-state physics, transition metal catalysis, and specific biochemical reactions [21].

3) Drug Design & Development: Quantum computing in drug development based on the structure of target proteins is listed chronologically. They include protein structure prediction, molecular docking, quantum simulation, and quantitative structure-activity relationship (QSAR) models [22]. The integration of expertise in bioinformatics, cheminformatics, medicinal chemistry, and quantum computing is making it possible to solve pharmaceutical problems that are challenging for classical computers.

4) Cybersecurity & Cryptography: Quantum computers have the potential to break existing classical cryptographic protocols, as it's based on the difficulty of factoring large numbers or solving the discrete logarithm problem. In view growing power of next-generation quantum computers, post-quantum cryptography has become a critical area for ensuring the safety and security of data and communication over the next few decades. [23]. The Post Quantum Cryptography algorithms are based on Lattice-based, Hash-based, Code Based, Multivariate-based, and Isogeny-based.

5) Financial Modelling: Quantum computing could be used to accelerate financial modelling and risk analysis by enabling faster Monte Carlo simulations. D-wave annealers and gate-model simulators demonstrate portfolio optimization with the highest expected return (profit) given a certain level of risk (i.e., the standard deviation of portfolio returns), or to minimize risk for a given level of expected return [24].

6) Logistics Optimisation: Many real-world optimization problems, such as those arising in logistics, scheduling, and finance, could be more computationally difficult for classical computers. Quantum computers have the potential to provide exponential speedup for solving these problems. Optimization in the logistics sector is experimented on quantum gate-based variational algorithms and quantum annealers [25]

7) Weather Forecasting: When it comes to handling substantial amounts of data, quantum computers offer a wide range of benefits. Quantum computers will be able to effectively handle modern Numerical Weather Prediction Models (NWP) that contain over one billion variables, including velocity, temperature, humidity, and other components.[26].

8) Better Batteries: The utilization of computational chemistry techniques based on quantum mechanics could lead to a more precise understanding of material properties, which in turn may assist the industry in the development of improved battery designs [27].

Overall Quantum has the potential to revolutionize many areas of science and technology and there is ongoing research into developing new applications and optimizing existing ones.

VI. CHALLENGES IN DEVELOPING QUANTUM COMPUTING

Quantum computing is a rapidly evolving field that holds promise for solving some of the most complex computational problems that classical computers cannot efficiently solve. However, there are several challenges in developing quantum computing are:

1) Building stable qubits: Qubits are the basic building blocks of quantum computers, and they are much more fragile than classical bits. Any external noise or interference can disrupt the fragile quantum states and lead to errors. Researchers are exploring different approaches to build qubits that are stable and resistant to errors, such as

superconducting qubits, ion traps, and topological qubits.

2)Reducing error rates: Errors in quantum computations can arise due to a variety of factors, such as noise, decoherence, and imperfect operations. Reducing the error rates is crucial for developing large-scale, fault-tolerant quantum computers. Researchers are exploring various error correction and fault-tolerance techniques to overcome this challenge.

3) Scaling up: To solve complex problems, quantum computers need to have a large number of qubits, and the ability to perform complex quantum operations. Scaling up quantum computers while maintaining the coherence of the qubits and the fidelity of the quantum operations is a significant challenge.

4) Developing quantum algorithms: Developing quantum algorithms that can effectively utilize the power of quantum computers is a major challenge. While some quantum algorithms, such as Shor's algorithm for factorization and Grover's algorithm for searching, have been developed, there is still a need to develop more quantum algorithms that can solve practical problems.

5) Reducing hardware costs: Developing and operating quantum computers is currently expensive. Researchers are working on developing more cost-effective hardware and exploring different architectures and technologies to make quantum computing more accessible.

Overall, quantum computing is a challenging field that requires multidisciplinary approaches and collaborations between researchers, scientists, engineers, and industry partners. Despite these challenges, the potential of quantum computing to transform computing, cryptography, and other fields is immense, making it an exciting area of research and development.

VII. ETHICAL AND SOCIAL IMPLICATIONS OF QUANTUM COMPUTING

Quantum computing has the potential to transform many fields of society, but it also raises ethical and social implications that need to be considered. Here are some of the key ethical and social implications of quantum computing:

1) Cybersecurity: Quantum computing can break many of the encryption methods used to secure digital communication and data storage, including those used for online banking, e-commerce, and government communications. This could lead to a loss of trust in digital systems and the potential exposure of sensitive information.

2) Economic Disruption: Quantum computing could disrupt industries by making it possible to solve problems that are currently unsolvable or impractical with classical computing. This could lead to significant economic shifts and job losses in some industries.

3) Ethical Use of Quantum Computing: As with any powerful technology, there is a risk of quantum computing being used for unethical purposes, such as hacking, cyberattacks, or the development of advanced weapons systems. It is important to ensure that the development and use of quantum computing leads to ethical and moral standards.

4) Access and Inequality: The development of quantum computing is expensive and requires specialized expertise, which could lead to unequal access and opportunity for individuals and organizations.

5) Environmental Impact: Quantum computing requires significant amounts of energy and generates a lot of heat, which could have an impact on the environment. It is important to consider the environmental implications of the development and use of quantum computing.

In order to mitigate these ethical and social implications, it is important for researchers, policymakers, and industry leaders to collaborate and address these issues in a proactive and transparent manner. This will require collaboration across a range of stakeholders, including scientists, policymakers, and the public.

VIII. CONCLUSION

The quantum computers power to perform calculations very quickly and perform tasks that classical computers will never be able to practically achieve. For now, at least, the world of cryptography is safe because the quantum computer is proving to be very difficult to implement. The very thing that makes them powerful, their reliance on quantum mechanics, also makes them extremely fragile. A quantum computer thus has the theoretical capability of simulating any finite physical system and may even hold the key to creating an artificially intelligent computer.

Quantum computation promises the ability to compute solutions to problems that, for all practical purposes, are insoluble by classical computers. However, the quantum promise is still a long way from achieving practical realization. Some properties of quantum mechanics that enable

quantum computers superior performance also make the design of quantum algorithms and the construction of functional hardware extremely difficult. We need to imply some solutions to improve the quality of qubit technology by increasing the coherence time of qubits and the speed of quantum operations. We also need to correct the state of the qubit for quantum error correction.

While quantum computers do not offer any extra advantages over classical computers in terms of computability, quantum algorithms for specific problems exhibit substantially lower time complexities than their corresponding classical algorithms.

References

[1] Feynman, R.P. Simulating physics with computers. Int J Theor Phys 21, 467–488 ,1982. https://doi.org/10.1007/BF02650179

[2] Moore, Gordon E.. "Cramming more components onto integrated circuits" (PDF). intel.com. Electronics Magazine. Archived (PDF) from the original on 2019-03-27. Retrieved April 1, 2020.

[3] Shukla Prashant, A Study on the basics of Quantum Computing, Book

[4] T.M. Forcer, A.J.G. Hey, D.A. Ross, and P.G.R. Smith. Superposition, entanglement and quantum computation. Quantum Information and Computation, 2(2):97–116, 2002.

[5] Schrödinger, Erwin. "Die gegenwärtige Situation in der Quantenmechanik (The present situation in quantum mechanics)". Naturwissenschaften. 23(48): 807–812, 1935. Bibcode:1935NW..23..807S. doi:10.1007/BF01491891. S2CID 206795705.

[6] N. Bohr. Can quantum-mechanical description of physical reality be considered complete? Physical Review, 48(8):696–702, 10 1935.

[7] Robert H. Lavroff, Doran L. Pennington, Ash Sueh Hua, Barry Yangtao Li, Jillian A. Williams, and Anastassia N. Alexandrova, Recent Innovations in Solid-State and Molecular Qubits for Quantum Information Applications, The Journal of Physical Chemistry A 125 (44), 9567-9570, 2021. DOI: 10.1021/acs.jpca.1c08677

[8] Preskill, John. Quantum computing and the entanglement frontier, 2012. arXiv:1203.5813

[9] Preskill, John. "Quantum Computing in the NISQ era and beyond". Quantum. 2: 79, 2018. doi:10.22331/q-2018-08-06-79

[10] Arute, F., Arya, K., Babbush, R. et al. Quantum supremacy using a programmable superconducting processor. Nature 574, 505–510 2019. https://doi.org/10.1038/s41586-019-1666-5

[11] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser , A Quantum Adiabatic Evolution Algorithm Applied to RandomInstances of an NP-Complete Problem, 2001

[12] Ivan L. Chuang and Michael A. Nielsen, Quantum Computation with Qubits, 2000.

[13]John Martinis, Sergio Boixo, Sergei V. Isakov, et al., A Blueprint for Demonstrating Quantum Supremacy with Superconducting Qubits,2017.

[14] Andris Ambainis, Janos Kempe, and Alexander Rivosh (2003), Quantum Walks on Graphs.

[15] Helmut G. Katzgraber, Federico M. Spedalieri, and Ivan L. Chuang , Quantum Annealing for Clustering, 2014.

[16] Edward Farhi and Jeffrey Goldstone, Quantum Computing for Finance: Overview and Prospects, 2014

[17] Thomas G. Wong, Artur Ekert, and Jonathan J. O'Brien, Quantum Image Processing , 2016

[18] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd , Quantum Machine Learning, 2017.

[19] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, and A. Melville, Quantum Annealing with More Than One Hundred Qubits, 2021.

[20] Mishra, Nimish, Kapil, Manik, Rakesh, et al. Quantum Machine Learning: A Review and Current Status .2019. 10. 13140 /RG.2.2.22824.72964

[21] McArdle, S., Endo, S., Benjamin, S., & Yuan, X. Quantum computational chemistry. ArXiv, 2018. https :// doi.org /10.1103/ RevModPhys.92.015003

[22] P. -H. Wang, J. -H. Chen, Y. -Y. Yang, C. Lee and Y. J. Tseng, "Recent Advances in Quantum Computing for Drug Discovery and Development,",in IEEE Nanotechnology Magazine, 2023. doi: 10.1109/MNANO.2023.3249499.

[23] I. Giroti and M. Malhotra, "Quantum Cryptography: A Pathway to Secure Communication," 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2022, pp. 1-6, doi: 10.1109/CSITSS57437.2022.10026388.

[24] S. S. Tripathy, N. Koul and H. Patel, "Comparative Study between Quantum and Classical Methods: Few Observations from Portfolio Optimization Problem," International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT), Pune, India, 2022, pp. 1-5, doi: 10.1109/TQCEBT54229.2022.10041475.

[25] R. Rana, R. Thingbaijam, J. Seshadri, P. Shah, A. Sinha and S. Poojary, "Quantum Powered Employee Transport and Agri-Logistics Optimization," 2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT), Pune, India, 2022, pp. 1-6, doi: 10.1109/TQCEBT54229.2022.10041701.

[26] Tennie, F., & Palmer, T. , Quantum Computers for Weather and Climate Prediction: The Good, the Bad and the Noisy., 2022 ArXiv. /abs/2210.17460

[27] Alan Ho, Jarrod McClean, Shyue Ping Ong, The Promise and Challenges of Quantum Computing for Energy

Storage, Joule, Volume 2, Issue 5, 2018, Pages 810-813, ISSN 2542-4351, https://doi.org/10.1016/j.joule.2018.04.021.
[28] Andreas Bayerstadler, Guillaume Becquin, Industry quantum computing applications Quantum Technology and Application Consortium – QUTAC, EPJ Quantum Technol., 8 1 (2021) 25. https://doi.org/10.1140/epjqt/s40507-021-00114-x

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-31

A Review: Medicative Flower Classification using Neural Network Classifier

Shubhangi S. Bhangale
Assistant Professor,
School of Computer Sciences, M.J.College,Jalgaon

Dr. Gouri M. Patil
Assistant Professor
Department of Computer Science,
Bhusawal Arts, Science and P. O. Nahata Commerce College, Bhusawal

ABSTRACT - *Classification of flowers is a difficult task because of the huge number of flowering plant spices, which are similar in shape, color and appearance. Identifying each of them requires a botanist with immense knowledge and skills. In this rising era of technologies most of the impossible are made possible by incorporating artificial intelligence into real world problems. By introducing machine learning algorithms such as convolutional neural networks for identifying flower species with just an image would be a great help for industries like pharmaceuticals and cosmetics. Ancient Ayurveda recognizes several medicinal benefits in most of the flowers. For classification different Convolutional neural network architectures are present such as MLCP, VGG-16, AlexNet and many more. This paper reviews the classifier architectures for recognition of medicative flowers which gives more accuracy.*

Keywords:
Convolutional neural network, MLCP, VGG-16, AlexNet.

I. Introduction

Due to the wide variety of flowers and complex natural environment the study of recognition of medicative flowers image is complicated. Flower classification has always been a challenging subject in the field of image recognition. Since flower images are fine-grained images with large intra-class differences and high inter-class similarity, it brings great challenges to their classification. With the rapid development of artificial intelligence technology, machine learning algorithms based on convolutional neural networks have begun to gradually replace manual methods for image recognition and classification tasks.

In traditional method raw images should be transform into the suitable format in which machine can easily extract handcrafted features such as a color, shape and texture while in deep learning network raw images can be fed to neural network directly without doing much preprocessing so these methods are efficient in recognition application.

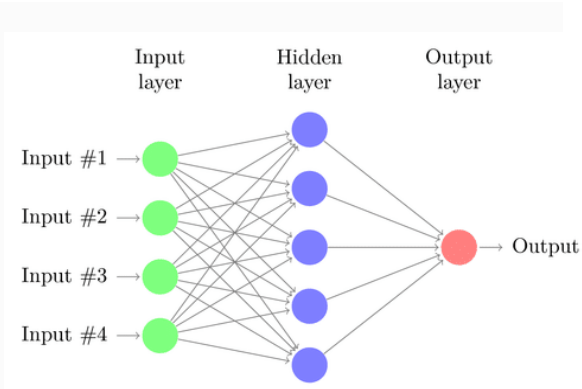
In the field of pharmaceutical industry, botany and agricultural there is a need of algorithm which will correctly classify the medicative flowers by processing its image since incorrect recognition leads to adverse effect. In this context, the literature review is done to study Classification techniques to identify medicative properties of flowers. Most of the researcher frequently used convolutional neural

network as classifier for the identification of flowers, plants.

II. Overview of Neural network:

A neural network is a method in artificial intelligence that teaches computers to process data in a way that is inspired by the human brain. It is a type of machine learning process, called deep learning, that uses interconnected nodes or neurons in a layered structure that resembles the human brain.

Simple neural network architecture



A basic neural network has interconnected artificial neurons in three layers:

1) Input Layer: Information from the outside world enters the artificial neural network from the input layers. Input nodes process the data, analyze or categorize it, and pass it on to the next layer.
2) Hidden Layer: Hidden layers take their input from the input layer or other hidden layers. Artificial neural networks can have a large number of hidden layers. Each hidden layer analyzes the output from the previous layer, processes it further, and passes it on to the next layer.

.

IV. Literature Review:

They considered color and GIST features for classifying the input image. A model which gave 85.93% accuracy was developed using SVM technique. A comparison of various models such as SVM, Random Forest, KNN and multi layer perceptron in flower image classification was done[1].

They proposed a textural feature-based method to categorize flower images. They analyzed the efficiency of their method on 35 types of flowers. They could get a maximum result of 79% [2].

II. Output Layer:

The output layer gives the final result of all the data processing by the Artificial neural network. It can have single or multiple nodes. For instance, if we have a binary (yes/no) classification problem, the output layer will have one output node, which will give the result as 1 or 0. However, if we have a multi-class classification problem, the output layer might consist of more than one output node.

Most Common Types of neural networks:

Artificial neural networks can be categorized by how the data flows from the input node to the output node. Below are some examples:

- Perceptron
- Multilayer Perceptron
- Convolutional Neural Network
- Recurrent Neural Network
- Long Short-Term Memory Network
- Generative Adversarial Network

Most popular machine learning method is deep learning, where multiple hidden layers are used in a model to recognize flowers using CNN. Convolutional Neural Networks (CNN) is an advanced technique for image classification. Lots of CNN models have been used for the classification of objects in the images. CNN is trained using profound learning algorithms that have made some enormous achievements in the recognition of large-scale identification methods in the field of machine learning. Recognition of medicative properties of flowers is a combination of both Recognition and Image Classification, as the system must detect a flower in the image as well as recognize which species it belongs to. ANNs are not suitable for images because these networks lead to over-fitting easily due to the size of the images

In this approach they proposed a method based on the transfer learning approaches. They used a pre-trained model of Inception v3 and could get the accuracy of 94% on the oxford 102-flower dataset [3].

In two pre-trained networks of AlexNet and LeNet were tested on flower datasets collected from Flickr, Google, and Kaggle. The modified network of LeNet could achieve a higher accuracy of 57.73% for image classification [4].

Automated method for the recognition of medicinal plants using computer vision and machine learning techniques was developed. Morphological features and shape-based features from the leaves of

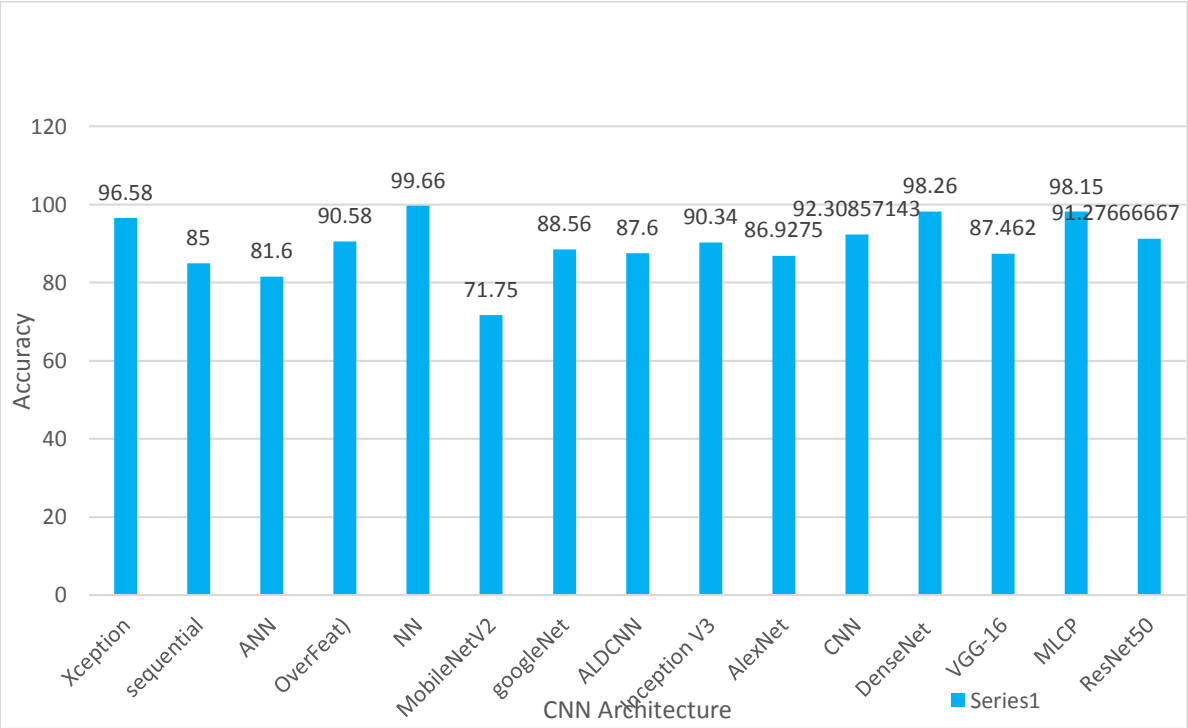
medicinal plants were extracted from each leaf. Random forest classifier obtained the 90.1% accuracy using a 10-fold cross validation technique performed better than other machine learning approaches. [5].

A technique using Convolution Neural Network for identification of rare medicinal plants was proposed. They got 90% accuracy by using TensorFlow on their own dataset which was created from scanned images of leaves and flowers used in Sri Lankan Ayurveda medicine [6].

Summary of literature review based on commonly used CNN architectures for flower classification is given in following Table.

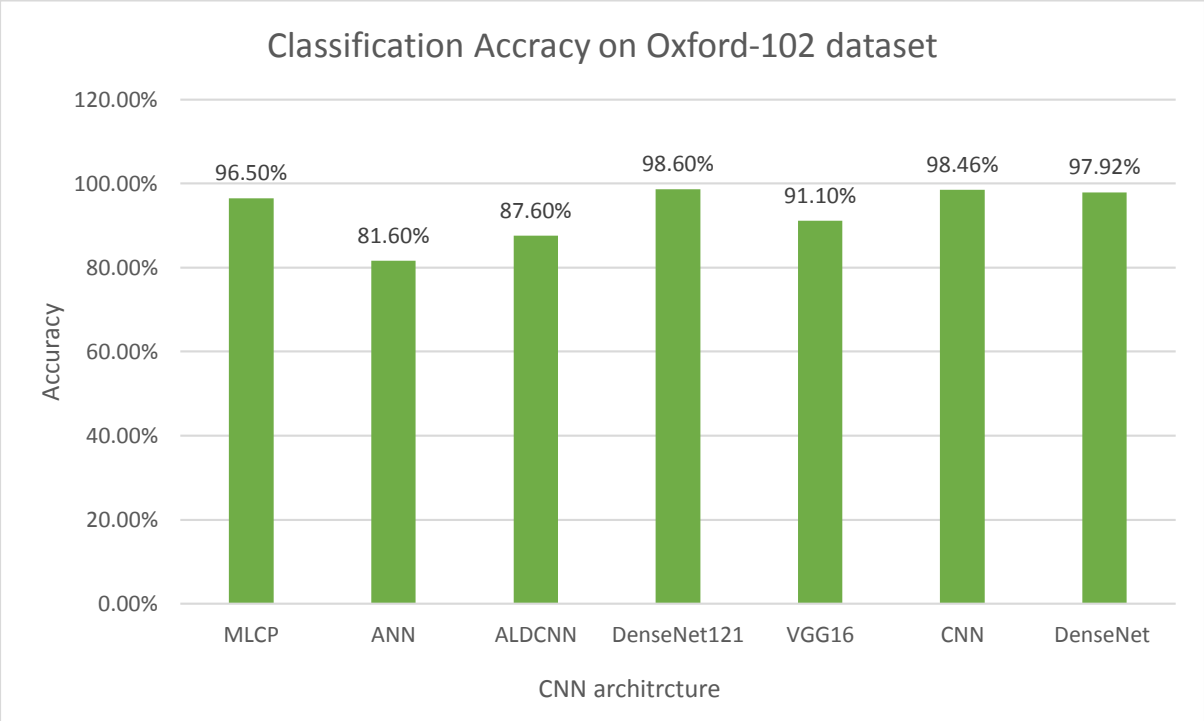
Author	Dataset	Classifier	Accuracy
Mete et. Al.(2019)	Oxford-17	MLCP SVM KNN Random Forest	99.8% 99.6% 99.1% 95.4%
	Oxford-102	MLCP SVM KNN Random Forest	96.5% 98.5% 96.2% 87.6%
Gogul et al. (2017)	Flowers28	Inception-v3 Xception OverFeat	97.68% 96.58% 90.58%
Shukla et al. (2020)	Iris	NN LR SVM KNN	99.66% 99.66% 92.12% 96.67%
Narvekar et al. (2020)	Kaggle	CNN VGG-16 ResNet50 MobileNetV2	91.00% 89.35% 92.12% 71.75%
Qins et al. (2019)	Oxford-102	ALDCNN	87.60%
Shankar et a. (2021)	Own Dataset	sequential Inception V3 VGG-16	85% 83% 81%
Jayalath et al. (2020)	Own Dataset	CNN	98%
Zhang et al.(2019)	Oxford-17	Ramdom Forest	81.37%
Alipour et al. (2021)	Oxford-102	DenseNet121	98.6%
Togacar et al. (2020)	Open-access	AlexNet GoogleNet ResNet50 VGG-16	87.06% 88.56% 89.51% 86.51%
Lv. Rongxin et al. (2021)	Oxford-102	VGG-16	91.1%
Peryanto et al. (2022)	Own dataset	CNN SVM	91.6% 78.3%
Parvathy et al. (2020)	Oxford-102	CNN	98.46%
Desai et al. (2022)	Oxford 17	CNN	91.1%
Kulkarni (2022)	ASIRRA Kaggle Pleura	Alex Net Alex Net Alex Net	99.01% 86.64% 75%
Solanki et al. (2022)	Oxford 102	DenseNet	97.92%

Most commonly used CNN architectures for flower classification



Accuracy of different CNN architectures for classification of Flowers.

From the above graph it is observed that, Neural Network, MLCP and DenseNet, and Xception gives higher accuracy on different dataset of flowers.



Accuracy of different CNN architectures on same dataset (Oxford-102)

Above graph shows that when different CNN architectures are applied on the same dataset (Oxford-102), DenseNet121 architecture gives higher accuracy.

Conclusion:

Recognition of Medicative properties of flowers is very momentous and important research area further accurate medicinal flowers detection and classification is very important. During the review it was observed that various classification techniques were used by many researchers for identification and classification, out of them most of researcher used CNN (Convolution Neural Network) because CNN are able to learn hierarchical features automatically with the goal of classification of image. Convolutional Neural Networks (CNN) is an advanced technique for image classification. There are many variants of CNN architectures have been used for the classification of objects in the images.

Out of the various CNN architectures Neural network, DenseNet, MLCP and Exception architecture can give higher accuracy but it also depends on dataset which is used for classification and feature extraction techniques. Even on same dataset with different CNN architectures gives different accuracy since number of features extracted are different.

References

1. Mete, B. R., & Ensari, T. (2019, October). Flower classification with deep cnn and machine learning algorithms. In 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-5). IEEE.
2. D.Guru, Y.Kumar, and S.Manjunath, "Textural features in flower classification", 54(3-4): p. 1030-1036, 2011
3. X.Xia, C. Xu, and B. Nan., "Inception-v3 for flower classification", IEEE, 2017 [2nd International Conference on Image, Vision and Computing (ICIVC)].
4. K.Mitrović and D. Milošević., "Flower Classification with Convolutional Neural Networks", IEEE, 2019 [23rd International Conference on System Theory, Control and Computing (ICSTCC)].
5. Begue, A., Kowlessur, V., Mahomoodally, F., Singh, U., &Pudaruth, S. (2017). Automatic recognition of medicinal plants using machine learning techniques. International Journal of Advanced Computer Science and Applications, 8(4), 166-175.
6. Jayalath, A. D. A. D. S., Amarawanshaline, T. G. A. G. D., Nawinna, D. P., Nadeeshan, P. V. D., &Jayasuriya, H. P. (2019, December). Identification of Medicinal Plants by Visual Characteristics of Leaves and Flowers. In 2019 14th Conference on Industrial and Information Systems (ICIIS) (pp. 125-129). IEEE.
7. Peryanto, A., Yudhana, A., & Umar, R. (2022). Convolutional Neural Network and Support Vector Machine in Classification of Flower Images. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, *8*(1).
8. Desai, S., Gode, C., & Fulzele, P. (2022, February). Flower Image Classification Using Convolutional Neural Network. In *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)* (pp. 1-4). IEEE.
9. Solanki, A., & Singh, T. (2022). Flower Species Detection System Using Deep Convolutional Neural Networks. In *Emerging Technologies for Computing, Communication and Smart Cities: Proceedings of ETCCS 2021* (pp. 217-231). Singapore: Springer Nature Singapore.
10. Gogul, I., & Kumar, V. S. (2017, March). Flower species recognition system using convolution neural networks and transfer learning. In 2017 fourth international conference on signal processing, communication and networking (ICSCN) (pp. 1-6). IEEE.
11. Alipour, N., Tarkhaneh, O., Awrangjeb, M., & Tian, H. (2021, May). Flower image classification using deep convolutional neural network. In 2021 7th International Conference on Web Research (ICWR) (pp. 1-4). IEEE.
12. Qin, M., Xi, Y., & Jiang, F. (2019, December). A new improved convolutional neural network flower image recognition model. In 2019 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 3110-3117). IEEE
13. Shankar, R. S., Srinivas, L. V., Raju, V. S., & Murthy, K. V. S. S. (2021, February). A Comprehensive Analysis of Deep Learning Techniques for Recognition of Flower Species. In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 1172-1179). IEEE.
14. Kulkarni, A. D. (2022). Deep Convolution Neural Networks for Image Classification.
15. Toğaçar, M., Ergen, B., & Cömert, Z. (2020). Classification of flower species by using features extracted from the intersection of feature selection methods in convolutional neural network models. Measurement, 158, 107703.
16. Lv, R., Li, Z., Zuo, J., & Liu, J. (2021, January). Flower classification and recognition based on significance test and transfer learning. In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 649-652). IEEE.

International Journal of Computer Research & Technology
IJCRT
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-32

Advances In Eeg Technology And Analysis Techniques:A Comprehensive Review

Mr.Ashok A. Pawar
S.S.V.P.S.Science College,Dhule
ashokpawar09@gmail.com

Dr.Rakesh S. Deore
S.S.V.P.S.Science College,Dhule
rakeshsdeore@gmail.com

ABSTRACT - *The electrical activity of the brain is recorded by electroencephalography (EEG) signals. The fundamentals of EEG data capture, signal processing, and analysis are covered. The application of EEG signals in several fields, including clinical diagnostics, cognitive neuroscience, brain-computer interfaces, and sleep studies, is covered in the article. It also discusses the difficulties associated with analysing and deciphering EEG signals, such as artefacts, noise, and individual variability. The study focuses on current developments in EEG signal processing methods, including machine learning and deep learning, which have increased the precision of EEG-based diagnosis and prediction. Lastly, it addresses the potential applications of EEG signals in a variety of domains, such as medicine, neuroscience, and engineering.*

Keywords:
Convolutional neural network, MLCP, VGG-16, AlexNet.

I. Introduction

A non-invasive neuroimaging technique called electroencephalography (EEG) is used to measure and record electrical activity in the brain. EEG signals are produced by neurons firing in unison, which results in minute electrical currents that electrodes affixed to the scalp can detect. The signals produced by these electrodes, which are commonly grouped in arrays, are amplified, filtered, and digitalized for analysis.EEG data offer a window into how the brain works, showing alterations in neuronal activity linked to various mental states, cognitive functions, and emotional states. Because it can record variations in neural activity on a millisecond timescale, EEG is very helpful for researching the temporal dynamics of brain activity.The clinical diagnosis and treatment of neurological conditions like epilepsy and sleep disorders, as well as studies into cognitive and affective processes, are just a few of the many uses for EEG. The creation of brain-computer interfaces (BCIs), which enable people to control external devices using their brain activity, also makes use of EEG.EEG has significant restrictions despite its various uses. Its limited spatial resolution, for instance, makes it challenging to determine the specific site of brain activity. Furthermore, noise and artefacts like muscle activity and interference from the environment can alter EEG signals. Nonetheless, EEG continues to be a useful tool for examining brain activity and has the potential to further knowledge of the human mind.

II. EEG Electrode Design

EEG electrode design is an important aspect of EEG technology, as it determines the quality and accuracy of EEG recordings.A group of 10–20 electrodes are put on the scalp in the conventional EEG electrode configuration. High-density EEG (HD-EEG) electrodes, however, can record brain activity with greater spatial resolution because to recent advancements in electrode technology. Up to 256 electrodes can be used in HD-

EEG systems, which enables the localisation of neural sources and the detection of fine-grained brain activity.EEG electrodes are typically made of metal or a conductive polymer and are placed on the scalp at specific locations to record electrical activity from the brain.

III. Artifact Correction

Muscle activity, eye movements, and external noise are only a few of the noise sources that pollute the EEG signal. These artefacts are eliminated and pertinent data is extracted from the EEG signal using signal processing techniques. Current developments in signal processing approaches have concentrated on the application of blind source separation (BSS) algorithms and deep learning-based artefact reduction techniques.

IV. Feature Extraction

The EEG signal comprises a significant quantity of data, which is reduced to a more manageable set of pertinent features using feature extraction algorithms. The application of time-frequency analysis, connectivity analysis, and graph theory-based approaches has been the focus of recent developments in feature extraction.

V. Machine Learning-Based Classification

The collected features are subsequently utilised to train machine learning models for the classification of various brain states, such as different sleep stages, motor imagery, and epileptic episodes. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which have demonstrated promising results in a variety of EEG-based applications, are two deep learning models that have been the focus of recent breakthroughs in machine learning-based classification.

VI. Conclusion

In summary, current developments in EEG technology and processing methods have greatly aided our comprehension of the brain and its operations. The identification of fine-grained brain activity and the localisation of neural sources have been made possible by high-density EEG electrodes, sophisticated signal processing technologies, and classification algorithms based on machine learning. The creation of more advanced BCI systems as well as new avenues for the detection and treatment of neurological illnesses are now possible because to these developments.

References

1. Bashashati, A., Fatourechi, M., & Ward, R. K. (2015). A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. Journal of Neural Engineering, 12(4), 041006.

2.Gargiulo, G., Babiloni, F., Cichocki, A., De VicoFallani, F., &Valenza, G. (2017). Advancements in EEG signal analysis for brain–computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(8), 1077-1092.

3.Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., &Sejnowski, T. J. (2000). Analysis and visualization of single-trial directions. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(8), 1596-1615.

event-related potentials. Human Brain Mapping, 9(1), 1-8.

4.Lal, T. N., &Debener, S. (2016). Neurophysiology and innovation: Using EEG to study the brain in real-world environments. NeuroImage, 8, 1-5.

5.Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., &Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. Journal of Neural Engineering, 4(2), R1-R13.

6.Wang, Y., Zhang, X., Li, Y., Wang, Y., Zheng, H., Li, X., & Li, X. (2018). A survey on EEG signal processing for BCI: Current status and future

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: LIB-01

Digital and Traditional Library: The Comparative Analysis

Dr. Ashish Vishnu Chaudhari,
Library Incharge,
Bhusawal Art's science and P.O.Nahata commerce college Bhusawal
cashish126@gmail.com

ABSTRACT - *The resources available on the Internet are growing exponentially, leading to a steady increase in Internet use in education and research. Over the past few years, free online resources such as journals, e-books, databases have grown exponentially. In the past, information and knowledge were conveyed orally or by hand, and communication was a slow process. Today, it is transmitted from one person to an unlimited number of other users through multiple media and formats enabling faster and faster data distribution. This paper discusses the use of digital resources by technology and research students. The study also shows that most faculty members learn the skills needed to use digital resources through self-study. Governments in all sectors of the Tribes and Legislatures have been advised to formulate policies and drive bills in their respective order that can advance the current state of the National grid and improve Internet technology infrastructure which is a major factor in digitizing.*

Keywords - *Digital library, Traditional library.*

I. INTRODUCTION

The term 'digital library' means information obtained through electronic sources such as the Internet. Unlike traditional libraries, libraries are not limited by place or time. Libraries have changed with the advent and use of information technology (IT). Libraries take on the role of teachers - teaching users to access, evaluate and use library and electronic data. As e-library usage continues to rise, users are expected to improve their reading skills. The transition from traditional libraries to digital is not just about technological evolution, but it does require a change in the mindset of people who access and interact with it.

The number of different books produced since printing began by no more than one million. (The number of books published annually is less than one million.) If the middle book has 500 pages per 2,000 characters per page, then even if it can be pressed it can be stored properly on onemegabyte. Thus, a billion megabytes is enough to store all the books. This is 1015 bytes, or one petabyte. At a price of $ 20 pergigabyte, this amount of disk storage capacity can be purchased for $ 20million. So it is very possible to think about keeping all the books digitally. A billion-dollar database, each of which occupies one megabyte, is huge but unimaginable. Once one is free of such a size, it is possible to imagine a thousand such details, or to visualize all the releases of the same global collection. This amount of storage is not enough for the house not only all the books, but all of the following: photographs, legal equipment, court decisions, museum items, recorded music, theater, including opera and ballet, lectures, movies and videotape

When information is entered into a computer and is accessible through a network, it is unreasonable to talk about its "location", even though it technically resides on at least one storage device elsewhere, and that device is connected to at least one computer. If the information is available on most of the screen sites, it is less useful to say it is "in place". While traditional libraries measure their size by the number of books, periodicals and other objects, the appropriate digital library statistics are the size of the copy that its users can access. This means that digital libraries will want to increase their "capture" by

sharing digital links with other libraries. Unfortunately, there seems to be very little distribution of this kind happening right now.

How can we understand libraries' dislike of content sharing? The question goes back to the old measure of the size of the traditional library the number of books it holds. When a library spends money to integrate computer-generated services, it loses part of its reputation (or thinks it does) by allowing other libraries to copy or access its own information. Ultimately, however, all material should be available in all libraries.

In 1998, Deputy President Gore said, "A new wave of new technologies allows us to capture, store, process and display unprecedented information about our planet and a variety of natural and cultural contexts. I believe we need a 'Digital Earth.', where we can embed many of the specifics of the geo "(Gore 1998). He then asked scientists to create a digital world map with a one-meter solution. Such a project will require the development of technology in addition to what is required and a digital library containing all the books ever written. The area of the earth by a square meter is approximately 5 x 1014. Storing two megabytes of data per square meter (which may include location data, photos, landscape and other relevant information) will require 1018 bytes, the equivalent of the current digital storage value on earth

Objectives of the Study:

1. To study the Digital and Traditional Libraries
2. To compare the usage pattern of Digital and Traditional Libraries

Hypotheses of the Study:

H_o- There is no significant difference among the usage of Digital and Traditional Libraries

H_1- There is significant difference among the usage of Digital and Traditional Libraries

II. SCOPE OF THE STUDY:

This study is based on the using of the faculty members and research scholars of Jalgaon District. The scope of the study is limited to the use of digital as well as traditional resources and to fulfil the teaching and research purpose of the faculties. It covers the availability of digital resources and services in Jalgaon District.

In the fast changing world; academic institutions are now responding to these global changes by adopting institutional digital repositories in their various institutions.

III. RESEARCH METHODOLOGY OF THE STUDY:

The study is based on critical evaluation and analysis of basically Primary Data. The primary sources include faculty members and research scholars of Jalgaon District. A study is undertaken in the sampled regions to see its impact for which a detailed questionnaire is prepared to collect relevant information from the primary source for the guidance of the researchers. With the help of the questionnaire, detailed discussions were made with the certain sources of primary data to understand their views, thinking and attitude which would help to give the researchers useful recommendations, if any. The questionnaire is processed with the help of statistical tools like tabulations, grouping, percentages, averages, testing of hypothesis etc.

As far as usage of Digital and Traditional Libraries are concerned, following factors are taken into consideration viz, access speed , Difficulty in finding relevant information , too long to view/ download pages , Too much information retrieved , Difficulty in using digital resources due to lack of IT knowledge , Limited access to computers , unavailability of the required reference at the time of need, maintenance of traditional library etc

Research Area

Researchers selected faculty members and research scholars from colleges from Jalgaon District. Sample sizes of 150 faculty members and research scholars from Jalgaon District. Researcher collects data through Primary and Secondary sources. Researcher distributed 150 questionnaires among the respondents.

IV. REVIEW OF LITERATURE

Library Resources' Digitization and Its Impact on the Services of Academic Libraries: The Case of John Harris Library, University of Benin by Oghenevwogaga David in International Journal of Education and Evaluation ISSN 2489-0073 Vol. 3 No. 7 2017: This research investigated the impacts of library resources digitization on the services of academic libraries using John Harris Library, University of Benin, Benin City. As part of the study, research questions were formulated so as to achieve the aim and specific objectives of the study; also various literatures of scholars and authors in the field were reviewed. In order to achieve the purpose and objectives of this study, questionnaire was designed to elicit information from the respondents. The sample size for the study was forty-seven (47) library staff (professional and para-professional) which is 100% of the total population. After the total collection of data, critical analysis of the study was carried out and the major findings revealed the following: the major reasons for the digitization of library resources at John Harris Library include the need to preserve library resources for long use and to have better search and retrieval facilities for library materials; that the main benefits of digitizing library resources are that digitization enables greater access to collections of all types and give the ability to search for library resources electronically without difficulties among others; that the impacts of digitized library resources on the services of academic libraries include helping

to offer more online services to library users and providing quick and easy methods of delivering services to them and that that there are many challenges facing digitization of library resources in academic libraries. These include deterioration of digital media, funding, inadequate technology infrastructures, technological obsolescence and constantly changing hardware and software amongst others. Finally, the study was concluded that despite the tremendous benefits accrued to digitization, both to library staff and users as well as library services, the challenges facing digitization processes in academic libraries makes it a discouraging process.

A Comparative Study of use of the Library and the Internet as Sources of Information by Graduate Students in the University Of Ghana by Cynthia H. Kumah in Library Philosophy and Practice (e-journal). 1298: This study was conducted to compare Internet use and library use among graduate students. It was based on the assumption that graduate students use the Internet more than the library. Literature on library and the Internet were reviewed. The researcher adopted convenient sampling technique to select the sample for the study. Data was collected using questionnaires. Collected data was statistically analysed and interpreted using the Statistical Package for the Social Sciences. Findings of the study indicated students do not bypass the library in satisfying their information need. They use both the library and the Internet, although Internet usage was more than the library, hence the Internet was the most preferred source of information. It was recommended that the library should be upgraded to meet recent advancement in research.

Limitations of the study

1. The study is based on limited geographical area.
2. Further variables could be added for the purposes of detail study.

V. DATA ANALYSIS

Researcher prepared the questionnaire for respondents and distributed it among them. After receiving the questionnaire researcher analyse the questionnaire.

Table No1: Information of questionnaire

Sr. No	Respondents	Questionnaire distributed	Questionnaire received	Questionnaire rejected (due to incomplete, wrongly filled etc)	Net Sample size for study
1	Faculty members and research scholars	200	195	6	189

Testing of Hypothesis

H_0: There is no significant difference among the usage of Digital and Traditional Libraries

H_1: There is significant difference among the usage of Digital and Traditional Libraries

Mathematically

	DIGITAL LIBRARY	***TRADITIONAL LIBRARY***
Mean	8.08	4.46
Variance	0.75	0.43
Observations	189	189
Pooled Variance	0.59	
Hypothesized Mean Difference	0	
df	376	
t Stat	45.88	
P(T<=t) one-tail	1.9651E-156	
t Critical one-tail	1.65	
P(T<=t) two-tail	3.9302E-156	
t Critical two-tail	1.97	

****Here level of significance is 0.05**

Thus, our null hypothesis *"There is no significant difference among the usage of Digital and Traditional Libraries"* is rejected. Alternatively we accept our alternative hypothesis *"There is no significant difference among the usage of Digital and Traditional Libraries"*

VI. FINDINGS

1. The important issue about use of Digital and Traditional Libraries among the respondents' perception is that, they have started to get in flow with digital formats as compared to traditional one which helps library systems to get reenergised
2. The additional point is came out from this analysis in which respondents were responding about maintenance of traditional library.

VII. CONCLUSION

From the above analysis, we can conclude that, in any case, the respondents 'attitudes have changed with regard to the use of digital and traditional libraries. It also suggested that, Awareness should be done through online journals to get the latest information; University administrators should make plans and infrastructure to train its staff in ICT especially on the use of digital resources etc.

REFERENCES

1. Library Resources' Digitization and Its Impact on the Services of Academic Libraries: The Case of John Harris Library, University of Benin by Oghenevwogaga David in International Journal of Education and Evaluation ISSN 2489-0073 Vol. 3 No. 7 2017
2. A Comparative Study of use of the Library and the Internet as Sources of Information by Graduate Students in the University Of Ghana by Cynthia H. Kumah in Library Philosophy and Practice (e-journal). 1298
3. S. C. Gupta and V.K.Kapoor (2006). Fundamentals of Mathematical Statistics, Sultan Chand & Sons Educational Publishers, New Delhi
4. S. C. Gupta and V.K.Kapoor (2006). Fundamentals of Applied Statistics, Sultan Chand & Sons Educational Publishers, New Delhi

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-33

Covid Patient Health Monitoring Quarantine Based using IOT using AIML

Hiralal Bhaskar Solunke
hiralal.solunke@sandipuniversity.edu.in

Vipin Y. Borole
vipinborole@sandipuniversity.edu.in

Vaibhav P. Sonaje
vaibhav.sonaje@sandipuniversity.edu.in

School of Computer Science & Engineering, Sandip University Nashik, Maharashtra, India

ABSTRACT - *Medical Surveillance results are the most important in the brief developing country crowd enhances demands for caregiving. Covid- 19 is as a cover contagious it's veritably important to counter blockade covid19 humans but at the equal time medical observers need to check fitness of covid- 19 victims also. With the boosting kind of cases, its long hauls turning into tough to save a tune on the health and well- being problems of several quarantined humans. Below the encouraged machine design of a Wi- Fi detector network grounded upon IoT invention. It's typically used for accumulating as well as moving the unique detectors tracking data regarding the mortal beings in healthcare centres. This software consists of Wireless primarily grounded network (Wi- Fi), having absolutely exceptional seeing widgets related to the transmitter area the bones are Heart beat seeing unit, Temperature stage seeing unit blood pressure detector and palpitation oximeter. These detectors are straight down connected to the affected man or woman and accumulate the customer issues by using way of the use of seeing bias. The same statistics is transferring out wirelessly to the receiver position this is with the medical agent and by way of that receiver trouble he will surely reap all updates in their guests. And also, it's going to authentically deliver voice word to humans to take their drugs suitable time. And one sharp buzzer will in fact there at case with a view to nearly endorse exigency state of affairs of guests. When case will press exigency button also the buzzer will be ON.*

Keywords - *oximeter, wifi, IOT, Heart beat sensor, medical data.*

INTRODUCTION:

Health is constantly a primary worry in every growth the mankind is advancing with reference to era. Like the modern coronainfection attack that has wrecked the economic state of affairs of China to a quantity is an instance precisely how health care has grown to be of predominant significance. In such locations wherein the epidemic is spread out, it's miles typically a far higher concept to screen those patients theusage of far-off health and fitness tracking generation. So, Internet of Things (IoT) primarily based health tracking device is the existing choice for it. Remote Person Monitoring affiliation encourages statement of customers beyond ordinary medical setups (e.g., In your own home), which broadens get entry to to human services workplaces at lower charges. The middle reason of this method is the design in addition to executionof a smart man or woman health and health tracking machine that utilizes Sensors to trackaffected character health and well-being and also makes use of web to educate their preferred ones in case of any sort of issues. The objective of setting up monitoringsystems is to restriction healthcare costs by using minimizing scientific expert workplace goes to, clinic stays, similarly to diagnostic finding out gadget each of our bodies uses temperature degree as well as additionally pulse acknowledging to read comprehendingwellness. The sensing devices are related to amicrocontroller to music the reputation that is therefore interfaced to a LCD show and also further a ways off association with have the functionality to trade alarms. If shape locatesany unexpected adjustments in comprehending coronary heart beat or body temperature, the structure finally upsets the purchaser approximately the patient's condition over IOT and moreover shows diffused elements of pulse in addition to temperature of consumer live in the net. In this way IOT set up tolerant wellbeing adhering to form viably uses net to display silent health dimensions and further lingers time. There is a sizable capability among

SMS primarily based absolutely individual thriving

viewing and additionally IOT based definitely man or woman inspecting framework. In IOT primarily based framework, subtle elements of the consumer thriving may be seen with the useful resourceof several clients. The description at the backof that is the info need to be tested by way of way of passing with the useful resource of a internet site or URL. While, in GSM based definitely affected person seeing, the thriving parameters are dispatched using GSM via techniques for SMS. In most of the agricultural regions, the medical middle might no longer be in a hand reap distance for the natives. So commonly people forget any type of shape of teenybopper fitness problemsthat's displayed in onset via variant of vital components like frame temperature degree, heartbeat fee and many others. When the fitness hassle has been raised to a vital component and the lifestyles of the character is endangered, after that they take clinical help, which could motive an unneeded waste in their revenues. This likewise enters account specifically whilst precise epidemic is unfolded in a place in which the achieve ofdoctors is hard. So, to keep away from the unfold of sickness, if a clever sensor is obtainable to human beings, who may be stored track of from a distance might be a practical treatment to store loads of lives. Thehealth care services the Internet of Things (IoT) has been widely applied to interconnectavailable medical resources and provide reliable, effective and smart healthcare service to the elderly and patients with achronic illness.

LITERATURE SURVEY:

Internet of Things (IoT) development brings new possibilities in many programs, together with smart towns and smart healthcare. Currently, the primary utilization of the IoT in health care can be labelled as remote monitoring and additionally real-time health structures. Regulating and managing alarming activities, which includes the best in 2020 while the corona virus situation (COVID-19) took manager of the globe, may

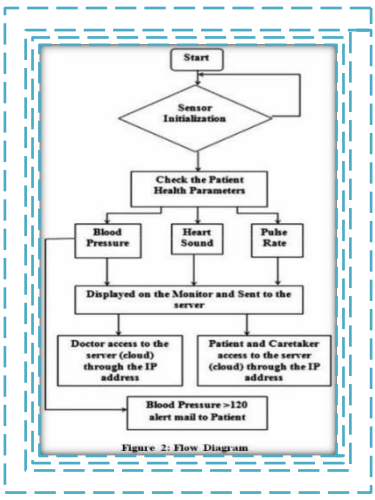


Figure 2: Flow Diagram

be completed with the help of IoT systems, without implementing crucial regulations on human beings in addition to sectors. COVID-19 reasons breathing signs and symptoms andadditionally seems more infectious in assessment to SARS in 2003. One method to manipulate the spread of viruses, up until a vaccine is to be had, is to examine physical (or social) distancing. By enforcing better systems for surveillance, healthcare, and transportation, infectious contamination maywant to have an entire lot tons less opportunity of spreading out. An IoT device,included with Expert system (AI), may additionally use the adhering to contributions even as considering an endemic: 1) improving public protection and security making use of protection and photo acknowledgment structures, 2) utilizing drones for supply, shipping, or disinfection,contact mapping further to proscribing human being's get proper of access to public places via apps in addition to systems encouraged with AI. An IoT system for healthcare is typically composed of lots of sensing gadgets related to an internet server; it presents actual-time surveillance of a putting or customers. In a virus, AI assisted sensors may be used to resource are watchingfor whether or not or now not or now not people are infected with the infection, primarily based on signs and symptoms consisting of frame temperature, coughing styles, and additionally blood oxygen degrees. Tracking human's relocation may be an additional useful characteristic. During the outbreak of a transmittable scenario, tracking the range in among people can supply vital info. Using era, together with Bluetooth, we can acquire an affordable estimate of how heaps range humans keep while walking in public places. These statistics may be made use of to alert individuals who aren't actually distanced interior an info array, 2 m for example, of an character, in addition to consequently, possibly forestall similarly transmission of the contamination. During the improvement of such structures, it's far additionally important to bear in mind safety and additionally records tracking notably to stop misuse of man or woman data Governments may also attempt to utilize those systems and facts for irreversible surveillance after a plague to alter and additionally tune people's behaviour.

Independent in addition to trouble- trouble-lose, wholesome and balanced dwelling is the goal of any type of human being irrespective of their age, intercourse, and place or health reputation. Nevertheless, there are limitations because of age, disease, medication, a medical institution live, epidemic, and pandemic further to special situations. Health and properly-being surveillance systems have surely developed to assist convenient wholesome and balanced dwelling, extra without difficulty available conversation between healthcare organizations further to human beings for near tracking, dimension of vital fitness requirements,

ordinary exam in addition to common healthful dwelling. Additionally, with the recent advances in information and interplay innovations (ICT) with the adoption of Net of Things (IoT) contemporary technology, clever health and well-being monitoring and also assist structures now have a greater facet of development and acceptability for extra appropriate healthful residing.

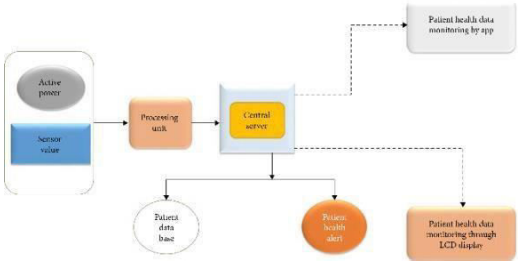
The studies finished through Zikali, discovered out that with the fast boom within the populace of older or senior residents, customers that need well-being surveillance have certainly likewise raised greatly. The equal research takes a look at predicts that by using manner of the twelve months 2045 the type of senior residents who're considered the maximum prone in way of existence will surpass the number of children further to more youthful people as a contemporary populace census indicates a rise in older humans. However, a lack of house health andnicely-being helpers, nursing aides and additionally residence hospital therapy giversare looming international, that makes deal with the aged luxurious. As a give up end result, a fitness and wellbeing tracking system can play a crucial function in lessening physical get in touch with, a medical institution live, appointment time, queuing listing and overall health rate for a patron at the identical time as likewise reducing artwork, fear as well as strain and anxiety on medical personnel. Advancements in information and verbal exchange upgrades for connectivity anywhere in addition to on every occasion make an important contribution to the development of the currenthealthcare machine used in telemedicine answers and moreover diverse extraordinary portable scientific systems.

The arrival of smart house innovations recommends healthful dwelling in addition to stepped forward top incredible of healthcare help answers for the aged as well as handicapped for impartial as well as comfortable way of lives while in theresidence, as opposed to nursing homes, hospitals or distinct arrest facilities. The hospital therapy thing, as part of the smart residence automation machine, will beautify medical institution treatment facilities for individuals whilst at home or in distant places outside the healthcare centres. Therefore, there can be a lower in anxiety that emerges from solitude in the medical institution wards for patients. The doctors can test people fromtheir place of business, prescribe drug in addition to sight determined important health specifications for a remote scientific diagnosis. Likewise, the fast enhancement of software program software program and

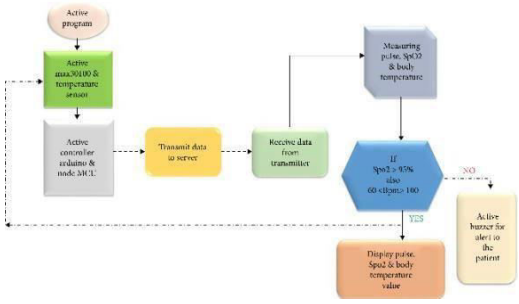
system modern-day technology within the smart house healthcare gadget, makes it viable for humans, in particular the senior or impaired, to manage particular house devices quite absolutely from equipment together with smart phones, pill laptop structures, laptops, net, etc.

METHODOLOGY:

When the cycle stream was created, it directed essential periods of any futureactivities from the beginning to the end of a system. Circuit diagrams were utilized for the planning, development, and support of electrical and electronic gears. For a well-developed system, these diagrams were truly significant. Figure 1 shows a block diagram of the proposed system. The system block diagram shows that when the power of the system is switched on, the sensor starts taking values. Here, the system has two types of sensors for measuring SpO2, pulse rate, and temperature.



The sensors measure the physiological data from a human body and then pass the analog values to the Arduino, which converts them into digital data. The server sends the measured data to the mobile application and displays the data through an LCD display simultaneously. From the mobile application and device, users can monitor the temperature level, oxygen saturation (SpO2), and pulse rate.



When the power of the device is activated, it starts measuring values and sendsit to the main controller, Arduino Uno, and Node MCU. The Node MCU transmits the measured value to the fixed server. The system then displays measured temperature, pulse rate, and SpO2; if the measured oxygensaturation is below 95% and the pulse rate is less than 60 or greater than 90, the system sends an alert to both the doctor and the patient. Users can see the measured value through a mobile application, and simultaneously, users can see the value through the device's LCD display.

Arduino Uno, node MCU, Bluetooth module, SpO2 sensor, temperature sensor, and power supply of the system. This circuit diagram is designed using the Proteus Design Suite software. The active power button turns on the fully automated system. The sensors measure the data and send them to the

processing unit, after which the data will be available on the mobile application as discussed earlier.

The system consists of two parts: the equipment and the mobile application. Both parts are fundamental to the system. The health monitoring system can measure oxygen saturation, pulse rate, and human body temperature. This multifunction system requires several components to be implemented. Implementation is achieved by performing activities portrayed in a work arrangement. For making the system successful, design implementation plays an important role.

CONCLUSION:

In this challenge, an Iota frame is offered to check people' health and fitness problems as well as notify them to keep fleshly distancing. The suggested device incorporates a wearable Iota knot with a cell phone software program, via which the Iota detector knot can collect someone's health norms, inclusively with temperature stage and also blood oxygen achromatism, and the cellular smart phone attaches to the network to transport out the statistics to the net garçon. The paper encouraged a Super high frequence distance- shadowing system which runs every for innards as well as outside surroundings to warn individualities to maintain the physical distancing. Applying ML factors on body specifications makes it feasible to check player' s' heartiness situations and to warn people in real time. A voice coughing- sensor always maintains track of the patron's voice and documents the variety and also inflexibility of coughing. The fog- grounded completely garçon is carried out to reuse gotten information from an Iota knot the operation of a cell community or woman hyperlink. In addition, regionally enriching the data makes it possible to make use of the Iota knot inside the atmospheres without net connection or fog- grounded completely networks. The contrivance can help members in covering their regular scores in addition to lessen the threat of hype to the Corona contagion.

REFERANCE:

1. Olutosin Taiwoa,, Absalom E. Ezugwub, "Smart healthcare support for remote patient monitoring during covid-19 quarantine", Information in Medicine Unlocked, volume.20,2020, pp.100428.

2. Mirza Mansoor Baig, Hamid Gholamhosseini “Smart health monitoring systems: an overview of design and modelling”, j Med Syst, volume. 37, 2013, pp.9898.

3. Katharina Rasch, Fei Li, Sanjin Sehic, Rassul Ayani ,Schahram Dustdar, “Context-Driven personalized servicediscovery in pervasive environments”, world wide web, volume. 14(4), 2011,pp.295-319.

4. Aleksandar Kotevski, Natasa Koceska, Saso Koceski, “E-health Monitoring System”, International Conference on Applied Internet and Information Technologies ,2016,pp.259-265.

5. Sparsh Agarwal, Chiew Tong Lau,“Remote Health Monitoring Using Mobile phones and web services”, telemedicine and e-health journal, volume.16(5),2010,pp.603–607.

6. Prosanta G., Tzonelih H., “BSN-care: a secure IoT-based modern healthcare system using body sensor network”,IEEE Sensors journalvolume. 16(5),2016, pp.1368–1376.

7. Minh P., Yehenew M., Ha D., Weihua S. “Delivering home healthcare through a cloud-based smart home invironment”, volume.81, 2018, pp.129–140.

8. Eren D., Erdem K., Radosveta S., Burhan Ş.“Smart home assistant for ambient assited living of elderly people with dementia”, International Workshop on IoT, M2M and Healthcare (IMH2017), volume.113, 2017, pp. 609-614

9. Bilal Ghazal, Khaled A1- Khatib,“Smart home automation system for elderly and handicapped people using XBee”, International Journals of smart Home, volume.9(4),2015, pp. 203-210.

10. Md. Raseduzzaman Ruman, Amit Barua, Waladur Rahman, Khan Roushan Jahan, Md. Jamil Roni; Md. Foyjur Rahman, ”IoT Based Emergency Health Monitoring System”, International Conference on Industry 4.0 Technology (I4Tech) IEEE, 2020, pp.159-162.

11. Zahir bin Sulaiman Al Brashdi1, Shaik Mazhar Hussain, Kamaluddin Mohammad Yosof, Shaik Ashfaq Hussain, Ajay Vikram Singh, ” IoT based Health Monitoring System for Critical Patients and Communication through Think Speak Cloud Platform”, International Conference on Reliability,Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) IEEE, 2018, pp.652-658.

12. Kovuru Chandu Chowdary, K. Lokesh Krishna, K Lalu Prasad, K. Thejesh, “An Efficient Wireless Health Monitoring System”, Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2018) IEEE Xplore, 2018, pp.373-377.

13. Jose Reena K, R. Parameswari, “A Smart Health Care Monitor System in IoT Based Human Activities of Daily Living: A Review”, International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) IEEE, 2019, pp.446-448.

14. Ashwini Gutte, Ramkrishna Vadali, “IoT Based Health Monitoring System Using Raspberry Pi”, Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) IEEE, 2018.

15. Alvee Rahman, Tahsinur Rahman, Nawab Haider Ghani, Sazzad Hossain, Jia Uddin, “IoT Based Patient Monitoring System UsingECG Sensor”, International conference on Robotics, Electrical and Signal Processing Technologies IEEE, 2018.

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-34

A Study of different application domains of GIS

Ms.Rupali Narkhede
Asst.Professor
MCA Department,
I.M.R.,Jalgaon
rupalinarkhede@gmail.com

Prof.Dr.Varsha M.Pathak
Professor
MCA Department,
I.M.R.,Jalgaon
Varsha.pathak@imr.ac.in

ABSTRACT - *This paper is mainly the study of GIS system in different application domains like Environment management, Tourism management, Marine fisheries management, Public health management, Real Estate application, Military services, Forest management, Traffic management etc. More GIS applications are developed for the specific purpose such as use by military or companies. GIS is an effective information tool to deal with spatial data and complex interactions in a particular city. GIS is found as decision-making support tool in mine development. GIS is the fastest growing area but also an integrated technology bringing together information, systems, applications and people from different fields.*

Keywords: GIS, Spatial data management, NLP, GeoAI

Introduction

The GIS (Geographic Information System) has potential to design and optimize in terms of Spatial data management, decision-making support and multi-parameter consideration. An oftenly quoted definition[5] of GIS in the literature has come from Burrough (1986). Burrough in 1986, stated that "GIS is a powerful collection of tools for gathering, storing, retrieving at will, transform and presenting spatial data from the real world"[20]. GIS is a system for making a map and an analysis tool. The geographical or spatial data represent fact from the real world in terms of their position with respect to a known coordinate system [2]. GIS integrates all kinds of information and applications with a geographic component into one manageable system. The integration capability of GIS technology allows organizations to make better and more informed decisions based on all associated factors.

GIS system is also useful to bring out information on the geographic context of documents using NLP (Natural Language Processing) that can help to identify place and location properties[3]. The GIS technology greatly improves the efficiency and quality of work with spatially distributed data as compared to traditionally paper cartographic methods [19]. GIS is characterized by large variety of applications. It is used in industry, government, emergency services, medicine, ecology and environmental management, scientific research, business, education that is in all areas where data research is needed.[4]. GIS is having it's toolkits which bring out to manipulate and analyze the data spatially and visualize data in the form of maps, tables, graphs, animations and virtual landscapes.

Geographic data refers to data about earth's surface and the objects found on it. This information comes in three basic forms: spatial data, tabular data, and image data. Spatial data contains the location and shapes of map features. Tabular data is gathered and compiled for specific area and is descriptive data that GIS links to map features. Image data includes diverse elements such as satellite images, aerial photographs and scanned data.

1. History and Development of GIS

GIS mainly deals with maps creation which are now more advanced than they have ever existed. In the previous era, cartography was used and continued to develop through conducting various tests in the course

of ensuing centuries, newer technologies also came to develop all together.

In the 1950s, William Garrison a well-known geographer and analyst of transportation issues. He worked at the University of Washington and came up with a brilliant idea to make use of statistics and computers to study and better identify with spatial problems. Waldo Tobler, a student of William Garrison and contestant in the Quantitative Revolution, devised a model called Map In – Map Out (MIMO) in 1959 which was to apply computers to cartography. This model was the basis for GIS and contained all the prime features of GIS software available and in use today. During the 1960s, first, in the year 1960 saw the development of the world's first true operational GIS in Ottawa, Ontario, Canada by the federal Department of Forestry and Rural Development. Developed by Dr. Roger Tomlinson, it was called the Canada Geographic Information System (CGIS)[21]. Tomlinson has become known as the "father of GIS", particularly for his use of overlays in promoting the spatial analysis of convergent geographic data[10]. In the subsequent year, Howard Fisher created the Harvard Lab for Computer Graphics.

At London's Royal College of Art, 1967 saw the development by the CIA of an Auto Mapping System, in 1968 Robert Tweedie based in Albany and working for the New York State Department of Transportation invented a Transportation Information System, in 1969, Jack Dangermond established the Environmental Systems Research Institute (ESRI), Jim Medlock established Integraph Corporation, Laser-Scan was founded by 3 academics from Cambridge's Cavendish Laboratories in UK and the publishing of Ian McHarg's book Design With Nature which popularized overlay map techniques.

In the 1970s; the start of the 70s was marked by four significant events that shaped the history of GIS. The Roger Tomlinson led Canadian project began to make full use of the Canada Geographic Information system (CGIS) developed, E L. Amidon an employee of the Berkley based American Department of Agriculture developed a. grid manipulated Map Information and Display System that was able to tabulate data for single as well as overlaid maps; Arie Shachar, working at the Hebrew University in Jerusalem generated "The Urban Atlas of Jerusalem" by applying the grid manipulation system to a block inventory data bank, Dieter Steiner of Switzerland created the Geographic Mapping Program (GEOMAP) which was grid manipulation based and similar to Howard Fisher's SYMAP and could produce shaded maps; The first ever symposium on GIS took place in Ottawa, Canada. In 1971, Roger Tweedie made use of a database that contained such information as physical attributes of roads, traffic volumes, bridge records, and road inventories to create the Highway Inventory Information System for the state of New York. In 1986, Mapping Display and Analysis System (MIDAS), the first desktop GIS product [13] was released for the DOS operating system. This was renamed in 1990 to MapInfo for Windows when it was ported to the Microsoft Windows platform. This began the process of moving GIS from the research department into the business environment, the end of the 20th century the development

photozincography, which allowed maps to be split into layers, for example one layer for vegetation and another for water .

The Future of GIS: The continuous developments taking place in this field, it is forecasted, will result in the increased usage of this technology in fields like science, industry, government, public health, archeology, urban and rural planning, logistics, transportation, environmental sustainability, defense, etc.

3. GIS Components

GIS consists of mainly 5 components: people, hardware, software, data and methods.

Figure 1[7]

1. Hardware: is the computer on which GIS operates. Today GIS software runs on a wide range of hardware types from centralized computer servers to desktop computers used in standalone or networked configurations.
2. Software: provides the functions and tools needed to store, analyze and display geographic information[22].
3. Data: is possibly the most consequential component of GIS. A GIS will integrate spatial data with other data resources and can even use a DBMS [23].
4. People: develop and manages plans to apply it to real world problems using GIS technology.

5. Methods: GIS operates to a well-designed plan and business rules which are the models and operating practices unique to each organization [24].

A working GIS integrates these five key components.

4. Need and Scope of GIS

GIS is needed to combine spatial and non-spatial data. In GIS , we can use data that has place in the world that is spatial data and combine with it's facts about a place i.e. non-spatial to solve large much problems. We can create data, store data, manage data, analyze data and display data with the help of GIS. While this is possible with the help of paper maps and Excel files, GIS combines the two to solve the problems in a rapid and repeatable way. GIS technology can be used for scientific investigations, resource management and development planning. GIS is used in Computer Science, Civil Engineering, Mathematics, Statistics & Operations Research. With GIS, we can create new approaches that help us to understand the relation between man and the environment. GIS can be thought of as a system, it digitally creates and manipulates spatial areas that may be jurisdictional, purpose or application oriented for which a specific GIS is developed[8].

5. Tools and Techniques

We have studied various tools and techniques that are used in GIS system by reviewing some articles of GIS given below in literature review. In these articles, we studied Forest Atlas Software, Photo Tracker and ArcGIS Tool, MapInfo Software and GPS for different application areas of GIS. The use of these tools and techniques are necessary to retrieve geographic information in emergency cases like flood hazard management, health surveillance or in real estate management or in routing school buses, transportation network etc. GIS tools and techniques play an important role to fetch data related to land, river, the environment and traffic etc.

6. Methodology

We have studied more than 50 related articles to this area and after going through the abstracts and conclusions part of all the articles, we selected 21 articles relevant to this topic under the study Viz. "Different applications domains of GIS ". From these articles, 11 relevant articles are selected for further study. The study is focused to categorize these articles based on their domain areas, applications, software tools and techniques that are used to gain accurate results.

7. Literature Review

Article 1: Sonti SH [9] has introduced the use of GIS technology in forest management which uses the forest atlas software to improve the quality and availability of information in the forest sector to support transparent and participatory decision making across the Congo basin.

Article 2: Vishakha Shukla et al.[10] has focused to find out a solution to ensure the safety of route to parents when their children are travelling to school and back to home. They have used two spatial technical tools, "Photo Tracker" and "ArcGIS", which records and analyzes vehicle routes. By using these devices and tools, we can compare the actually recorded routes and optimal routes.

Article 3: Debashis Das et al.[11] has proposed the road network analysis of Guwahati city using GIS after experiencing major issues like traffic congestion, delay, pollution, increased vehicle operating cost and road accidents. The shortest route is generated between required origin and destination point. ArcGIS tool is used for this analysis. Thus Network Analysis is one of the most powerful tools to deal with the real time transportation problems.

Article 4: Yosoon Choi et al[12] has focused on GIS based methods and applications utilized for mine development.Mine planning operations and environmental management were examined to determine the role of GIS as a decision making support tool in mine development.

Article 5: S. Tamilenthi et al[13] has introduced flood hazard zone study in which damage to life, natural resources and environment and losses to economy and health have occurred. They proposed contour map creation through GIS with the field data by GPS and used ArcGIS 9.2, GPS , and Sufer pack 8 tools for the study. Their study on the mapping of the flood affected zones would serve as a basis for all the above five said elements of flood risk management. The primary purpose of producing these kind of maps is for public dissemination of flood maps which will serve to increase public awareness.

Article 6: [14] GIS in marine fisheries is useful for habitat mapping, species distribution and abundance, oceanographic modellings, fisher's activities and fisheries management.

Article 7: Marilyn O'Hara Ruiz, Arun Kumar Sharma[15] have focused on the public health domain. They stated that GIS system can be used for better surveillance and control of human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS) and various diseases.

Article 8: Wei Wei[16] has focused on GIS system application in tourism management. He presented various problems in existing GIS applications in tourism management. He focused to combine the expert system technology with GIS to enhance the capacity of the system to solve travel problems is studied.

Article 9: Chengda Lin et al[17] presented the concept and technology of GMN and it's applications in developing real estate management software. They have developed real estate software which provides complete solutions for spatial indexing, selecting and clipping data in the perspective space, dynamic loading, office automation etc.

Article 10: J. V. Iyengar[18] has focused on potential uses and applications of GIS. He proposed GIS is a influential tool for military trainers, environmentalists and natural resource planners.

Article 11: Maged N. Kamel Boulos et al[19] focused on GeoAI Tool which has been used to model and capture the environment around us. An emerging role for in health and healthcare as location is an integral part of both population and individual health. GeoAI has been used to design and record the environment around us, linking locations in which we live, work, and spend our time to these exposures.

3. Review Analysis

Article No. & Reference Paper	Domain Name	Applications	Software/Tools used	Results
Article 1 [9]	Forest Management	Improve productivity, to save time, money and man power in forest management activities.	The Forest atlas software	Real-time online data capture and query in the field
Article 2 [10]	Scheduling and routing school buses	Minimize the fear and stress of parents	Photo tracker and ArcGIS tool	Actual recorded routes and optimal routes are compared.
Article 3 [11]	Transportation Network	Road Network analysis is possible	ArcGIS	The shortest route between the required origin and destination points is achieved.
Article 4 [12]	Mine development	Decision-making support tool in mine development	----------	It was observed that GIS-based methods, including spatial database, spatial analysis, spatial mapping, and visualization, can be effectively used at all mine development stages at the global, regional, and mine scales.
Article 5 [13]	Flood Hazard Management	To increase public awareness about flood affected zones to save human life.	ArcGIS 9.2, GPS , and Sufer pack 8.	As data management and map representations tools of GIS helped in exploring new potions its integration with Remote Sensing, enhance the ability for preparing flood hazard map and forecasting. Thus GIS can be very useful to minimize flood hazards.
Article 6 [14]	Marine Fisheries	To locate sites for new marine aquaculture operations, habitat mapping, species	CAD, Digital cartography,	Fisher's activities are tracked, habitat mapping ,marine operations etc. are carried out with the help of GIS system.

		distribution, and abundance, Fisher's activities, etc.	Geostatistics, Photogrammetry, MapInfo Software	
Article 7 [15]	Public Health	GIS system can be used for better surveillance and control of human diseases on time.	----------	Many health and medical communities along with universities have adopted geospatial technologies and methods to address a broad range of health applications.
Article 8 [16]	Tourism Management	GIS plays important role in tourism management & produce a comprehensive thematic map.	RS (Remote Sensing), GPS	Combined the expert system technology with GIS to enhance the capacity of the system to solve travel problems.
Article 9 [17]	Real Estate Application	GIS provides geographic information and 3D visualization capabilities to traditional real estate software systems. GMN has a great prospect on GIS applications, not limited to real estate, but also applicable to building construction, city planning, mineral resource detection etc.	The real estate software	The real estate software we have developed has provides complete solutions for spatial indexing, selecting and clipping the data in the perspective space, dynamic loading, office automation, capabilities for managing large volumes of information as well as real-time applications.
Article 10 [18]	Military Application	GIS is used in military training and environmental/natural resources management	ArcView, AutoCAD	GIS is a influential tool for military trainers, environmentalists and natural resource planners.
Article 11 [19]	Health Domain	GeoAI has been used to design and record the environment around us, linking locations in which we live, work, and spend our time to these exposures.	GeoAI Tools like SAS Visual Analytics	Prediction of disease occurrence is possible due to GeoAI. New data linkages are formed and hypothesis generation is possible.

Table 1.1

8. Results and discussion

The scope of GIS is not limited for one area but it is used in various application domains. In this study,21 relevant articles are reviewed and the observations are noted as above in tabular form 1.1. Our study shows that GIS used mostly in routing and transportation purpose as well as in public health domain. In 2 papers, we found GIS system can be used for better health surveillance and can help to control human diseases in time. In 13 papers show use of GIS observed in scheduling, transportation, tourism and routing purposes. They mainly use algorithms to find the shortest distance and to reach within time to destination. In 4 papers, GIS is used to manage environmental/natural resource management like forest, marine fisheries, military applications and Flood Hazard management. In 1 paper, GIS is used as a decision-making tool in mine development. In 1 paper, GIS is also used in real estate management as it provides geographic information and 3D visualization capabilities to traditional real estate software system. Various soft wares/Tools are used for GIS system in various domains.

Domain-based analysis of GIS system

Domains / Article No.	Health Domain	Scheduling, Routing, Transportation	Environmental\Natural Resource Management	Decision-Making Tool	Real Estate Management
1			✓		
2		✓			
3		✓			
4				✓	
5			✓		
6			✓		
7	✓				
8		✓			
9					✓
10			✓		
11	✓				

Table 1.2

8. Conclusion

In this paper, we found important points for the study of literature related to different application domains of GIS. This study mainly focuses on tools and techniques used by different GIS applications and the results gained by the researchers. GIS technology has become more user-friendly with time. Continuous development has taken place with time in this field. The GIS technology is used in various fields like science, industry, government, public health, archeology, urban and rural planning, logistics, transportation, environmental sustainability, defense, etc. It is not limited to only one specific field.

References:

1. "Review of GIS-Based Applications for Mining: Planning, Operation, and Environmental Management", Yosoon Choi , Jieun Baek and Sebeom Park
2. "AN OVERVIEW OF GEOGRAPHIC INFORMATION SYSTEM AND ITS ROLE AND APPLICABILITY IN ENVIRONMENTAL MONITORING AND PROCESS MODELING", Miroslav RUSKO, Roman CHOVANEC, Dana ROŠKOVÁ
3. "Geographic Information Extraction using Natural Language Processing in Wikipedia Texts", Edson B. de Lima, Clodoveu Augusto Davis Jr., Proceedings XVIII GEOINFO, December 04th to 06th , 2017, Salvador, BA, Brazil. p 122-127.
4. "The use of GIS technology in modern conditions", D V Andreev 2020 IOP Conf. Series: Earth and Environmental Science 421 (2020) 042001 IOP Publishing doi:10.1088/1755-1315/421/4/042001.
5. Burrough, P. A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. New York: Oxford University PressGIS Development by JRB Dein Techblog.htm
6. http://en.wikipedia.org/wiki/Geographic_information_system
7. https://www.rst2.org/ties/GENTOOLS/comp_gis.html
8. "GIS: Scope and Benefits", Dhirendra Pandey, Anil Kumar Shukla, Abhinav Shukla, International Conference on Advances in Engineering & Technology - AET - 2013 - ISBN-978 - 81 -927082-1-7.
9. **"**Application of Geographic Information System (GIS) in Forest Management", Sonti SH, J Geogr Nat Disast ISSN: 2167-0587 JGND, an open access journal Volume 5 • Issue 3 • 1000145.
10. "GIS-BASED SOLUTION OF SCHEDULING AND ROUTING SCHOOL BUSES- A THEORETICAL APPROACH", Vishakha Shukla, Ashutosh Pandey, Kritika Pandey, Mohammed Saim, International Journal of Advanced Research in Science & Engineering Vol No.4,Special issue(01),November 2015.
11. "Road network analysis of Guwahati city using GIS", Debashis Das, Anil Kr. Ojha, Harlin Kramsapi,,Partha P. Baruah, Mrinal Kr. Dutta, SN Applied Sciences (2019) 1:906 https://doi.org/10.1007/s42452-019-0907-4

12. Received: 17 February 2019 / Accepted: 10 July 2019 / Published online: 23 July 2019 © Springer Nature Switzerland AG 2019
13. "Review of GIS-Based Applications for Mining: Planning, Operation, and Environmental Management", Yosoon Choi , Jieun Baek and Sebeom Park, Appl. Sci. 2020, 10, 2266; doi:10.3390/app10072266 www.mdpi.com/journal/applsci
14. "Application of Gis in Flood Hazard Zonation Studies in Papanasam Taluk, Thanjavur District, Tamilnadu", G. Thilagavathi, S. Tamilenthi., C. Ramu and R. Baskaran, Adv. Appl. Sci. Res., 2011, 2 (3):574-585 ISSN: 0976-8610 CODEN (USA): AASRFC
15. "The current status of GIS in marine fisheries", Geographic Information Systems to support the ecosystem approach to fisheries.
16. "Application of GIS in Public Health in India: A Literature-Based Review, Analysis, and Recommendations", Marilyn O'Hara Ruiz, Arun Kumar Sharma, © 2016 Indian Journal of Public Health | Published by Wolters Kluwer – Medknow, [Downloaded free from http://www.ijph.in on Saturday, December 3, 2022, IP: 103.4.233.24]
17. "Research on the Application of Geographic Information System in Tourism Management", Wei Wei, 2011 International Conference on Environmental Science and Engineering (ICESE 2011) Procedia Environmental Sciences 12 (2012) 1104 – 1109 doi: 10.1016/j.proenv.2012.01.394
18. "APPLICATIONS AND RESEARCH ON GIS FOR THE REAL ESTATE", Chengda Lin, Lingkui Meng, Heping Pan, paper presented at the 22nd Asian Conference on Remote Sensing,5-9 November 2001,Singapore
19. "Application of geographical information systems", J. V. Iyengar Jackson State University, Journal of International Information Management, Volume 7,Issue 2,Article 9
20. "An overview of GeoAI applications in health and healthcare", Maged N. Kamel Boulos , Guochao Peng and Trang VoPham , Kamel Boulos et al. International J Health Geographics (2019) 18:7 https://doi.org/10.1186/s12942-019-0171-2, International Journal of Health Geographics.
21. www.ncbi.nlm.nih.gov
22. en.wikipedia.org
23. www.rst2.org
24. www.grin.com
25. shivkumardas.wordpress.com

International Journal of Computer Research & Technology
ISSN 2454-7719
** A Blind Peer Reviewed Journal **
Volume-8, Issue-1, Jan-June 2023
Paper ID: CS-35

A Review Paper on: Epidemic Prediction

Deepali Yogiraj Kirange
Asst.Professor
KCES's Institute of Management and Research, Jalgaon

Prof.Dr.Varsha M.Pathak
Asst. Professor
KCES's Institute of Management and Research, Jalgaon
Varsha.pathak@imr.ac.in

ABSTRACT - *Epidemic disease outbreak had caused community to raise their great concern over the infectious disease controlling, preventing and handling methods to diminish the disease dissemination percentage and infected area. The epidemic prediction had done using different machine learning, deep learning and AI approaches. In this paper different machine learning algorithm had discussed which were used for epidemic prediction. This study investigates epidemic prediction based on factors affecting for disease spread. In this study the algorithmic analysis has done for epidemic prediction.*

Keywords: Introduction, Epidemic, factors affecting disease spread, algorithm analysis, conclusion

Introduction

Epidemic disease outbreak had caused nowadays community to raise their great concern over the infectious disease controlling, preventing and handling methods to diminish the disease dissemination percentage and infected area [1].

An epidemic is the quick spread of disease to a large number of persons in a given population community, or region within a short period of time. For example plague, yellow fever, Flu, polio, AIDS, H1N1 Swine Flu, Ebola, Zika, Influenza, measles, diphtheria, cholera, diarrhea, malaria , dengue and COVID 19 etc. Epidemic diseases are the contagious diseases that are possible to be spread into the entire nation if the contagion measurement had reached the outbreak level and manage to wipe out the entire population. This contagious diseases had caused a major world health issues and was believed to be the one of the major factors that had caused the life lost globally [1]. The entire worlds also had experienced some of epidemic diseases outbreak such as COVID 19, dengue, hepatitis, chikungunya and many more. However, recently from 2019 to till date, we faced the outbreak of COVID 19 in entire world. The infection of epidemic disease can be spread vigorously by the active mobilization of the pathogen and the rapid production and stimulation of the pathogens. [2]

There is need to apply machine learning & Artificial Intelligence (AI) technique to detect the possibility of epidemic events occurring in specific region, population.

For the detection of epidemic SVM, Naive Bayes, Multinomial Naive Bayes, k-Nearest Neighbor, Logistic Regression, Decision Tree, Deep Neural Network, and Random Forest can be used [3].

2. LITERATURE REVIEW: -

There are several research that have been already done in the field of epidemic prediction. Some of the studies are discussed in this section.

Najihah Ibrahim et al. [1] addressed the backpropagation method was used for the counter measure and prediction analysis of the epidemic disease. The data collection activities done by using the Electronic Health Record (EHR). The predictive analysis based on the backpropagation method can be determine via machine learning process that promotes the artificial intelligent in pattern recognition, statistics and features selection. This computational learning process will be integrated with data mining by measuring the score output as the classifier to the given set of input features through classification technique.

R. Sujath et al. [4] had worked on the linear regression, multilayer perceptron and vector auto regression method for desire on the COVID-19 Kaggle data to anticipate the epidemiological example of the

disorder and pace of COVID-2019 cases in India. Expected the potential patterns of COVID-19 effects in India dependent on data gathered from Kaggle.

Juhyeon Kim et al. [5] proposed that, the newly emerging infectious infections could be detected using the news article data from Medisys. Support Vector Machine (SVM), Semi-supervised Learning (SSL), and Deep Neural Network (DNN) were used for forecast to examine the use of information embedded in the web articles: and to identify the pattern of emerging infectious disease.

Sina F. Ardabili et al. [6] presents a comparative analysis of machine learning and soft computing models to forecast the COVID-19 outbreak. Among a wide range of machine learning models investigated, two models showed promising results (i.e., multi-layered perceptron, MLP, and adaptive network-based fuzzy inference system, ANFIS). This study advises machine learning as an effective tool to model the outbreak.

Rahele Kafieh et al [7] selected deep learning models including multilayer perceptron, random forest, and different versions of long short-term memory (LSTM), using three data sources to train the models, including COVID-19 occurrences, basic information like coded country names, and detailed information like population, and area of different countries.

Chamalka Seneviratne Kalansuriya, et. al [8] presented the review of machine learning approaches to the dengue prediction was done. They consider environmental factors such as climate and vegetation data, dengue case data along with the population of a specific geographic area for the disease outbreak predictions.

Loukas Samaras et. al [9] inspect a highly infectious disease, measles, as there is no extensive literature on forecasting measles using Internet data. The data collected from ECDC official web site and Google Trends. Author compared regression models predicting the development of measles in the five countries. The measles can be estimated and forecast through Google Trends in terms of time, volume and the overall spread. Python scripts used within the framework of an integrated Internet surveillance system for tracking epidemics as the one addressed here.

Samuel Lalmuanawma et al [10] proposed the widely review the role of AI and ML as one significant method in the arena of screening, predicting, forecasting, contact tracing, and drug development for SARS-CoV-2 and its associated epidemic.

Chimmula, V. K. R et al [11] had develop a Canadian based forecasting model using time-series employing Deep learning algorithm for the long-short-term-memory network, the studies found out a key factor intended for predicting the course with an ending point estimation of the current SARS-CoV-2 epidemic in Canada and all over the globe.

Miodrag Zivkovic et. al [12] proposed prediction model that signifies a hybridized approach between machine learning, adaptive neuro-fuzzy inference system and enhanced beetle antennae search swarm intelligence metaheuristics. The greater beetle antennae search is utilized to regulate the parameters of the adaptive neuro-fuzzy inference system and to improve the overall performance of the prediction model. This model test on Covid-19 & datasets of weekly influenza confirmed cases in China and the USA.

Nicholas Thapen et. al [13] proposed DEFENDER, a software system that integrates data from social twitter's live streaming API and news media from RSS and incorporates algorithms for outbreak detection, situational awareness and forecasting. As part of this system they developed a technique for creating a location network for any country or region based purely on Twitter data. They present a disease forecasting approach which leverages counts from multiple symptoms, which was found to improve the forecasting accuracy by 37 percent over a model that used only previous case data.

On the basis of literature review there is Dengue, Influenza, Chikungunya and COVID 19 etc. are the epidemics. These epidemics, have lot of common symptoms. There are many factors they affect the disease spread.

3. FACTORS AFFECTING THE DISEASE SPREAD :-

The disease transmission and spread show interactions with climate, population, environmental factors like water, cleanliness, food and air quality are vital elements in the spread of communicable diseases and in the spread of illnesses prone to cause epidemics. On the basis of literature review we found the variables, research area, description and evaluation for location based Dengue, Influenza, Chikungunya and COVID 19 epidemics. These are mentioned into the following Table 1.

System (name)	Variables used and area researched	Description	Result
A DEFENDER, system that integrates data from social and news media for outbreak detection.[13]	1. Diarrhea 2. Vomiting 3. Gastroenteritis • Area- UK	Using ARIMA creating a location network based on Twitter data and disease forecasting approach which leverages counts from multiple symptoms.	The best performing models were found to be two commonly used forecasting algorithms. ARIMA model and additional regressor.
Prediction of dengue outbreak for the best machine learning model [14]	1. Climate data •Area-Selangor Malaysia	The predictive modelling was performed using several data mining models, namely Decision Trees (CART), Artificial Neural Network (MLP), SVM (LINEAR, POLYNOMIAL, RBF), and Bayes Network (TAN). The models were evaluated in the analysis node.	The SVM, using a linear kernel, best predicted dengue outbreaks without overfitting.
COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach[15]	1. Covid 19 Case Data • Area-China	To predict new COVID-19 cases by retaining hybridized algorithm between machine learning, adaptive neuro-fuzzy inference system (ANFIS) and enhanced beetle antennae search (BAS) swarm intelligence metaheuristics.	The Cauchy exploration strategy BAS (CESBAS) be a robust metaheuristics that significantly outscored all other approaches including the original BAS.
Machine learning-based prediction of COVID-19 diagnosis based on symptoms[16]	1. Age 2. Sex 3. Symptoms 4. Other Information • Area- Israel	A Baseline model that forecasts a positive SARS-CoV-2 infection in a RT-PCR test by asking eight basic questions.	The model can be implemented globally for effective screening and prioritization of testing for the virus in the general population.
Cluster Based epidemic control through smartphone based body area network[17]	1. Vital Signs • Area- USA	CNI & UGP & approximation method is used.	Develop effective inter & intra cluster based epidemic control strategies based on real social contact networks.
Effective Epidemic Control and Source Tracing Through Mobile Social Sensing over WBANs[18]	1. Vital Signs • Area- USA	Genetic Algorithm use for epidemic source tracking & for controlling Dominating set identification algorithm used.	GA has accuracy 80% & shown by graph.
Predicting and controlling infectious disease epidemics using temporal networks[19]	1. Epidemic case study • Area- Japan	For detection of disease spreading the SIR, SEIR, SI models are used.	Efficient immunization protocol used for controlling.

Epidemic Disease Propagation Detection Algorithm using MapReduce for Realistic Social Contact Networks[20]	1. Epidemic case study • Area- Social network	The parallel self-detection schemes which uses MapReduce algorithm & Hadoop framework to find the infected node in social network.	Social contact network is represented as bipartite graph.
Application of Artificial Neural Networks for Dengue Fever Outbreak Predictions in the Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico-2018 [21]	1 Population size 2.Climate Data • Area-Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico(USA)	Successful results were obtained by using ANN with genetic algorithms to predict dengue cases for the research area. The predictive power was above 70%.	ANN is highly versatile and deals well with nonlinear data.
Kernel-Based Machine Learning Models for the Prediction of Dengue and Chikungunya Morbidity in Colombia-2017 [23]	1. Dengue Case Data • Area-Colombia	Kernel Ridge Regression and Gaussian Processes were used to predict the future outbreaks.	Gaussian kernel outperforms for linear regression but kernel ridge regression does not have a simple interpretation though it achieves more accuracy.
Prediction of Dengue Cases in Paraguay Using Artificial Neural Networks-2017 [22]	1. Climate data 2. Dengue case data 3. River level • Area-Paraguay	Use Artificial Neural Networks (ANN) along with regression method comparisons.	Have the ability of accuracy maintenance when some data are missing ANN also produces acceptable results when data are noisy and incomplete.
Using C-support Vector Classification to Forecast Dengue Fever Epidemics [26]	1. Climate Data 2. Dengue Fever case data • Area-Taiwan	Uses classification of object in data set based on C-SVM kernel with the use of RBF and linear kernels, grid search method is used for the selection of hyper parameters.	Good results were obtained from SVM but according to result comparison of some researches more accurate outcomes could be obtained through other methods also.
Classification Rules Using Decision Tree for Dengue Disease [28]	1. Climate Data 2. Dengue Case Data • Area- India	Uses unsupervised clustering as well as supervised model of classification, higher accuracy in prediction is achieved through decision tree classification model.	Overall performance can be considered good possession of the suitability in order to discover rules in data mining.
Analysis of Significant Factors for Dengue Infection Prognosis Using the Random Forest Classifier [30]	1. Dengue Case Data • Area- India	With the use of RF classification tree to determine and visualize factors which are significant to identify the dengue patients in order to improve stability and accuracy.	High accurate results could be obtained from Random Forest compared to other techniques of statistical models.

4. Machine Learning Approaches:-

"Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding those data and information in the form of observations and real-world interactions."

Machine Learning algorithms are classified as –

1) Supervised Machine Learning Algorithms

Machine learning algorithms that make forecasts on a given set of samples. Supervised machine learning algorithm searches for patterns within the value labels allocated to data points. Some popular machine learning algorithms for supervised learning comprise SVM for classification problems, Linear Regression for regression problems, and Random forest for regression and classification problems.

2) Unsupervised Machine Learning Algorithms

There are no labels related with data points. These machine learning algorithms establish the data into a group of clusters to describe its structure and make complex data look simple and organized for analysis.

3) Reinforcement Machine Learning Algorithms

These algorithms select an action, based on each data point and later learn how good the decision was. Over time, the algorithm changes its strategy to learn better and achieve the best reward.

Generally the supervised machine learning algorithms are used for prediction purpose.

1. **Support Vector Machines (SVM):** SVM is a popular binary classifier built upon the concept of decision planes that define decision boundaries. In this approach, original training data is transformed into a higher dimension using a nonlinear mapping. Within this new dimension, a linear optimal separating hyperplane is searched to minimize the distance between hyperplane points and maximize the margin between the classes.

2. **Naive Bayes:** It is a classification algorithm for binary and multiclass classification problems. This algorithm makes a naive assumption that there are no predictors i. e. features are independent of each other, and one feature's impact on predicting class does not depend on the presence of another feature. The probability model can be used to make predictions for new data using Bayes Theorem. When the data is real-valued it is common to assume a Gaussian distribution.

3. **Logistic Regression (LR):** LR is a statistical technique for analysing a dataset for a binary classification problem. It helps in discovering the relationship between a dependent binary variable and at least one independent variable. Each independent variable is multiplied with weights and summed up. This outcome will sum up to a sigmoid function to get the result in the range of 0 and 1. The values below 0.5 are considered as 0, and those above 0.5 are considered as 1. In this manner, optimization techniques aim to find the best regression coefficients and weights. Logistic regression is mathematically constrained to produce probabilities in the range 0 & 1. Also, it can converge on parameter estimates relatively easily.

4. **Decision Tree:** Decision Tree is a flowchart-like tree structure, where each non-leaf node represents a test on an attribute, each branch denotes an outcome of the test, and each terminal node holds a class label. Attributes values of an unlabeled sample, X, are tested against the decision tree to predict its class. A unique path is traced from root (topmost node) to a terminal node based on attributes' values, which holds the predicted class for the unlabeled sample.

5. **Random Forest:** Random Forest is an ensemble learner which improves the accuracy of the model by combining a collection of decision tree classifiers (forest) to generate the aggregated result. It uses classification and regression trees (CART) methodology to grow the trees. At each node, attributes are randomly selected to determine the split and generate individual decision trees. The values of the random vector sampled are responsible for determining each tree. During classification, votes are cast by each tree, and the class with maximum votes is returned.

6. **K-Nearest Neighbour:** k-Nearest-Neighbour is an instance-based statistical analysis method to perform classification. Its implementation requires an integer k, a set of training data and measure closeness metric. The given training set is used as an input vector to form different regions for different classes. When given an unlabelled object, this classifier searches for k training sets in pattern space that are closest to the unlabelled object. These k training sets are the k "nearest neighbours" of the unlabelled object.

7. **Linear Regression:** Linear regression is most well-known algorithms in statistics and machine learning. Predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explain ability.

There are different types of machine learning algorithms are discussed in Table 2—algorithm analysis. Many studies emphasize the accuracy of the prediction of the disease is dependent on the data set and the mechanisms [21].

Table 2 Algorithm Analysis

Paper No.	Name of algorithm	Brief description
[5], [14], [26]	Support Vector Machine (SVM)	Classifier method which defines decision boundaries through the decision plane concept.
[21], [22]	Artificial Neural Networks(ANN)	Computational algorithm which is based on functions and structure of neural network of biology. Works in the manner of human brain [32]
[28]	Decision trees	A decision tree is a tree-like structure [33]
[7]	Random forests	A classifier with a collection of tree-structured classifiers.
[4] , [23]	Liner Regression	Statistical technique for analyzing a dataset for a binary classification problem.
[11]	Deep Learning	Multiple layers of neural networks to perform in processing data and computations on a large amount of data.
[6] , [12] ,[15]	Adaptive network-based fuzzy inference system (ANFIS)	Neuro fuzzy technique where the fusion is made between the neural network and the fuzzy inference system.

2. Conclusion:-

Epidemic global health burden with the factors affecting the epidemic spread is evaluated in this review paper. Different computational approaches are mentioned for the epidemic predictions. In this paper, different researches carried out for epidemic prediction with the use of computational method of machine learning are critically evaluated. Finally, machine learning-approached dengue, COVID 19, Influenza and Chikungunya research models with different classification algorithms with their pros and cons. In this review, the decision tree methods and random forests of machine learning approach are emphasized due to the superior performance. The contributory factors affecting the diseases change from the location to location so with best accurate approach, variable selections, and accurate data sets will provide best location-based models. These models will help the government, public, and the decision makers of health sector to prevent and minimize future epidemic outbreaks. Through that millions of valuable human lives could be saved as well as the money utilized for disease treatment could also be saved.

3. References:-

[1] Najihah Ibrahim, Nur Shazwani Md. Akhir ,et al. ,Predictive analysis effectiveness in determining the epidemic disease infected area, AIP Conference Proceedings 1891, 020064 (2017); doi.org/10.1063/1.5005397.

[2] Andrick, B., et al. Infectious Disease and Climate Change: Detecting Contributing Factors and Predicting Future Outbreaks. in Geoscience and Remote Sensing, 1997. IGARSS '97. Remote Sensing- A Scientific Vision for Sustainable Development. 1997 IEEE International. 1997

[3] Aakansha Gupta, Rahul Katarya,Social Media based Surveillance Systems for Healthcare using Machine Learning: A Systematic Review(2020) doi:10.1016/j.jbi.2020.103500

[4] R. Sujath , Jyotir Moy Chatterjee ,et al., A machine learning forecasting model for COVID-19 pandemic in India ,Springer-Verlag GmbH Germany, part of Springer Nature (2020), doi.org/10.1007/s00477-020-01827-8

[5] Juhyeon Kim, InsungAhn, Infectious disease outbreak prediction using media articles with machine learning models, Scientifc Reports (2021), doi.org/10.1038/s41598-021-83926-2

[6] Sina F. Ardabili , Amir Mosavi, et al.,COVID-19 Outbreak Prediction with Machine Learning , (2020) medRxiv preprint doi.org/10.1101/2020.04.17.20070094;

[7] Rahele Kafieh,Roya Arian, COVID-19 in Iran: Forecasting Pandemic Using Deep Learning, Hindawi , Computational and Mathematical Methods in Medicine Volume 2021, Article ID 6927985, 16 pages ,doi.org/10.1155/2021/6927985

[8] Chamalka Seneviratne Kalansuriya, et. al ,Machine Learning-Based Approaches for Location Based Dengue Prediction: Review, Springer, Fourth International Congress on Information and Communication Technology pp 343-352.

[9] Loukas Samaras et. al ,Predicting epidemics using search engine data: a comparative study on measles in the largest countries of Europe, Samaras et al. BMC Public Health (2021) , doi.org/10.1186/s12889-020-10106-8

[10] Samuel Lalmuanawma , Jamal Hussain , Lalrinfela Chhakchhuak, Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review, Elsevier, https://www.sciencedirect.com/science/article/pii/S0960077920304562

[11] Chimmula, V. K. R., & Zhang, L. (2020). Time Series Forecasting of COVID-19 transmission in Canada Using LSTM Networks. Chaos, Solitons & Fractals, 109864. doi:10.1016/j.chaos.2020.109864

[12] Miodrag Zivkovic , Nebojsa Bacanin ,et. al, COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach,Elsevier,2021

[13] Nicholas Thapen, Donal Simmie , Chris Hankin , Joseph Gillard, DEFENDER: Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response, Christopher M. Danforth, University of Vermont, UNITED STATES, 2016

[14] NurulAzam Mohd Salim , Yap BeeWah , Caitlynn Reeves , Madison Smith et. al,Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques, Scientifc Reports(2021)

[15] Miodrag Zivkovic , Nebojsa Bacanin ,et. al, COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach,Elsevier,2021

[16] Yazeed Zoabi , Shira Deri-Rozov and Noam Shomron ,Machine learning-based prediction of COVID-19 diagnosis based on symptoms, Digital Medicine(2021)

[17] Zhaoyang Zhang , Honggang Wang et. al,Cluster-based Epidemic Control Through Smartphone-based Body Area Networks, IEEE Transactions on Parallel and Distributed Systems,2015

[18] Zhaoyang Zhang, Honggang Wang, et. al,Effective Epidemic Control and Source Tracing Through Mobile Social Sensing over WBANs , 2013 Proceedings IEEE INFOCOM

[19] Naoki Masuda and Petter Holme, Predicting and controlling infectious disease epidemics using temporal networks, F1000Prime Reports 2013

[20] Rakesh Ranjan, Rajiv Misra, Epidemic Disease Propagation Detection Algorithm using MapReduce for Realistic Social Contact Networks, IEEE , 2014 , 978-1-4799-5958-7/14

[21] A. Laureano-Rosario, A. Duncan, P. Mendez-Lazaro, J. Garcia-Rejon, S. Gomez-Carro, J.Farfan-Ale, D. Savic, F. Muller-Karger, Application of artificial neural networks for dengue fever outbreak predictions in the northwest coast of Yucatan, Mexico and San Juan, Puerto Rico. Trop. Med. Infect. Dis. 3(1), 5 (2018)

[22] V. Ughelli, Y. Lisnichuk, J. Paciello, J. Pane, Prediction of dengue cases in paraguay using artificial neural networks, in 3rd International Conference on Health Informatics Medical Systems (2017), pp. 130–136

[23] W. Caicedo-Torres, D. Montes-Grajales,W. Miranda-Castro, M. Fennix-Agudelo,N.Agudelo-Herrera, Kernel-based machine learning models for the prediction of dengue and chikungunya morbidity in Colombia. Commun. Comput. Inf. Sci. 735, 472–484 (2017)

[24] N. Méndez, M. Oviedo-Pastrana, S. Mattar, I. Caicedo-Castro, G. Arrieta, Zika virus disease, microcephaly and Guillain-Barré syndrome in Colombia: Epidemiological situation during 21 months of the Zika virus outbreak, 2015–2017 (2015)

[25] B. Jongmuenwai, S. Lowanichchai, S. Jabjone, Prediction Model of Dengue Hemorrhagic Fever Outbreak using Artificial Neural Networks in Northeast of Thailand (2018)

[26] D. Rahmawati, Y.P. Huang, Using C-support vector classification to forecast dengue fever epidemics in Taiwan, in 2016 IEEE International Conference on System Science and Engineering (ICSSE, 2016)

[27] G. Zhu, J. Hunter, Y. Jiang, Improved Prediction of Dengue Outbreak Using the Delay Permutation Entropy, in Proceedings on 2016 IEEE International Conference Internet Things (iThings);

IEEE Green Computing and Communications (GreenCom); IEEE Cyber, Physical and Social

Computing CPSCom; IEEE Smart Data Smart Data (2016), pp. 828–832

[28] N.K.K. Rao, G.P.S. Varma, D. Rao, P. Cse, Classification rules using decision tree for dengue disease. Int. J. Res. Comput. Commun. Technol. 3(3), 2278–5841 (2014)

[29] R. Babu, Decision tree model for dengue data analysis, Int. J. Res. Sci. Comput. Eng. 3(1) (2017)

[30] A.S. Fathima, Analysis of significant factors for dengue infection prognosis using the random forest classifier. Int. J. Adv. Comput. Sci. Appl. 6(2), 240–245 (2015)

[31] T.M. Carvajal, K.M. Viacrusis, L.F.T. Hernandez, H.T. Ho, D.M. Amalin, K. Watanabe, Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. BMC Infect. Dis. 18(1), 183 (2018)

[32] R. Gholami, N. Fakhari, Support vector machine: principles, parameters, and applications, in

Handbook of Neural Computation (Elsevier, 2017), pp. 515–535

[33] C. Petri, Decision Trees (2010)

[34] L. Tanner, M. Schreiber, J.G.H. Low, A. Ong, T. Tolfvenstam, Y.L. Lai, L.C. Ng, Y.S. Leo, L. Thi Puong, S.G. Vasudevan, C.P. Simmons, M.L. Hibberd, E.E. Ooi, Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS Negl.Trop. Dis. 2(3), e196 (2008)

[35] L. Breiman, Machine Learning (Kluwer Academic Publishers, Dordrecht, (2001)

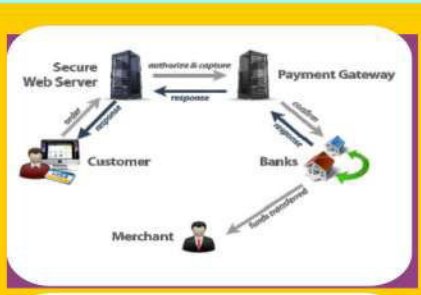
[36] A.A. Annakutty, A.C. Aponso, Review of brain imaging techniques, feature extraction and classification algorithms to identify alzheimer's disease. Int. J. Pharma Med. Biol. Sci. 5(3), 178–183 (2016)

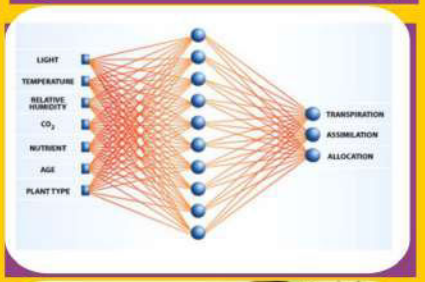
ISSN NO
2454-7719

International Journal of Computer Research & Technology (IJCRT)

Blind Peer Reviewed Half Yearly Research Journal

Vol -8 Issue -1
January-June,2023





Special Issue

Proceeding of the 8th National Conference on "Emerging Trends in Computer and Information Technology" held on 27th March 2023.

Editor in Chief

Dr.B.H.Barhate

Department of Computer Science and Information Technology

Bhusawal Arts Science & P.O.Nahata Commerce College, Bhusawal - 425 201 Maharashtra