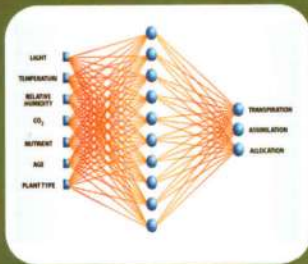# International Journal of Computer Research & Technology ( IJCRT )

## A Peer Reviewed Half Yearly Research Journal

**Vol -7 Issue -1**
**January-June,2021**

**Editor in Chief**
**Dr.B.H.Barhate**

# Special Issue on
# 7th National Conference on "Emerging Trends in Computer and Information Technology"

**Held on 29th May, 2021**

**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 areas with the servants from all over India, Bhusawal serves as a slice of the nation; and honorably has unity in diversity. It is 25 kms away from the district, Jalgaon, famous as a city of gold; and few kms away fromYawal 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 the city 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 a home 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, 50 kms 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 Madhukarrao Chaudhari, ex-speaker of Maharashtra Legislative Assembly, a group of 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: *Vidyadanammahatpunyam.* 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 *vidyadanam*is*mahatpunyam.* The dedicated faculty, the sublime view of the management soonstarted to produce good academicians. Inspired by the results the trust purchased a barren land of 7 acres out of the city which issoon 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 3$^{rd}$ 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. 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 a platinum 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.

# ETCIT-2021
# Organizing Committee

## *Patrons*

| | |
|---|---|
| **Dr. Mohan Phalak** <br> President, Tapti Education Society, Bhusawal | **Mr. Mahesh Phalak** <br> Chairman, Tapti Education Society, Bhusawal |
| **Mr.Vishnu Chaudhari** <br> Secretary, Tapti Education Society, Bhusawal | **Mr.Sanjaykumar Nahata** <br> Treasurer, Tapti Education Society, Bhusawal |

## *Co-Patrons*

| | | |
|---|---|---|
| *Prof. Dr. M. V. Waykole* <br> Principal, BASPONC College, Bhusawal | | |
| *Dr. S. V. Patil* <br> Vice Principal, <br> BASPONC College, Bhusawal | *Dr. A. D. Goswami* <br> Vice-Principal, <br> BASPONC College, Bhusawal | *Dr. N. E. Bhangale* <br> Vice-Principal, <br> BASPONC College, Bhusawal |

## *Convener*

### *Prof. Dr. B. H. Barhate*
Vice-Principal, IQAC Coordinator, HOD of Computer Science BASPONC College, Bhusawal

| *Organizing Secretary* | *Co-ordinator* |
|---|---|
| **Dr. Gouri M. Patil** <br> Assistant Professor, BASPONC College, Bhusawal | **Dr. G. R. Wani** <br> Associate Professor, BASPONC College, Bhusawal. |

## *Inaguaration Speech*

**Prof. Dr. B. V. Pawar** is presently Pro-Vice-Chancellor (Acting)**,** KBC North Maharashtra University, Jalgaon (MS), 425001. He is former Officiating Registrar at KBC North Maharashtra University, Jalgaon & Director of School of Computer Sciences, KBC North Maharashtra University, Jalgaon (MS) during 2013-2019.

## *Resource Person*

1. **Prof. Dr. Parag Anil Tamhankar** is the faculty from Abasaheb Garware College, Pune. He has completed M.Sc. (Computer Science), Ph.D. (Computer Science), NET-2012, SET-2002. His experience is more than 21 years in college teaching, 15 years as Corporate Trainer and visiting faculty at 13+ colleges. He published 12+ research papers in journals, conferences and events. His area of specialization is Digital Image Processing, Machine Learning, and Artificial Intelligence.

2. **Mr. AtulKahate is** an Author and Visiting Faculty of Computer Science. His overall Corporate and Teaching experience is over 25 years. He is Author of 70 books (30 in English and 40 in Marathi). His book titled "Cryptography and Network Security" has crossed the mark of 1.50 lakh copies. He has completed MBA after BSc (Statistics) and earlier worked in Information Technology (IT) industry - Handled various responsibilities at Syntel, American Express, Deutsche Bank, L&T Infotech, and i-flex Solutions

### Organizing Committee Members

| Prof. Harshal V.Patil | Prof. Dr. Swati P. Phalak | Prof. Poonam M.Mahajan |
|---|---|---|
| All Teaching and Non-Teaching Staff of DOCS | | |

### *President's Message*

       I am very happy to know that Department of Computer Science and I.T., Bhusawal Arts, Science and P. O. Nahata Commerce College, Bhusawal is organizing a one day online 7th National Conference "Emerging Trends in Computer And Information Technology" (ETCIT-2021) on 29th May, 2021. I assure that this online conference will provide a platform to bring together researchers from various research and industrial organizations and educational institutions under a common environment and discuss the emerging trends in Computer and Information Technology in this COVID pandemic. My best wishes to all the participants and wish the conference good luck and grand success.

       Stay Home ! Stay Safe

## *From principal's Desk........*

The Department of Computer Science and Information Technology is happy to announce that, following the success of the online 7th National Conference ETCIT-2020, the Department of Computer Science and Information Technology will host a one-day online 7th National Conference on "Emerging Trends in Computer and Information Technology" (ETCIT-2021) on May 29, 2021.

COVID pandemic is and not will be able to stop the researches by researchers due to such type of online conferences organized by Department of Computer Science and Information Technology.

All the staff of Department of Computer Science and Information Technology is always make such innovative and challengeable events successful beyond our expectation and make me proudly. Passion of organizing committee of the national conference ETCIT-2021 is very admiring.

I ensure enormous success for this one day online 7th National Conference ETCIT-2021 from my bottom of heart.

# *Convener's Message*

Acquaintance of advance and enhanced technology is useful in the world of modernization. Nowadays researches with emerging trend in Computer and Information Technology together make on hand the source which is useful for assembly and moving towards the auspicious intend. The event 7th one day online National Conference entitled "Emerging Trends in Computer Science and Information Technology - (ETCIT-2021)" is organized by Computer Science and IT department. This conference is the platform for researchers as well as students.

We feel grateful to all who submit their research work for this conference. The selected papers will be published in the International Journal of Computer Research and Technology (IJCRT), a peer reviewed, half yearly research journal, Volume-7, Issue-1, Jan-Jun 2021, ISSN: 2454-7719. We are pleased to our patrons, principal, resource persons, vice principal for their blessing and moral support. I assure great success of the National Conference.

Stay Home, Stay Safe, Save Live!

**Dr. B. H. Barhate**

**Vice - Principal, HOD of Comp. Sci. and Info. Tech.,**

**BASPONC College, Bhusawal**

**Mobile: +91-9890966830**

# I N D E X

Paper ID: CS-01

# Artificial Intelligence to Enhance Education System

## Suresh J. Raut. (MCA, MCM, DCM)

*STNC – School Training Nodal Co-ordinator, PGT-CS*
*Reliance Foundation School, Nagothane. Dist:-Raigad, Maharashtra, India.*
*Email- sjraut111@gmail.com*

**ABSTRACT -** *Artificial intelligence has the power to enhance the education system. It has capabilities to address some of the most critical challenges in learning and teaching activities. Not only it will impact the overall development and performance of students, but also help to reduce the stress and pressure on educators. AI solutions bring forth student performance analytics, personalized learning, and student engagement tools to enrich the traditional classroom experience.*

*Keywords - AI, AR, VR, CAGR, AI-Powered, Real-Time, GPS, CCTV, RFID.*

## I. INTRODUCTION

### What is Artificial Intelligence in Education?

Artificial Intelligence in education is more about empowering education system with advanced tools and technologies that would help to cater the growing needs of students and the rising demand for personalizededucation.Butit is also important to understand that artificial intelligence in education doesn't replacing teachers and tutors with a machine or robot in traditional classrooms.
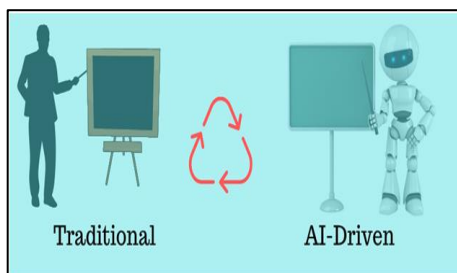


The virtual facilitators and learning environments applications are projected to have the largest market size by 2023.

The global AI in education market is projected to reach USD 3.68 billion by 2023, at a CAGR(Compound Annual Growth Rate)of 47% during the period 2018 till 2023.

## II. AI SOLUTIONS

AI solutions can be used to enhance the education system by using:

**a) E-Content:**IT is an important part of the digitized curriculum. Many schools, colleges and universities are taking efforts to make the online curriculum available using different digital platforms to make it
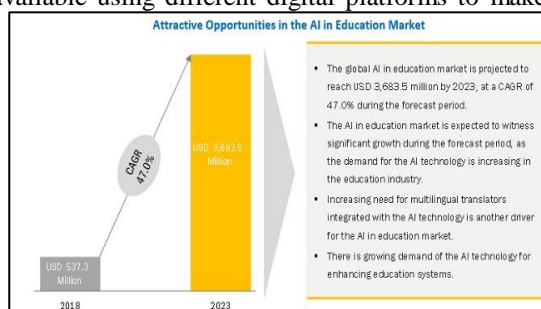


more interactive and easily accessible. With AI-based tools, learning materials such as e-books, audio and video presentations can be easily categorized and grouped according to the educational needs and preferences. This way the learning process and the curriculum could be more student-centric.

**b) Personalized learning using AI based tools:** The traditional teaching and learning process in education system targets the 80% of students and ignores the 20% students, AI-based systems servethestudents of all capabilities and discover their full potential.

AI Tools are useful for Schools and universities in the classroom for personalized recommendations on the sessions, their performance, daily assignments and quizzes, and class tests or examinations. Also, Students can have access to AI-powered e-content systems where they can have the advantage of personalized study plans, accessible from any user device or location.
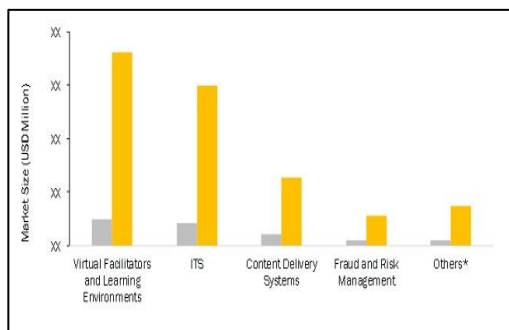


c) **Voice Tools:**Voice Tools like Apple's Siri, Amazon's Alexa, and Google Home can be helpful in the teaching and learning activities in the class room. These interactive voice assistants can assist with different learning material in the absence of a teacher, and can be made learning possible anywhere and anytime. Apart from supporting learning, integration of such voice assistant technology within the school/colleges and universities campus system can be the effective way to engage the students.

d) **Global Learning Platforms:** Technology is bridge to joins different learning platforms. It connects people of different geographical locations to learn and share the knowledge. This could be useful for the students to communicate and share ideas with their peers from around the world. AI-based tools are helping them to organize class-based curriculum and customize the experience to match the student's needs. It can provide access to education materials. Language translations and Automated Subtitle creation enable teachers and students to communicate online without any language barriers. The future of global education looks bright as these AI tools get further fine-tuned to support more languages and regional dialects.

e) **Institution and ResourceManagement:**AI technologies can optimize and manage processes for educational institutions. For Example, AI-based tools can help in the proper designing and planning of classrooms according to the number of students. With data-driven insights, school management can be timely distributed resources.

f) **AI based RFID:**It stands for Radio Frequency Identification. It can be introduced with the student's ID cards or attached to their uniforms, every time when student passes through a specific placed checkpoint; their presence is noted, also it notifying all the stakeholders – parents, management and the School ERP software.

It is useful for teachers because it removes the headache of marking individual attendance in every class. It notifies absent students automatically received. It has ability to notify parents about absentees. It can communicate with students who aren't present in class about assignments and homework

It is useful for schools because automatic report generation helps in analysis of final attendance. It saves room that is required to store large files with student attendance and records. Get notified if students are in the restricted area where they are not given access to and identify the students.

It is useful for students because it ensures accurate attendance is recorded as system is updated in real-time. Grievances about attendance can be instantly raised. It reduces pressure on students for roll call; giving students more time to learn. It takes review attendance analysis and reports in different formats

It is useful for parents because it updates on entry and exit of the student to and from school. Daily, monthly, term and yearly attendance percentage automatically calculated. Alerts on absenteeism through SMS and in the school/institute/collegesoftwaresystem notification, dashboards and reports.

g) **AI in Early Learning:** it is an adaptive learning platform with AI games and toys for small children. AI based games are useful to educate the small children of Nursery, LKG, UKG classes.

h) **AR and VR:** Students can get a chance to learn in augmented and virtual reality. Machine learning will continue to improve the VR experience in many ways including gesture and voice recognition, image processing and better collaborative experiences.

**i)AI for Assessment:** AI tool can improve assessment in many ways including adaptive testing using MCQs and short answer questions, faster grading, tracking steps in problem-solving and monitoring student progress including hard to measure skills. It will also be key to interoperability and combining multiple sources of formative assessment.

**j)AI BasedAnalyticsPlatform:**It can be used to give early warnings and notifications. Analytics platforms may be used in higher education to keep track on students' performance especially for K-12 and University students.

**k) Transportation:** Bus scheduling just got easier with AI-powered mobile apps. Vehicle with GPS tracking system and CCTV is more secure than normal vehicle.

## III. CONCLUSION

Education system could be managed more efficiently by AI powered applications. AI tools can be used to monitor and analyzethe Student Progress in Real-Time. It Saves Time and Improves Efficiency. It gives more personalized learning experience. It provides Convenient and Improved Student-Teacher Interactions. AI can be used in Simplifying Administrative Tasks.AI technology application has good potential in Education System in future, hence AI apps programmer should focus on to develop more AI based apps to enhance the traditional education system.

## REFERENCES

1. Website:https://clofus.com/features#Robot
2. Web site:https://www.gettingsmart.com/2018/08/32-ways-ai-is-improving-education/
3. Statistics:https://tinyurl.com/AI-apps-use-Statics
4. Reliance Foundation School, Nagothane,Raigad,Maharashtra

# An Overview of Clustering Techniques

**Ankita Sanjay Deo**

*Dept. of Computer Mngmt.,
Systel Institute of Mngmt. and
Research, Dhule.
Maharashtra, India*
Email-0607deoankita@gmail.com

**Rupali P. Patil**

*Dept of Computer Sci.,
S.S.V.P.S's L.K. Dr.P.RGhogrey
Science College, Dhule.
Maharashtra, India*
Email-rupalipatilh@gmail.com

**Hansraj M. Patil**

*Dept. of Computer Mngmt.,
Systel Institute of Mngmt. and
Research, Dhule.
Maharashtra, India*
Email-hmpatil.systel@gmail.com

*ABSTRACT -Data mining is a process of extracting the useful information in huge data repositories and mold it into an understandable structure for future use. Clustering is an important and widely used technique in data analysis and data mining application. Clustering is a process of grouping data objects into clusters based on similarity measures. Data mining can be handled by using supervised and unsupervised learning. The clustering is an unsupervised learning task where one seeks to recognize a finite set of categories termed clusters to describe the data. Clustering algorithms can be classified into partitioning algorithm, hierarchical algorithm, density-based algorithm and grid based density algorithms. Partitioning clustering algorithm divides the data points into k partition i.e. cluster. Hierarchical clustering algorithm divide the similar data set by build a hierarchy of clusters. Density based algorithms identify the cluster in a data space is a contiguous region of high point density, divided from other clusters by contiguous region of low point density. Grid based density algorithm generally have a fast processing time, which uses the multi-resolution grid data structure and uses dense grid to form clusters. In this paper an overview of different clustering techniques is done and also discuss the advantages and disadvantages of each technique along with some clustering applications.*

*Keywords– Data Mining, clustering, supervised, unsupervised, semi supervised learning, classification of clustering*

## I. INTRODUCTION

Data mining is the process of discovering or mining knowledge from a huge amount of data. Data mining is also known as KDD (Knowledge discovering from Data). It attempts to extract hidden patterns and trends from huge databases. Data mining includes the utilization of refined data analysis tools to discover previously unknown, valid patterns and relationships in huge amount of data. These tools can constitute statistical models, mathematical algorithms, machine learning techniques. Thus, data mining can be used for abandon analysis and prediction, depending on different technologies from the intersection of machine learning, statistics etc.Professionals in data mining have dedicate their careers to better understanding of how to process and make conclusions from the huge amount of data. For this reason various data mining techniques have been developed and used, including association, classification, clustering, prediction, sequential patterns and regression. In association a pattern is find based on a relationship between items in the same transaction. Classification helps to classify data in different classes. Clustering is the process of building a group of abstract objects into classes of similar objects. Prediction discovers relationship between dependent and independent variables. Sequential patternsrecognize similar patterns, regular events or trends in transaction data over a business period. Regression analysis is used to identify and analyze the relationship between variables. In this paper, we have done overview of various clustering techniques used in data mining.

In data mining threetypes of learning sets are used, they are supervised learning, unsupervised learning and semi supervised learning

*Supervised learning* – A supervised learning algorithm learns from labeled training data and helps us to expect outcomes for unexpected data. It can be compared to learning which take place in the existence of a supervisor. It recognize to collect data or produce a data output from the preceding experience. It helps to solve different types of real world computation problems. It is simpler method. Its result is highly accurate. Classification and regression are the type of supervised learning technique.It also includes support vector machine, neural network, linear and logistics regression and classification trees.

*Unsupervised learning* – Unsupervised learning does not need to supervise the model, need to allow the model to work on its own to discover information. It cover with unlabeled data. It finds all kinds of unknown patterns in data. It helps to find characteristics which can be useful for categorization. It is complex method compared to supervised learning. Clustering, association are the types of unsupervised learning technique. It support k-means algorithm, hierarchical clustering etc.

*Semi supervised learning* - Semi supervised learning combines a small amount of labelled data with a large amount of unlabeled data during training. Semi supervised learning can be used in document classification speech analysis etc.

## II.   CLUSTERING

Clustering is a type of unsupervised learning method of machine learning. In the unsupervised learning method, the inferences are worn from the data sets which do not contain labelled output variable. It is an exploratory data analysis technique that allows us to examine the multivariate data sets. Clustering is the process of partitioning a set of data into relevant similar subclasses called clusters. Clustering is the process of grouping the set of objects in such a way that the object ofsamecharacteristics are grouped together i.e when doing clustering or cluster analysis.We first partition the data into groups based on the similarity between them. Let us take an example, Library. In a library there will be large number of books in various topics at different shelf. So it is very difficult to pick a particular book. So here comes clustering techniques. By using clustering technique we can remain the books that have similarities in one shelf or cluster and label it

meaningfully. Then it will be easy to grab the book. This is the simple example for clustering.
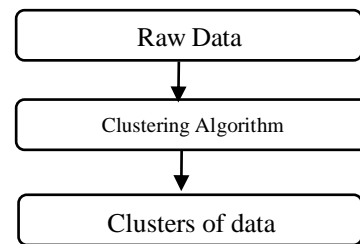


*Figure 1: Stages of clustering*

Raw data apply clustering algorithm and the output you getas clusters of data.A cluster is a group of data objects that affiliated to one another withinthe same group and unrelated to the objects in the other groups. Cluster analysis is to identify similarities between data according to the features found in the data and grouping similar data objects into clusters.

There are two types of clustering, hard clustering and soft clustering. In hard clustering each data point is either belongs to a cluster completely or not. In soft clustering, instead of putting each data point into a different cluster, a probability or likelihood of that data point to be in those clusters is assigned

## III.   CLASSIFICATION OF CLUSTERING ALGORITHM

Clustering is an important task of data mining.Various algorithms are used for clustering. Clustering algorithms can be categorized into partitioning, hierarchical, density and grid based algorithm.

### A.   *Partitioning Algorithm*

Partitioning clustering method divides the information into multiple groups depending on the features and similarity of the data. The algorithm requires the analyst to specify the number of cluster to be generated. In partitioning method when database(D) that contains multiple(N) objects then the partitioning method constructs user-specified(k) partitions of the data in which each partition represents a cluster and $k \leq N$. It will classify the data into k groups, each group contain at least one object and each object must belong to exactly one group. For a given number of partitions(k), the partitioning method will create an initial partitioning. Then it uses the iterative relocation technique to improve the partitioning by moving objects from one group to other.

There are many partitioning method are used; they are k-means, k-medoidsetc.

### 1) K-means Algorithm -

K-means is the most popular clustering method. It is centroid based technique. Initially, selecting k objects from the data set are chosen in which each of the objects represents a cluster mean or centroids. Each of the remaining objects is assigned to its nearest centroids, where generally nearest is determined using the Euclidean distance between the object and the cluster mean. The new mean of each of clusteris then computed with the added data objects. Now that the centers have been recomputed, each observation is looked again to see if it force be near to a different cluster. All the objects are reassigned again utilize the updated cluster means. These steps are iteratively repeated until the convergence is achieved. Although k-means has the advantage of living easy to implement and fast algorithm,it can be effective to deal with very large data sets. There are some drawbacks, it presume previous knowledge of the data and need the analyst to choose the relevant number of cluster in advance. It's sensitive togive the initial cluster.

### 2) K- medoids Algorithms

It is related to K-means clustering for partitioning a data set into k groups or clusters. In K-medoids, each cluster is represented by one of the data point in the cluster that is cluster medoids.

The medoids declares an object within a cluster for which average dissimilarity between it and all the other the members of the cluster is minimum. It communicate to the most centrally located point in the cluster. K-medoids algorithm is less sensitive to outliers and noise, compared to k-means, because it apply medoids as cluster centers alternately of means.

The general k-medoids clustering method is the PAM algorithm(Partitioning around Medoids). Initially selecting randomly an object as medoids for each of the k clusters. Then, each of the unselected objects is grouped with the medoids to which it is the similar. It then reiteratively return one of the medoids by one of the non-medoids objects. There are some drawbacks, It is improper for clustering non-spherical groups of object. It get distinct results for different runs on the same dataset because the initially k medoids are chosen randomly.

*Advantages of Partitioning Clustering Algorithm –*

1. It is simple, fast and robust.
2. It is easy to understand and implement
3. Interactive for datasets with compact spherical clusters that are well-separated.
4. K-means may be computationally fast with a large number of variables.
5. K-means is based on the exclusive clustering then its produce tighter clusters.



**Before k-means**          **After k-means**
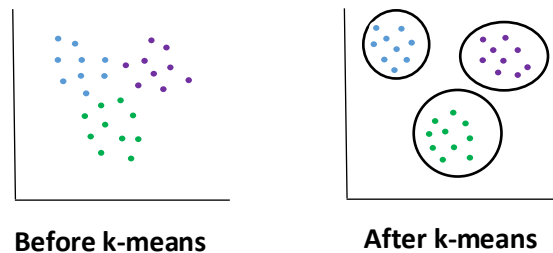
*Figure 2: k-means Algorithm*

*Disadvantages of Partitioning Clustering Algorithm –*

1. High sensitivity to initialization phase, noise and outliers.
2. Poor cluster descriptive.
3. Applicable only when mean is defined i.e. fails for categorical data.
4. Not suitable to find clusters with arbitrary shapes.
5. Result and total run time depends upon initial partition.

### B. Hierarchical Algorithm

Hierarchical clustering is a method of cluster analysis which separates the similar dataset by constructing a hierarchy of clusters. This algorithm starts by treating each data points as a separate clusters. First, point out the two nearest clusters are merged into the similar cluster. This repetitive process continues until all the clusters are merged together.

In hierarchical clustering, the result produced is a nested clusters called dendogram. A dendogram is the graphical representation of the hierarchy, a single all inclusive clusters are at the top and singleton clusters of separatepoints at the bottom.The hierarchy can be formed in bottom-up (agglomerative) or top-down (divisive) way.

## 1) Agglomerative Hierarchical Clustering –

It is also known as AGNES(Agglomerative Nesting). It's a bottom-up approach. Initially each object as a singleton cluster. Take the two nearest clusters and make them one cluster. This procedure is repeated until left with only one cluster. AGNES is
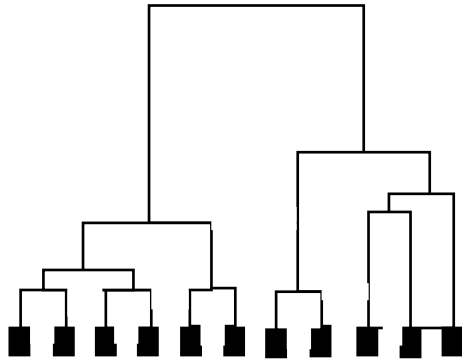


*Figure 3: Dendogram*

good at recognize small clusters.

## 2) Divisive Hierarchical clustering -

It is also known as DIANA(Divisive Analysis Clustering). It's a top-down approach. It is exactly opposite to Agglomerative clustering. Initially an including all objects in a single cluster. Then divide the cluster to two least similar clusters. The process is repeated until all object are in their own cluster. DIANA is good at identifying large cluster.

### Advantages of Hierarchical Clustering Algorithm –

1. Easy to understand and easy to implement.
2. We do not require to specify the number of clusters needed for the algorithm.
3. Dendogramis great for visualization.
4. Relevancy to any attributes type.

### Disadvantages of Hierarchical Clustering Algorithm –

1. The algorithm does not reverse any preceding steps.
2. Vagueness of termination criterion.
3. If we have a huge dataset, it can become difficult to discover the accurate number of clusters by the dendogram.
4. Deficiency of interpretability regarding the cluster descriptors.

### C. Density-Based Clustering Algorithm

Density based clustering is an example of unsupervised learning method that identify distinctive clusters in the data. It is well organized to discover the clusters of arbitrary shapes, which is containing noise and outliers. It is based on density.

**DBSCAN (Density Based Spatial Clustering of Application with Noise)** is most widely used density based algorithm.
The DBSCAN algorithm need two parameters- eps($\epsilon$)and MinPts.

1. eps($\epsilon$) – A distance will be used to discover the points in the neighborhood of any point.
2. MinPts – The minimum number of data points clustered together for a region to be considered dense.

These parameters can be recognized if we examine two concepts called Density Reachability and Density Connectivity.

*Density Reachability* – A point "a" is said to be density reachable from a point "b" if point "a" is within $\epsilon$ distance from point "b" and "b" has sufficient number of points in its nearest which are within distance $\epsilon$.

*Density Connectivity* - It is a chaining process. A point "a" and "b" are said to be density connected if there exist a point "c" which has sufficient number of points in its nearest the both points "a" and "b" are within the $\epsilon$ distance. So, if "b" is nearest to "c", "c" is nearest of "d" , "d" is nearest of "e" which in turn is nearest of "a" implies that "b" is nearest of "a" .

In DBSCAN algorithm, we have three types of data points:

*Core Point* – It is taken that particular point which in epsneighborhood, has more than MinPts.

*Border Point*–It is taken particular point which inepsneighborhood, has less than MinPts. But it is in the nearest of a core point.

*Noise/Outlier* - A point which is neither a core point nor a border point.sws

### DBSCAN Algorithm-

- Initially, discovers all the nearest data points within eps($\epsilon$) and pick out the core point s or visited with more than MinPts nearest.
- For every core point if it is not ahead assigned to a cluster, make a new cluster.
- Then discover repetitive all its density connected data point and allocate them to the same cluster as the core point.
- Repeat through the remaining unvisited data points in the dataset.
- Those data points that do not apply to any cluster are consideredas an outlier.

*Advantages of Density Based Clustering algorithm –*

1. DBSCAN algorithm is able to discover arbitrary size cluster.
2. Able to handle noise and outlier.
3. Discover more sub-cluster of data huge cluster number is specified.
4. The order of the point in the database is insensitive.

*Disadvantages of Density Based Clustering algorithm –*

1. DBCAN algorithm fails in case of different density clusters.
2. Cannot performwell in case of high dimensional data.
3. Do not handle non-globular data of various size and densities.

### D. Grid Based Clustering Algorithm

Grid density based clustering is united with the value space that surrounds the data points and not with the data points. This clustering approach uses a multi-resolution grid data structure and utilize dense grid to form clusters. The object space is quantized into finite number of cells which form the grid structure and on which all of the operations for clustering are performed. Grid based clustering maps the unlimited number of data records in data streams to finite numbers of grid. The grid based clustering methods have the fastest processing time that typically depends on the size of the grid instead of the data objects. These methods utilize a single uniform grid mesh to divide the entire problem domain into cells and the data objects discover within a cell are represented by the cell using a set of statistical attributes from the objects. Clustering is, then, performed on the grid cells, instead of the data set itself, because the size of the grid is usually much less than the number of data objects, the processing speed can be significantly enhance. The grid based clustering algorithms are CLIQUE(Clustering In QUEst), STING(Statistical INformation Grid), Wave

Cluster. These all methods use a uniform grid mesh to cover the entire problem. But these methods are organized only for low dimensions. Around the large number of cells most are empty and several may be failed with one point. For the problems with high dimensional data distributions, the resolution of the grid mesh must be fine to recover a good clustering quality. A finer mesh can result in the mesh size close to or even exceed the size of the data objects, which can significant increase the calculation load for clustering. The algorithm OPTICS is proposed for the purposed of high dimensional data.

The Grid based clustering algorithms typically involve the following five steps:
• Creating the grid structure or partition the data space into a finite number of cells,
• Calculating the cell density for each cell,
• Sorting of the cells according to their densities,
• Identifying cluster centers and traversal of neighbor cells.
• The Grid density algorithm perform well over the time complexity as well as on the high dimensional data.

*Advantages of Grid Based Clustering algorithm –*

1. The major advantage of this method is fast processing time.
2. It is dependent only on the number of cells in each dimension in quantized space.
3. It offers good performance to deal with arbitrary shaped clusters.
4. Its computational time is independent to the number of data points.

*Disadvantages of Grid Based Clustering algorithm –*

1. The accuracy of the clustering result may be reduce at the expense of the simplicity of the method.
2. It needs a large number of parameter.
3. Not suitable for non-convex data and relatively sensitive to the number of clusters.

### IV. APPLICATION OF CLUSTERING

Clustering has a large number of applications spread across various domains. Some of the most popular applications of clustering are:

1. Image Processing
2. Document Classification
3. Pattern Recognition
4. Spatial Data Analysis
5. Market research
6. Fraud Detection Application

### V. CONCLUSION

The main purpose of this paper is to study commonly used clustering algorithms,there advantages and disadvantages. In our study we have covered various clustering algorithms like partitioning clustering algorithm, hierarchical clustering, density based clustering and grid based clustering. Partitioning clustering algorithms are

very useful when the clusters are of convex shape be recognized earlier. These paper analyses the two partitioning algorithms such as k-means and k-medoids. Hierarchical clustering algorithms are used when the disability in predicting the number of clusters in advance. Hierarchical clustering algorithm represent their output in form of dendogram.These algorithms are effective but the cost of formation ofdendogram is very high for huge datasets. There are two types of hierarchical clustering algorithms like agglomerative hierarchical clustering and divisive hierarchical clustering. In agglomerative algorithms, the hierarchy results from the clusters of increasing similarity are grouped to form larger ones and that larger clusters are divide into smaller having similar size and the number of cluster can ones of decreasing dissimilarities in divisive algorithms. Density based clustering algorithms are very useful for huge datasets because they can easily recognize noise and can distribute with clusters of arbitrary shape. It is based on the density. The DBSCAN is most widely used density based algorithms.Grid based clustering is systematic and good technique to make cluster of data. Iteratively grids are generated and according to that clusters will be created. It significantly reduces the computational complexity, especially for clustering huge data sets.This study gives readers a clear view of important data analysis method, i.e. clustering.

## REFERENCES

[1] Amandeep Kaur Mann, SurveyPaper on Clustering Techniques, International Journal od Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4, April 2013

[2] Amandeep Kaur Mann &Navneet Kaur, Review Paper on Clustering Techniques, RIMT, Global Journal Of Computer Science And Technology Software & Data Engineering, Vol 13, Issue 5, 2013.

[3] DeeptiSisodia , Clustering Techniques: A Breif Survey of Different Clustering Algorithms, International Journal of Latest Trends in Engineering and Technology(IJLTET)

[4] Wei-keng Liao, A Grid-based Clustering Algorithm using Adaptive Mesh Rfinement, Appears in the 7th Workshop on Mining Scientific and Engineering Datasets 2004.

[5] NidhiAshwin Shah, Survey On Different Grid Based Clustering Algorithms of Data Mining, IJARIIE-ISSN(O)-2395-4396,Vol-3 Issue-1 2017

[6] SauravKaushik, An Introduction to Clustering and different methods of clustering, November 3, 2016

[7] Mythili S, Madhiya E, An Analysis on Clustering Algorithms in Data Mining, International Journal of Computer Science and Mobile Computing, Vol. 3, Issue. 1, Jan. 2014

[8] Dr. S. K. Jayanthi, C. KaviPriya, Clustering Approachfor Classification of Research Articles Based on Keyword Search,International Journal of Advanced Research in Computer Engineering & Technology(IJARCET), Volume 7, Issue1, January 2018

[9] MridumurchanaKashyap, A Comparative Study on Partition-based Clustering Methods, International Journal of Creative Research Thoughts (IJCRT), Vol. 6, Issue 2, April 2018

[10] Rashi Chauhan, Clustering Techniques : A Comprehensive Study of Various Clustering Techniques, International Journal of Advanced Research in Computer Science (IJARCS) Vol. 5, May-June-2014

[11] SwarndeepSaket J, An Overview of Partitioning Algorithms in Clustering Techniques, International Journal of Advanced Research in ComputerEngineering& Technology (IJARCET), Vol.5, Issue 6, June2016

[12] T. SoniMadhulatha , An Overview On Clustering Methods, IOSR Journal of Engineering, Apr. 2012,

Paper ID: CS-03

# Review of Detection of Fruit Diseases Using Image Processing and Machine Learning Techniques

**Priyanka Vijay Barhate**,

*Research Scholar,*

*DOCS, DNCVPs Shirish Madhukarrao Chaudhari College, Jalgaon*

**Dr. Bhojraj Hanumant Barhate,**

*Asso. Professor,*

*Dept. of Computer Science, P. O. Nahata College Bhusawal*

**ABSTRACT -** *Diseases in fruit cause devasting problems and results in economic and agricultural industry loss. Earlier detection of fruit diseases is done manually. Nowadays Computer Vision plays an important role to detect diseases in fruits. Image processing and machine learning techniques are used to identify defects in fruits. In the first step region of interest is separated using image segmentation technique with clustering algorithms which is followed by extraction of some features and finally image is classified either normal or diseased. The main objective of this paper is to present survey of existing methodologies and uncover the future research issues.*

*Keywords -BPNN, KNN, SVM, K-means clustering, LBP, CCV.*

## I. INTRODUCTION

In India millions of people depend on agriculture sector and it is the backbone of economy of our nation. Food is essential part of human's life and one of the best natural sources of food is fruit. India is the second largest producer of fruits in the world [1]. However this production is greatly diminished because of the fruits affected by diseases, pests and make them non-edible or waste [2]. Fruit industry is one of the major drives to grow economy of country. Variety of fruits is grown as mango, banana, pomegranate, citrus, grapes, guava, pineapple and apple. Fruits provide lot of nutrients which are essential for our day to day life but there are various diseases which affects fruits. Diseases may appear at any stage growing season, harvesting, handling, transport, post-harvest storage and marketing conditions or after purchasing by the consumer. Diseases in fruit can cause major losses in marketing and it affects overall economy of nation. The traditional method of fruit disease identification consists of necked eye observation by experts. It requires skill, numerous times, expenses and continuous monitoring which is not feasible for human/ monitoring large farms. Hence automatic detection of fruit disease is very crucial as it helps to overcome drawbacks of manual detection [Karthikeyan 12] .

## II. LITERATURE REVIEW

[1] Work in two phases for grading and classification of Anthracnose fungal disease for fruits. Three fruits mango, grape and pomegranate were considered. In the first phase, various segmentation techniques thresholding, region growing, K-means clustering and watershed are used to find the affected area. In the second phase Runlength Matrix is used to extract texture features. MATLAB 7.0 tool box is used to develop algorithm. Dataset of 600 fruits image samples are used for experimentation. Neural network classifier is used to classify healthy and anthracnose affected fruit. ANN and BPNN classifiers are used. Segmentation using K-means clustering founds to be better technique over region growing. 87% highest recognition and classification accuracy is

obtained for normal mango and the lowest 71.5% for anthracnose affected pomegranate. The average accuracy fiund for normal type is 84.65% and 76.6% for anthracnose affected type. BPNN classifier is found suitable for this work.

[2] Described detection of disease and fruit grading using image processing techniques. Work has done on two fruits apples and grapes. Color, morphology and texture vectors are used to extract features. Artificial Neural Network and NN tool box was used for implementation. They had obtained the better result for color and morphology as compared to texture. Two methods were used for automated grading of fruit: spread of disease and automated calculation of mango weight. K-means clustering was used for spread of disease and weight of mango by pixel count was calculated by mathematical formula.

[3] Suggested image based processing classification approach for apple fruit disease. Diseases considered namely apple blotch, apple rot and apple scab for conducting the experience. For image segmentation K-means clustering is used. color, texture and shape based features are fused to form more distinctive feature. Color Coherence Vector (CCV), Histogram, Local Binary Pattern (LBP), Complete Local Binary Patterns are used for feature extraction. For fruit disease detection Multiple class Support Vector Machine (MSVM) is used. Dataset of 320 sample images were used for implementation.

[4] Presented the work to detect fruit diseases using Color, Texture and Artificial Neural Network(ANN). Grapes, Apples and Pomegranates were used to detect various diseases. System used two image databases. One for implementation of query images and other for training of already stored disease images. Four feature vectors color, morphology, texture and structure of hole on fruit used as feature extraction. Speed Up Robust Feature (SURF) algorithm and Blob detection method were used for feature extraction. K-means clustering method and ANN was used for segmentation and pattern matching, classification of diseases. OpenCV library is applied for implementation. Work can be extended to estimate the severity of the disease.

[5] Proposed the web based system to identify the bacterial blight disease on pomegranate fruit. For feature extraction parameters color, morphology and CCV are used. K-means clustering is used for image segmentation. To classify normal or

disease fruit Support Vector Machine (SVM) is used. Author has obtained the best result for morphology feature. The overall 82% system accuracy is obtained to identify the pomegranate disease.

[6] Presented a web based search tool AgroSearch for diseases and pest identification of pomegranate fruit. Color, Color Coherence vector and morphology features are extracted from images. K-means algorithm and Multi class Support Vector Machine were used for segmentation and classification. Results obtained as morphology features are best as compare to other features. Suggested work can be extended for advancement of technologies like cloud computing, Internet of Thing and train the system for other fruit disease detection.

[7] Proposed approach for identification and classification of fruit diseases. To find the region of interest fuzzy C-means and K-means clustering methods are used. Gray Level Co-occurrence Matrix is used for feature extraction. Dataset of 243 image samples of 10 various fruits were used for experiment. KNN algorithm is used to classify various diseases. Author has achieved the results better with K-means over C-means clustering algorithm.

[8] Presented the analysis of diseases caused due to fruit harvesting. This analysis shows that blight is the most common disease among most of the fruits. Various techniques as image pre processing, feature extraction, Image denoising, Gausian noise, shot noise , median filter , Mean filter, Contrast enhancement schemes, particle filter. This paper presents comparative analysis of techniques used to analyse fruit diseases.

[9] Presented the work to classify four types of orange diseases citrus canker, melanose, brown rot and stubborn disease and to calculate the disease severity. K-means clustering and SVM is used to classify disease. Fuzzy logic based model is proposed to measure disease severity. Authors have suggested work can be extended to achieve the better accuracy using soft computing technique and validated with more number of sample.

[10] Proposed the system which is used to grade the fruit and identify the defect for apple using image processing techniques. Image is pre-processed by rgb2gray method and median filter. For segmentation iterative tri-class threshold based on Otsu's method is used. Features texture and shape are

extracted by normalized symmetric GLCM method. Fruit is categorized into four categories by using KNN method with 95% accuracy.

[11] Proposed system to identify fruits and analysis of disease using machine Learning Techniques. 5 types of fruits apple, mango, orange tomato and pomegranate are taken for experimentation. Two common diseases are identified anthracnose and fruit rot. To access the features Gray Level Co-Occurrence Matrix (GLCM) is used. The Multiclass Support Vector Machine (MSVM) classifier is used to identify type of fruit and detect the disease appear on fruit and classified the type of disease. To measure the severity of diseases fuzzy logic is used. Dataset consists of 150 images. Proposed system consists of three phases. First phase identifies the type of fruit. Second phase detect the defected fruit and third phase identifies type of disease. Accuracy 92.17% is obtained for antracnose and fruit rot diseases. Author has suggested accuracy can be increased when the number of datasets increased with good quality images.

[12] Presented image processing based system to identify passion fruit diseases. Healthy and two types of passion fruit diseases as fruit scab and woodiness were considered for this experimentation. Three stages were used for segmentation such as mild, moderate and severe. K-means clustering was used for segmentation. Features are extracted with Local Binary Pattern and classification is done with Support Vector Machine algorithm. Average accuracy of 79% is obtained for identification of passion fruit diseases and 66% for various stages. Dataset of 87 images were considered for this approach.

[13] Proposed the framework for disease identification for apple. Three types of diseases rot, scab, blotch were considered. Texture, color and shape features are used . Improved sum and difference histogram (ISADH), color coherence vector (CCV), gray level co-occurrence matrix (GLCM), completed local binary pattern (CLBP) and zernike moments (ZM) features are used in order to identify and classify the diseased and normal fruits. To achieve the greater accuracy author has combined the features as ISADH+GLCM and ISADH+CLBP+ZM. Dataset of 280 sample images of normal and diseased apples were used. Upto 96% classification accuracy is achieved with combined features

[14] Developed the mobile application to determine various infections on leaves, fruits and stem of plant using image processing and cloud computing techniques. System generates the automated database to examine infections. Spots, color, shape, area and texture features are used for experimentation. K-Means, Neural Network and SVM classification techniques are used.

[15] Presented machine vision based recognition system to identify diseases of mango. Features are extracted using K-means clustering and classification is done with Support Vector Machine (SVM). System has obtained 94.13% accuracy.

**Survey of existing methodologies and results:**

| References | Fruits with diseases | Features | Segmentation | Classification | Result |
|---|---|---|---|---|---|
| Jagadeesh Pujari et al [2013] | Mango, Grape, Pomegranate, Anthracnose | Texture | Thresholding, region growing,K-means clustering, watershed, Runlength Matrix | ANN and BPNN | Normal: 84.65 % Affected: 76.6% |
| Monica Jhuria, et al [2013] | Apples: Apple Scab, Rot Grapes : black rot, Powdery Mildew | Color, texture morphology | Erosion | BPNN, ANN | Morphology gives better result 90% over other two features. |
| Shiv Ram Dubey, et al [ 2015] | Apple : Rot, Scab, blotch | Color, texture, shape | CCV, LBP, CLBP, Histogram | MSVM | Shape feature is not better suited for this puposc |
| ` | Grapes: black rot, powdery mildew, downy mildew Apple: scab,rot ,blotch Pomegr | Color, morphology, texture, structure of hole | Speed Up Robust Feature (SURF) algorithm and Blob detection, K-means | ANN | To do smart farming helps to take time to time decision and reduce loss of fruit |

| | | | | | |
|---|---|---|---|---|---|
| | anate: Bacterial blight, aspergillus rot, graymold | | | | due to disease. |
| ManishaBhange, et al[2015] | Pomegranate : blight | Color, morphology, CCV | K-means | SVM | Overall system accuracy is 82%. |
| Sunil More et al, [2016] | Pomegranate: bacterial blight, wilt,thrip, fruit borer, scab | Color, CCV, Morphology | K-means | MSVM | Morphology features are best as compare to other features. Accuracy of system is 84% |
| Y. H. Sharath Kumar, et al [2017] | Apple, banana, citrus, grape,guava, mango, papaya, peach, pomegranate, watermelon | Texture | Gray Level Co-occurence Matrix (GLCM), Fuzzy C-means, K-means | KNN | Achieved better result with K-means over C-means clustering algorithm. |
| SantiKumariBehera, et al [2018] | Orange : citrus canker, melanose, brown rot, stubborn | Texture | GLCM K-means, | Multiclass SVM | 90% classification accuracy |
| Patne&Ghonge, [2018] | Apple | Texture, Shape | Otsu's method for preprocessing, GLCM for Feature Extraction | KNN | 95% Accuracy |
| SantiKumariBehera, et al [2019] | Apple, mango, orange, tomato, pomegranate, Anthracnose, | Color, texture, shape | GLCM, Fuzzy Logic technique | SVM | 92.17% accuracy is obtained for antracnose and fruit |
| | fruit rot | | | | rot |
| SBDH Dharmasiri [2019] | Passion Fruit : Fruit Scab, woodiness | Morphology, texture, color | LBP K-means | MSVM | 79% accuracy obtained |
| Ayyub&Manjramkar [2019] | Apple | Color, Texture, Shape | Improved sum and difference histogram (ISADH), color coherence vector (CCV), gray level co-occurrence matrix (GLCM), completed local binary pattern (CLBP) and zernike moments (ZM) | Multi class SVM | Author has obtained average classification accuracy is upto 96% using feature combination. |
| Mia et al [2019] | Mango : Anthracnose, Bacterial Black Spot, Die Back, Mango Malformation and Powdery Mildew | | K-Means Clustering | SVM | System has obtained 94.13% accuracy |
| Ramya et al. [2020] | Mango : Powdery Mildew, Anthracnose and soothy Mould | Spots, color, shape, area and texture : K-means | K-means | NN, SVM | Accuracy indicated in alert box through mobile application |
| Mia et al. | Mango : Anthrac | | K-Means | SVM | System has |

| [2019] | nose, Bacterial Black Spot, Die Back, Mango Malformation and Powdery Mildew | | Clustering | | obtained 94.13% accuracy |
|---|---|---|---|---|---|

## III.    CONCLUSION

This survey help researchers for collecting knowledge about methodologies used for fruit disease detection such as segmentation, feature extraction, and classification techniques.

Most of the system comprises image pre-processing, segmentation, feature extraction, and classification for detection of normal or infected fruit. Various image processing and machine learning techniques were used to identify diseases in fruits. From literature it is evident that color, texture and morphological features are most suitable to identify and classify the diseases in fruits. Most commonly used classification techniques to identify fruit diseases are artificial neural networks (ANN) and support vector machine (SVM). Automatic detection of fruit diseases would solve the problem of expensive domain expert.

### REFERENCES

1. JagadeeshDevdasPujari ,RajeshYakkundimath and AbdulmunafSyedhusainByadgi "Grading and Classification of Anthracnose Fungal Disease of Fruits based on Statistical Texture Features" International Journal of Advanced Science and Technology Vol. 52, March, 2013
2. Monica Jhuria, AshwaniKunar, RushikeshBorse" Image Processing for smart farming: Detection of disease and fruit grading" IEEE Second International Conference on Image Information Processing 2013
3. Shiv Ram Dube, Anand Singh Jalal, "Apple disease classification using color, texture and shape features from images", SIVip, Springerseptember 2015.
4. AshwiniAwate, DaminiDeshmankar, GayatriAmrutkar, UtkarshaBagul, SamadhanSonavane" Fruit Disease Detection using Color, Texture analysis and ANN" IEEE 2015.
5. ManishaBhange , H.A. Hingoliwala "Smart Farming: Pomegranate Disease Detection Using Image Processing" Second International Symposium on Computer Vision and the Internet (Vision Net 15)
6. Sunil More, MininathNighot "AgroSearch: A Web Based Search Tool for Pomegranate Diseases and Pests Detection Using Image Processing" ACM Proceedings of Second International Conference on Information and Communication Technology for Competitive Strategies, March 2016.
7. Y. H. Sharath Kumar, G. Suhas" Identification and Classification of Fruit Diseases" International Conference on Recent Trends in Image Processing and pattern Recognition 2016, CCIS 709,382-399, 2017 (Springer)
8. Kawaljit Kaur. ChetanMarwaha " Analysis of Diseases in Fruits using Image Processing Techniques" IEE, International Conference on Trends in Electronics and Informatics ICEI 2017.
9. SantiKumariBehera, Lipsarani Jena, Amiya Kumar Rath and Prabira Kumar Sethy "Disease Classification and Grading of Orange using Machine Learning and Fuzzy Logic", IEEE, International Conference on Communication and Signal Processing, 3-5, April 2018.
10. G. Patne and P. Ghonge, "Automization of agriculture products defect detection and grading using image processing system," Int. J. Comput. Sci. Eng. Inf. Technol. Res., vol. 8, no. 3, pp. 25–32, 2018.
11. SantiKumariBehera, Amiya Kumar Rath, Prabira Kumar Sethy "Automatic Fruits Identificationand Disease Analysis using Machine Learning Techniques" International Journal of Innovative Technology and Exploring Engineering Vol 8, Issue 652, April 2019.
12. S.B.D.H. Dharmasiri, S. Jayalal" Passion Fruit Disease Detection using Image Processing" Smart Computing and System Engineering 2019.
13. S. R. N. M. Ayyub and A. Manjramkar, "Fruit Disease Classification and Identificatio using Image Processing," in 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) IEEE Xplore, Mar. 2019, pp. 754–758.
14. Ramya, P. Kumar, K. Sivanandam, and M. Babykala, "Detection and Classification of Fruit Diseases Using Image Processing Cloud Computing," in 2020 International Conference on Computer Communication and Informatics (ICCCI), Jan. 2020, pp. 1–6.
15. Md. R. Mia, A. C. Chhoton, M. H. Mozumder, S. A. Hossain, and A. Hossan, "An Approach for Mango Disease Recognition using K-Means Clustering and SVM Classifier," in 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), Nov. 2019, pp. 404–409.    doi:    10.1109/SMART46866.2019.9117273

**Paper ID: CS-04**

# Routing Protocols in VANET: A Comprehensive Study

**Ms. Poonam Murlidhar Mahajan**

*Asst. Professor, Dept. of Computer Science & IT, Bhusawal Arts, Science & P. O.Nahata Commerce College, Bhusawal Dist. Jalgaon (MS), India Email- amritamah@gmail.com,*

**Dr. Gouri Milind Patil**

*Asst. Professor, Dept. of Computer Science & IT, Bhusawal Arts, Science & P. O.Nahata Commerce College, Bhusawal Dist. Jalgaon (MS), India gourimpatil@gmail.com*

***ABSTRACT - In today's computing environment, VANETs communication provides a plethora of applications that can be divided into three categories: safety-oriented applications, value-added apps, and comfort-oriented applications. Putting a greater emphasis on safety-related applications there are times when emergency warnings, stopped vehicle warnings, lane shifting warnings, road conditions warnings, and other alerts must be sent to vehicles on the same route. However, due to the VANET's extremely dynamic nature, there is a risk of network failure. This review aims to determine the most appropriate technique for data dissemination in this dynamic vehicular context.***

***Keywords -Vehicular Ad Hoc Network, Cluster head, Surveillance, WSAN, WSN etc.***

## I. INTRODUCTION

VANET is a type of wireless multi-hop network that is constrained by the need for rapid topology changes due to high node mobility (shown in Figure 1[1]). Inter-vehicle communication is becoming a viable topic of research, standardisation, and growth as the number of vehicles equipped with computing technology and wireless communication devices grows. Routing protocols play an important role in network communication. Any routing protocol's principal goal is to identify the best way to communicate between nodes (vehicles).

Because of the high mobility of the vehicles, the topology of the VANET is continually changing. The majority of VANET vehicles these days are capable of merging their own system with other existing technologies such as GPS (GPS).

The distribution of essential medical emergency signals is the responsibility of VANETs. In order to preserve human lives, these signals must be delivered on time. For this extremely dynamic environment of VANET, a correct mobility model is essential in order to implement VANET efficiently and accurately.

Security risks are frequently overlooked in travel comfort applications since cooperative driving is assumed. Integrating security methods into VANETs is critical during the broadcast of safety data. Customers will not embrace warning systems until trust, security, and reliability are guaranteed. The most critical security concern within a VANET is the establishment of trust through providing trustworthy applications. Integrating security measures into VANET applications, on the other hand, will lengthen the time it takes for messages to arrive. VANETs can be used for a variety of purposes, including collision prevention, safety, blind crossing, dynamic route routing, real-time traffic condition monitoring, and so on.
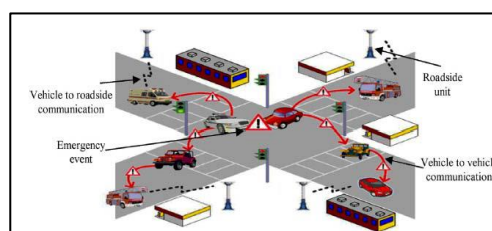


***Figure 1. VANET Architecture***

## II. ROUTING IN VANET

The development of a dynamic routing protocol that can help distribute information from one node (vehicle) to another is one of the primary issues in the architecture of a vehicular ad-hoc network. Only a few protocols created for the MANET environment have been tested on VANET.

The challenge remains, though, in reducing the time spent transferring information from one node to the next. Overcoming these obstacles in MANET protocols can aid in the development of real-time VANET applications.VANET can be divided into five distinct groups shown in Figure 2[2] namely as:

- Ad-hoc or Topology Based Protocols
- Position Based Routing Protocols
- Cluster Based Protocols
- Broadcast Based Protocols
- Geo-cast Based Protocols
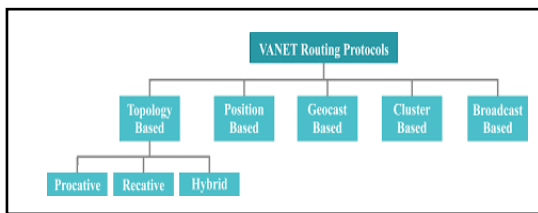


*Figure 2. Routing Types in VANET*

## III. TOPOLOGY BASED ROUTING

Optimized Link State Routing (OLSR) is a proactive or table-driven routing technology that works on the principle of link sensing. Frequent link failures result in lower QoS parameters such as throughput, end-to-end delay, latency, and link bandwidth utilisation, demonstrating that link quality is an important statistic to consider as a study topic when choosing a routing protocol. When making routing decisions, the OLSR evaluates the quality of the link.[3].

When compared to a black hole attack, the reactive routing system AODV is more vulnerable to flooding [3]. AODV has the disadvantage of requiring additional system resources and memory [4].For discovering a link failure, the reactive protocol DSR requires MAC layer support [5]. DSR performs route discovery by flooding the network with RREQ packets [4] [6].TORA, a hybrid routing protocol, is incompatible with vehicle ad-hoc networks. FSR and AODV, on the other hand, exhibited promising

performance in networks with high density nodes or full load [7].By boosting the efficiency and reliability of vehicle traffic information message delivery, the Ant Colony Hybrid Routing Protocol (ACOHRP) improves the service quality of intelligent transportation systems (ITS) [8].

## IV. POSITION BASED ROUTING

Position-based routing is a set of rules with varied degrees of sophistication that determines where inbound data packets from different places should be routed. The sender's outgoing ID is usually used to determine the location of data origin, however GPS position or signal triangulation can also be used [9].

The use of GPSR results in a reduction in route delay. GPSR cannot be used in its original state. To address this issue, a modified version of GPSR known as Advanced Greedy Forwarding (GPSR-AGF) [10] [11] [9] was presented. The HELLO packets in this version include information about the destination node's speed and direction.

The invention of Geographic Source Routing provided another technique to solve the shortcomings of GPSR (GSR). Road layouts are utilised in GSR to determine the destination route. It is mostly reliant on the availability of road maps, which are combined with shortest path finding algorithms.

Location Assisted Routing (LAR) [12] is another major location-based routing technique. It's a reactive protocol that employs node position information to reduce routing overheads that previous reactive protocols like AODV and DSR have highlighted. To calculate the next hop, LAR primarily uses two methods: one based on window size and the other based on distance variation.

Greedy Perimeter Coordinator Routing presents an alternative methodology that does not rely on external resources such as maps or infrastructure (GPCR). GPCR is founded on the idea that in an urban area, street pathways produce a planner graph. Packets are always routed to the junction edge's node (coordinator). The coordinator is the sole one who makes routing decisions.

Routing decisions in Connectivity Aware Routing (CAR) are based on the connected pathways between the source and destination pairs. Another position-based approach for urban environments is Anchor-Based and Traffic Aware Routing (A-STAR)

[13]. It uses usage data for a certain route to find the anchor path (number of junctions) that will provide the most connectivity.

## V. CLUSTER BASED ROUTING

CBR's fundamental concept is to build a network architecture based on small groupings of cars known as clusters, as depicted in Figure 3 [1]. One of the cars in a cluster serves as the cluster-head. The size of the cluster is determined by the routing algorithm's design, which might be dependent on the number of cars in the cluster or their geographical location [14].
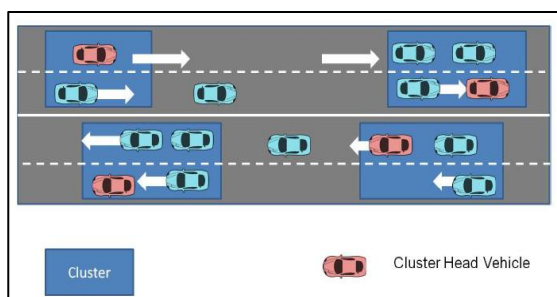


*Figure 3. Cluster Based Routing*

Introducing one cluster head, zero or more cluster members, and one or more gateways to communicate with other cluster heads, LORA-CBF is a reactive geographic routing based system for V2V communication. The LORA-architecture CBF's aids in reducing needless retransmission overheads. Because it updates the location information of clusters rather than individual vehicles, LORA-CBF may be ideal for highly mobile environments such as VANET [15] [16].

## VI. BROADCAST BASED ROUTING

When a message needs to be conveyed to a vehicle that is out of range, the broadcast strategy is typically utilised. The Nth-Powered P-persistent Broadcast protocol (NPPB) and BROADCOMM are two well-known protocols that use the broadcasting notion. A hierarchical roadway system is emulated in BROADCOMM, and the entire region is divided into virtual cells. Cell members form a hierarchy of Cell Reflectors (CRs), which operate as a base station, collecting messages for a specific call as well as messages from neighbouring cells.

A probabilistic broadcasting strategy is used in NPPB to reduce broadcasting storms in dense VANETs and effectively convey emergency messages. [17].

## VII. GEOCAST BASED ROUTING

Geo-cast routing's main purpose is to narrow down the search for a future hop to a predetermined Zone of Relevance (ZOR). For low-density situations, ROVER performs well. Only control packets are flooded into the network in ROVER, and data is distributed via a unicast manner, which improves the routing scheme's efficiency and dependability. A ZOR is a rectangular window that is produced at the back of the source node in ROVER.

Rather than using greedy selection criteria, GROOV is built on the idea of identifying the best suitable relay node. VCARP makes optimal routing decisions based on vehicle information such as destination, location, and packet cache status [18].Packet retransmissions are avoided by temporarily keeping them in a cache. VCARP produces improved PDR and lowers routing overheads in VANET settings, according to simulation tests. The concept of carry and forward was used by the designers of the Vehicular Ad-hoc Network (VADD) protocol [19].

## VIII. CONCLUSION

Vehicle-to-vehicle (V2V) and infrastructure-to-vehicle (I2V) communication rely heavily on routing. Various VANET routing protocols were discussed in this article. Designing an efficient routing mechanism for all VANET applications is quite tough. As a result, a thorough examination of numerous VANET protocols, as well as a comparison of their many aspects, is required in order to come up with a novel approach to building a VANET protocol. The performance of VANET routing protocols is influenced by a number of factors, including mobility model, driving environment, and so on.

As a result, the focus of this work has been on the thorough survey. As a result, the article proposes that future routing protocol research should incorporate the integration of sensor networks with wired networks (i.e. Internet). It's also worth noting that some protocols, such as TORA, aren't appropriate for use in a vehicular network.This article provides a number of protocols that can be used with VANET, the majority of which being Cluster-Based Protocols such as LORA-CBF.The data collected from the

sensor nodes must typically be transferred to a server for additional processing in security and environmental monitoring applications. Because each environment necessitates a unique set of routing methods, more study is required to address these issues.

## REFERENCES

[1] E. Khoza, C. Tu and P. A. Owolawi, "An Ant Colony Hybrid Routing Protocol for VANET," *IEEE Xplore,* 2019.

[2] S. u. Rehman, M. A. Khan, T. A. Zia and L. Zheng, "Vehicular Ad-Hoc Networks (VANETs) - An Overview and Challenges," *Journal of Wireless Networking and Communications,* vol. 3, no. 3, pp. 29-38, January 2013.

[3] R. Jain and I. Kashyap, "An QoS Aware Link Defined OLSR (LD-OLSR) Routing Protocol for MANETS," *Wireless Personal Communications,* p. 1745–1758, 2019.

[4] A. Kumar and M. Sinha, "Design and analysis of an improved AODV protocol for black hole and flooding attack in vehicular ad-hoc network (VANET)," *Journal of Discrete Mathematical Sciences and Cryptography ,* vol. 22, no. 4, pp. 453-463, 2019.

[5] S. Alaparthi, S. R. Parvataneni, C. S. Vaishnavi, P. Sathvika, M. Chandrika and P. Sharanya, "Dynamic Source Routing Protocol—A Comparative Analysis with AODV and DYMO in ZigBeebased Wireless Personal Area Network," *IEEE Xplore,* 2019.

[6] V. Soi and D. B. S. Dhaliwal, "Performance comparison of DSR and AODV Routing Protocol in Mobile Ad hoc Networks," *International Journal of Computational Intelligence Research ,* pp. 1605-1616, 2017.

[7] R. Sharma and D. K. Lobiyal, "Proficiency Analysis of AODV, DSR and TORA Ad-hoc Routing Protocols for Energy Holes Problem in Wireless Sensor Networks," *Procedia Computer Science,* p. 1057 – 1066, 2015.

[8] P. M. Kumar and R. Hariharan, "Performance of GPSR with different Routing Protocols in Wireless Sensor Networks," *International Journal of Engineering and Advanced Technology (IJEAT),* vol. 9, no. 1, pp. 6858-6863, October 2019.

[9] A. Silva, N. K. M. Reza and A. Oliveira, "An Adaptive GPSR Routing Protocol for VANETs," *15th International Symposium on Wireless Communication Systems (ISWCS),* August 2018.

[10] X. Yang, M. Li, Z. Qian and T. Di, "Improvement of GPSR Protocol in Vehicular Ad Hoc Network," *IEEE Access,* vol. 6, 12 July 2018.

[11] R. C. Muniyandi, F. Qamar and A. N. Jasim, "Genetic Optimized Location Aided Routing Protocol for VANET Based on Rectangular Estimation of Position," *Applied Sciences,* vol. 10, no. 17, 20 August 2020.

[12] N. K. Gupta, R. S. Yadav and R. K. Nagaria, "Anchor Based Geographical Routing in WSN," in *ICSCA 2020: Proceedings of the 2020 9th International Conference on Software and Computer Applications*, 2020 .

[13] S. P. Ardakani, "ACR A CLUSTER-BASED ROUTING PROTOCOL FOR VANET," *International Journal of Wireless & Mobile Networks (IJWMN),* vol. 10, no. 2, pp. 39-53, April 2018.

[14] P. K. Pagadala and D. N. S. Kumar, "ROUTING PROTOCOLS IN VEHICULAR AD HOC NETWORKS: A SURVEY," *GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES,* vol. 4, no. 12, pp. 93-107, December 2017.

[15] L. Zhou, G. Cui, H. Liu, D. Luo and Z. Wu, "NPPB: A broadcast scheme in dense VANETs," *Information Technol-ogy Journal,* vol. 9, no. 2, p. 247–256, 2010.

[16] S. Kumar and Kantveer, "Vehicle Assisted Data Delievery Technique To Control Data Dissemination In Vehicular AD - HOC Networks (Vanets)," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH,* pp. 36-41, OCTOBER 2015.

[17] 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.

[18] P. F. Zhao, K. Liu, Y. Zhang, T. Zhang and F. Liu, "A Clustering-Based Fast and Stable Routing Protocol for Vehicular Ad Hoc Networks," *Journal of Physics: Conf. Series 1060,* 2018.

[19] A. Yanez, S. Cespedes and J. Rubio-Loyola, "A baseline for context-aware system for safety messages dissemination in VANETs," *Revista Facultad de Ingeniería Universidad de Antioquia,* vol. 93, pp. 9-18, Oct./Dec. 2019.

Paper ID: CS-05

# A Review on Flower Image Identification and Classification

**Shubhangi S. Bhangale**

*Asst.Professor,*
*School of Computer Sciences,*
*M.J.College,Jalgaon.*
bhangaless03@gmail

**Dr. Gouri M. Patil**

*Asst. Professor, Dept. of Computer Science & IT,*
*Bhusawal Arts, Science & P. O.Nahata Commerce*
*College, Bhusawal Dist. Jalgaon (MS), India*
gourimpatil@gmail.com

**ABSTRACT -** *Automated flower recognition system has been studied from many years, difference between these studies comes from which features were extracted from the flower image and the recognition algorithm that was used to recognize the flower species. The most important attributes to be considered are color, texture, and shape for selecting the feature from flower images. For classification different classification models are present such as ANN, KNN, SVM, CNN and many more. This paper reviews the technologies and algorithms which are available for segmentation, feature extraction, classifying, detecting the flower images.*

## I. INTRODUCTION

Flowers are not only used just for their aesthetic sense but also for nutritive and medicinal properties. Due to vast existence, complex structure and unpredictable variety of classes in nature flower identification remain a challenge in Image processing and Computer Vision community. Due to natural complexities, it's highly undesirable to perform normal segmentation or feature extraction in moderate accuracy on benchmark datasets. Although some feature extraction techniques reaches state of the art accuracy in classifying flowers, still there is a requirement for a strong and efficient system to automatically identify and recognize flower species at a larger scale. But it is a very much challenging computer vision problem because of the large similarity between flower classes. Flowers from different species may seem same, for instance, Dandelion and Colt's Foot, as shown in Figure-1(a) and flowers from the same species may have different look, for example, the Pansy flower in Figure 1(b) [16].
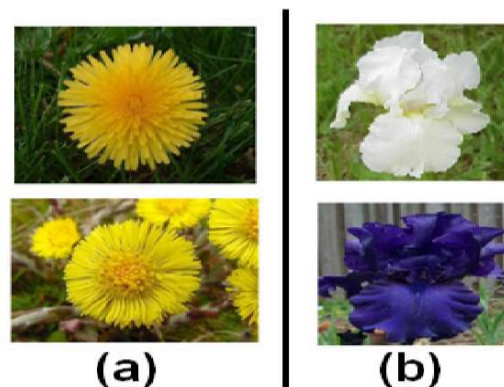


Figure-1 (a) Two visually similar flowers from the different species. (b)Two visually dissimilar flowers from same species

So there was need a model or a system that able to process and memories these large amounts of flower species data and can be trained easily and through which these above problems can be solved. This kind of model can develop with the help of image processing, where content-based image indexing techniques are used to analyze and describe images based on their visual content and appearance.

## II. LITERATURE REVIEW

To categorize various species of oxford-102 flowers dataset they used the transfer learning approach employing DenseNet121 architecture. They preprocess, normalize and resize images and then fed them to pre-trained model. They split the dataset into three sets that are train, validation, and test. They achieved the accuracy of 98.6% for 50 epochs which is better than other deep-learning based methods for the same dataset [1]. They divided the work into 4 main steps: image enhancement, cropping of images, image segmentation and features extraction. Back Propagation ANN was used for the classification part. They used Oxford 102 flowers dataset. They achieved more accuracy than other researchers [2]. They extract color, shape and different texture features from flower images and create three different feature vectors with Haralick, Tamura and Gabor textures. Then classify each with different classifiers such as SVM, Decision Trees and K-NN. In order to identifying Ayurveda plants the results are compared with each other's individual performances [3].

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 [4]. Flower classification system that uses active contour segmentation technique. After segmentation he extracted LBP and SURF features then concatenated them and send for classification with Quadratic SVM. As compared to other models this is significant, comparable and provided satisfactory Result [5]. Deep learning approach using Convolutional Neural Networks (CNN) to recognize flower species with high accuracy was proposed. They extracted flower features like color, shape and texture using HOG descriptor, GLCM as well as LBP's, Hu moments and Zernike moments respectively, also for segmentation grabcut segmentation algorithm was used. These global feature vectors are concatenated and then trained over a machine learning classifier. For classification and overFeat network was used. Transfer learning approach yields impressive accuracy [6].

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 [7]. They created dataset from scanned images of front and back side of leaves of commonly used Ayurvedic medicinal plants. Unique combinations of morphological, color and texture feature have been identified that maximizes identification rate of green leaves. They achieved highest identification rate using MLP Neural Network classifier for both dry and green leaves. Composition of both side feature increase identification rate when green leaves are used to identify plants [8]. They transform the image to the Lab color space. Then, apply the OTSU thresholding algorithm on each Lab component, independently. Finally, to choose the best result, they used a supervised measure for segmentation results evaluation. They obtain good results within short execution-time [9].

A hybrid methodology using MKL – SVM with multi-label classification algorithms was proposed. In this Otsu's methodology was applied for segmentation. With color, Texture feature morphology features also extracted from a flower image such as Disc flower, Corona, Petal type, No. of petals, Aestivation of a petal and flower class. This approach achieved relatively good classification accuracy with optimum possible extracted features [10]. The features are extracted from images of medicinal plant leaves and then classified. The features considered are the shape, textural and color features. Then some machine learning classification techniques such as KNN and SVM are used for classification and a comparison is made among their performances [11]. They created a good dataset using image processing techniques which include image filtering, segmentation, region detection and feature extraction for building neural network. They used RGB to HSV conversion for color feature, GLCM for texture feature and some categories are tested with neural network. They found that classification accuracy was affected by size of dataset [12].

A comparative study of image segmentation model DLISM for Accurate Identification of Medical Plants with a traditional ADDC method was done. They preprocess the image of leaf before classification to get only interested section of image, and then extract features of image using image processing techniques, afterword using GLC model and SVM DLNN model they classify the image. Comparison result shows that performance of proposed model is better in accuracy and time levels [13]. In Preprocessing they resize the image; segmentation was done by graph cut algorithm and RGB to grayscale conversion. In this system edge and color characteristics of flower images were considered to classify flowers. Hu's seven moment algorithm was applied to acquire edge characteristics. Using histograms they derived red, green, blue, hue, and saturation characteristics. K-nearest neighbor was used to classify flowers. The accuracy of this system

is more than 80% [14]. Feature extraction of flower are done using Scale Invariant Feature Transform (SIFT), and Segmentation-based Fractal Texture Analysis (SFTA) algorithms. Classification was done using Support Vector Machine (SVM) and Random Forests (RF) to recognize different kinds of flower. SVM provides better accuracy when used with SIFT algorithm While, RF algorithm provides its better accuracy with SFTA [15].

The comparative study of feature extraction techniques and classifiers are shown in Table-1.

| Sr. No | Year | Title | Segmentation/Feature Extraction methods | Methods/ Classifiers | Accuracy Rate |
|--------|------|-------|------------------------------------------|----------------------|---------------|
| 1 | 2019 [5] | A New Flower Classification System Using LBP and SURF Features | active contour technique based on chen- Vese method. LBP (Local Binary Patterns) with HOG descriptor and SURF (Speeded Up Robust Feature) | KNN,Ad aboost, Decision tree and Quadrati c SVM. | quadratic SVM outcomes with 87.2% accuracy |
| 2 | 2018 [2] | A Flower Recognit ion System Based On Image Processin g And Neural Network s | Segmentati on Chan-vese image processing segmentati on technique features extraction was done using color moments, GLCM and IM for color, textures and shape respectively | Back Propagati on ANN | **81.19%** |
| 3 | 2019 [18] | Supervis ed Learning Approac h for Flower Images using Color, Shape and Texture Features | Segmentati on was done by using threshold.C olor moments and Color histogram was used in color feature, whereas Gray level | artificial neural network | 96.0 %. |
| | | | co-occurrence matrix (GLCM) and Invarient moment (IM) was used for texture and shape feature. | | |
| 4 | 2014 [15] | An Automati c Flower Classific ation Approac h Using Machine Learning Algorith ms | OTSU threshold Scale Invariant Feature Transform (SIFT), and Segmentati on-based Fractal Texture Analysis (SFTA). | Support Vector Machine (SVM) and Random Forests (RF) to | When SVM used with SIFT it obtained better result than RF, and when RF used with SFTA it gives better results. |
| 5 | 2020 [19] | A Compute r Vision Approac h to Classify Local Flower using Convolut ional Neural Network | data collection process, image augmentati on method was used. | CNN model with "ReLu" activatio n function, "Adam optimizer " and the "Softmax " function to build the network layer. | 85% |
| 6 | 2017 [6] | Flower Species Recognit ion System using convoluti on Neural Network s and Transfer Learning | Segmentati on was done with grabcut segmentatio n algorithm. For color feature extraction Color Histogram descriptor along with HOG descriptor was used. For texture feature extraction Haralick | CNNs are applied for Compute r Vision applicati ons and overFeat network was used | CNN combine d with Transfer Learning approach yields impressiv e Rank-1 accuracie s of 73.05%, 93.41% and 90.60% using OverFeat , Inception -v3 and |

| | | | textures, which uses the concept of (GLCM) as well as LBP's were used. . For shape feature extraction Hu moments and Zernike moments were used. | | Xception architectures, respectively |
|---|---|---|---|---|---|
| 7 | 2017 [8] | Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples | leaf shape, color and texture features are extracted by HU invariant moments and Zernike moments, HSV respectively. | MLP Neural Network | Identification rates up to 99% have been obtained for this combination. |
| 8 | 2018 [20] | Medicinal Plant Leaf Recognition and Show Medicinal Uses Using Convolutional Neural Network | Canny algorithm is used as shape feature extraction | CNN and Inception-v3 as our pre-trained network | 98% |
| 9 | 2019 [22] | Flower Recognition Based on Convolutional Neural Network | Dataset is pre-process and augmented used. | CNN and Random Forest | 81.37% |

## III. CONCLUSION

In this work several papers were considered for survey and describe a general model for flower recognition. Natural image of flower contains complex background, to remove it use different segmentation techniques among these some techniques like graph cuts with MRF provides accurate result but it is slow and others like Otsu method is very fast but result may not be accurate for some images. Choosing the method for segmentationis totally depends on the problem doming and requirement of the research work [21].

Classifiers plays important role to test the data and check the accuracy of classification algorithm. To identify different flower images based on color and texture is challenging and most expensive task. It will be noticed that for classification many researchers took multiple feature attribute like shape, color, texture combine with different feature extraction method and found improvement on performance by taking multiple features than single feature. So it is good to take multiple features but it also very much important to choose the right set of features for the particular problem based on the characteristic of the problem and training dataset.

## REFERENCES

1. 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.
2. Almogdady, H., Manaseer, S., &Hiary, H. (2018). A Flower Recognition System Based On Image Processing And Neural Networks. International Journal Of Scientific & Technology Research, 7(11), 166-173.
3. Bandara, M., &Ranathunga, L. (2019, July). Texture dominant approach for identifying ayurveda herbal species using flowers. In 2019 Moratuwa Engineering Research Conference (MERCon) (pp. 117-122). IEEE.
4. 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.
5. Dhar, P. (2019). A new flower classification system using LBP and SURF features. International Journal of Image, Graphics & Signal Processing, 11(5), 13-20.
6. 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.
7. 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.
8. Manojkumar, P., Surya, C. M., &Varun, P. G. (2017). Identification of ayurvedic medicinal plants by image processing of leaf samples. In Third ınternational conference on research in computational ıntelligence and communication networks.
9. Najjar, A., &Zagrouba, E. (2012, June). Flower image segmentation based on color analysis and a supervised evaluation. In 2012 International Conference on Communications and Information Technology (ICCIT) (pp. 397-401). IEEE.
10. Patel, I., & Patel, S. (2019). Flower Identification and Classification using Computer Vision and Machine Learning

Techniques. International Journal of Engineering and Advanced Technology (IJEAT), 8(6).

11. Raghukumar, A. M., & Narayanan, G. (2020, March). Comparison Of Machine Learning Algorithms For Detection Of Medicinal Plants. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) (pp. 56-60). IEEE.

12. Siraj, F., Ekhsan, H. M., &Zulkifli, A. N. (2014, October). Flower image classification modeling using neural network. In 2014 International Conference on Computer, Control, Informatics and Its Applications (IC3INA) (pp. 81-86). IEEE.

13. Thella, P. K., &Ulagamuthalvi, V. A Comparative Analysis on Machine Learning Models for Accurate Identification of Medical Plants. PAIDEUMA JOURNAL Vol XIV Issue 4 2021 ISSN NO : 0090-5674

14. Tiay, T., Benyaphaichit, P., &Riyamongkol, P. (2014, March). Flower recognition system based on image processing. In 2014 Third ICT International Student Project Conference (ICT-ISPC) (pp. 99-102). IEEE.

15. Zawbaa, H. M., Abbass, M., Basha, S. H., Hazman, M., &Hassenian, A. E. (2014, September). An automatic flower classification approach using machine learning algorithms. In 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 895-901). IEEE.

16. Maria-Elena Nilsback, Andrew Zisserman, A Visual Vocabulary for Flower Classification, CVPR(2)2006: 1447-1454,2006.

17. Amira Ben Mabrouk, AsmaNajjar and EzzeddineZagrouba ,"Image Flower Recognition based on a New Method for Color Feature Extraction ",2014 International Conference on Computer Vision Theory and Applications (VISAPP), 2015 IEEE.

18. Galphade, A. V., &Walse, K. H. (2019). Supervised learning approach for flower images using color, shape and texture features. Int. Res. J. Eng. Technol, 6(5), 5682-5688.

19. Islam, S., Foysal, M. F. A., &Jahan, N. (2020, May). A Computer Vision Approach to Classify Local Flower using Convolutional Neural Network. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1200-1204). IEEE.

20. Sabarinathan, C., Hota, A., Raj, A., Dubey, V. K., &Ethirajulu, V. (2018). Medicinal plant leaf recognition and show medicinal uses using convolutional neural network. Int. J. Glob. Eng, 1(3), 120-127.

21. Zhang, X., Han, D., Bai, F., & Ma, Z. (2019, August). Flower Recognition Based on Convolutional Neural Network. In 2019 9th International Conference on Information Science and Technology (ICIST) (pp. 333-338). IEEE.

22. AsmaNajjar and EzzeddineZagrouba, Flower image segmentation based on color analysis and a supervised evaluation, 2012 International Conference on Communications and Information Technology (ICCIT),2012 IEEE.

**Paper ID: CS-06**

# A Review on Varietal Classification of Grain using Image Processing

**Shubhangi K. Patil**

*Assistant Professor, Department of Computer Science,Dhanaji Nana Mahavidyalaya, Faizpur*

*Shubhangipatil14@gmail.com*

**Dr. Gouri M. Patil**

*Asst. Professor, Dept. of Computer Science & IT, Bhusawal Arts, Science & P. O.Nahata Commerce College, Bhusawal Dist. Jalgaon (MS), India*

*gourimpatil@gmail.com*

**ABSTRACT - India is the world's largest producer of food grains. There are several different types of grains like rice, wheat, sorghum, barley, millets, etc. found in India. It is critical for producers, manufacturers, and consumers to understand the variety of agricultural products while marketing them. Producers need to know the variety of the grain they plant to undertake proper farming. At the same time, marketers seek to ensure that the product variety they sell meets the needs of their target markets. Therefore the use of computer vision technology has vital importance in the agriculture sector to classify and identify these products based on their varieties and quality. This paper highlights the recent use of computer vision technology in the classification of grain as per different varieties. The paper presents a review of various image processing techniques including machine learning and deep learning classifiers for food grains varietal classification and quality prediction.**

**Keywords - varietal classification, preprocessing, segmentation, feature extraction, image classification**

## I.    INTRODUCTION

India is a leading producer of food grains.Most species of grain are grown in various varieties, each with unique characteristics such as physical characteristics, amount of nutrients found, agronomic qualities, and processing quality attributes. There is frequently a price difference based on different varieties of grain.Grains, such as rice, wheat, millets, sorghum, corn, etc. make up the majority of the human diet because they're cheap sources of carbohydrates and protein. They're also a wonderful base for many kinds of grain production. The grain seeds of excellent quality and pure variety play an important role in increasing production. Low purity seeds will cost farmers and consumers a lot of money in terms of breeding, planting, and commodity quality. But nowadays there are many different types of food grainsavailable in the market. That's why it is

very important to classify them according to their different varieties.

Currently, the varieties of food grains and their qualityare determined by human experts through visual inspection. The human-inspector decision-making abilities are influenced by external factors. Also, there may be chances of errors because of human subjectivity and requires more time. The limitation of this conventional method of quality analysis and variety identification can be overcome by using various image processing techniques.

Image processing and computer vision have become the most promising technology for identification, classification, and quality analysis for various food products.Research has been carried out forthe past few years for automatic classification and identification systems of grain-based on quality. This paper presents a review of recentdevelopments of

image processing applications in the field of agriculture. This review highlights different steps of image processing such as image acquisition, preprocessing, segmentation, feature extraction, feature selection, and classification for varietal classification of grains.

## II. LITERATURE REVIEW

Ksh Robert Singh, Saurabh Chaudhary [1]proposed a system for the classification of rice grain based on morphology, color, texture, and wavelet features. This work included classification ofa single rice kernel using a cascaded network classifier. They got 97.75% accuracy, using morphological features compared to other features. Also for the three best-selected features, the proposed cascade classifier performed well with 96.75 % classification accuracy. It is observed that theresearcher's proposed classifier outperforms than other mentioned classifier on a standard dataset from the University of California, Irvin (UCI).

BirenArora, NimishaBhagat[2] classified four types of rice including boiled rice. Area, perimeter, major axis, minor axis, eccentricity features extracted from the images and classified by using machine learning algorithms like logistic regression decision tree classifier, random forest classifier, and SVM. They obtainedthe most promising accuracy by the logistic regression algorithm on the given dataset.

W. Wu et al.[3] were proposed a model for the detection and enumeration of wheat grain. The wheat grain image dataset was acquired for this study under different scenarios and scales. The wheat grain detection and enumeration model was trained using a faster R-CNN deep learning network with a transfer learning strategy. This model had a grain counting error rate of less than 3% and a running time of less than 2%.

J. Panigrahi, P. Pattnaik, B. B. Dash, and S. Ranjan Dash [4] were used the HOG texture extraction method, shadow-based method, external high and refraction-based method, and CNN for the classification of rice as rice or non-rice to predict rice quality. The features extracted by HOG were classified by KNN, SVM, and Linear SVM and they found that SVM helped to find optimum boundaries in between the outputs. H. Vu, V. N. Duong, and V.V. Vu [5] were utilized the motif discovery to extract the shape pattern of each rice species. The counter of each

rice seed image was extracted and normalized. These exploiting results were used to construct a similarity map of ninety rice species. In this work, fine-grained challenges were handled when distinguishing highly similar species varieties of rice.

K. Robert Singh and S. Chaudhury[6] were presented the local square pattern (LSP) Local Vertical and Horizontal pattern, Local block pattern, and average local binary pattern feature extraction scheme for the classification of eight varieties of rice grain. The features extracted by using these techniques were classified by KNN, LDA, NB, and BPNN. They obtained 99.75 % average accuracy using BPNN for given feature sets.

. K. Aukkapinyo, S. Sawangwong, P. Pooyoi, and W. Kusakunniran[7] were trained the Region-based convolution neural network for the localization and classification of Rice grain in an image. The watershed algorithm, Auto alignment, and contrast limited adaptive histogram equalization has been used for preprocessing. A single rice kernel was considered for this study.

Adib Ali & Iris Co-researchers [8] classified six different varieties of corn seeds by combining histogram, texture, and spectral features. To optimize features, a correlation-based feature selection technique with the Best First search algorithm was applied. Also, for the classification Bayes Net, Multilayer perception, Random Forest (RF), Logit Boost, and classifiers were employed. They obtained 98.93 % accuracy using the MLP classification algorithm.

Gulzar et al.[9] proposed a model that classifies 14 commonly known seeds by employing Convolutional Neural networks and transfer learning. The dataset used in this study includes 234 symmetrical images. Model checkpointing, decaying learning rate, and hybrid weight modification are some of the approaches used in this study.They obtained a 99% result.

AnisSufiyaHamzah and Azlinah Mohamed [10] provided a review on classification methods in determining quality white rice grain using an artificial neural network approach. The finding of this study shows that ANN using BPNN has the highest accuracy of 96 %.

Singh et al.[11] proposed anew adaptive technique using a fuzzy clustered random forest (FCRF)andimage processing techniques. The

morphological features such as area, perimeter, height, width, length and asymmetry coefficientwere extracted for the classification. To classify the wheat seeds based on the extracted features the FCRF model wasused. The researchers obtained 97.7% average performance gain of the proposed technique.

Ramirez-Paredes and Hernandez-Belmonte [12] combined the Local Phase Quantization texture descriptor with color histogram features and shape descriptors for the classification of malting barley grains images. The SVM-RBF classifier was employed for the classification. They obtained 99% average precision accuracy for the proposed model as compared to the improved Median Ternary Pattern and Median Robust Extended Local Binary Pattern techniques.

M. J. Asif, T. Shahbaz, S. Tahir Hussain Rizvi, and S. Iqbal [14] were combined the approach of canny edge detection technique and principal component analysis for rice grain identification and quality analysis. For rice quality analysis, many morphological features such as area, size, perimeter, length, and so on are used. In terms of classification and quality analysis, the results they obtained were 92.3% and 89.5%, respectively.

NadeeshaNagoda and LochandakaRanathung[15] provided an approach for the segmentation and classification of rice samples using SVM. The watershed algorithm was used for the segmentation of touching and overlapping rice. Kernels Linear Binary pattern histogram, texture, and spectral feature, and color features were used to predict rice samples using a linear kernel-based support vector machine. They obtained 96% and 88% accuracy for segmentation and classification respectively

Kaur et al. [16] proposed a machine vision system for the classification of turmeric rhizomes into three different varieties using ANN. The histogram thresholding is used for segmentation. The seventeen morphological features were extracted from the images. The ANN model is developed with 17 input layers, 5 hidden layers, and 3 output layers for classification. They obtained 73.33% accuracy. Priyanka Gupta, Mahesh Bundle [17]presents a novel approach for the classification of seven different varieties of Basmati rice. ROI was used for segmentation after that six morphological and three color features were extracted for classification. They

obtained 95-100% accuracy to classify India Gate Basmati rice.

After doing a literature review the image classification techniquesused by different authors are presented in Table 1.

| Ref. | Food Type | Features | Classifier | Accuracy |
|------|-----------|----------|------------|----------|
| [1] | Rice | Morphological, Color, Texture, and Wavelet | BPNN, Fuzzy, ANFIS, KNN, LDA, NB, Cascade Network | 97.75% for Morphological features 96.75% for best-selected features |
| [2] | Rice | Morphological | LR, LDA, KNN, CART, NB, SVM, RF | LR 96% |
| [3] | Wheat | -- | Faster RCNN | 91% |
| [4] | Rice | HOG + Convolution Pooling | CNN, KNN | -- |
| [6] | Rice | Texture ( LSP, LBP ) | BPNN, KNN, LDA, NB, | BPNN 99.63 % |
| [7] | Rice | -- | Mask R-CNN | 80 % |
| [8] | Corn | Texture Spectral | RF, BN, LB, MLP | MLP 98.93% |
| [10] | White rice | Morphological Texture, Color | BPNN, SVM,KNN | BPNN 96% |
| [11] | Wheat | Morphological | FCRF | 97.7% |
| [12] | Barley | Shape Texture, Color | SVM | 99% |
| [13] | Rice | Morphological | PCA | 92.3% |
| [14] | Rice | Color, Texture | SVM | 96% segmentation 88% classification |

Table-1Image classification techniques used in researches

### III. CONCLUSION

By the review it is observed that the varietal classification of grain has been done with the steps of image acquisition, image preprocessing, feature extraction and image classification. It is also possible to classify turmeric by using the same. Also it is noted

that most of the grain classification work has been carried out by capturing the images in the controlled environment hence there still a scope to work on real life scenario images. Most of the researchers got higher accuracy by extracting the morphological features where as in some work the color and texture features play an important role in grain classification. In addition to varietal classification some work focuses on quality analysis of grain.

The different machine learning classifiers like SVM, KNN, NB, and Random forest are employed for the classification of rice, wheat and seeds. Also CNN is the most popular for the classification of different types of grain but it has varying architecture as per application and author. Hence a robust architecture is needed for the classification of grain as per the different varieties.

## REFERENCES

5. K. Robert Singh and S. Chaudhury, "A cascade network for the classification of rice grain based on single rice kernel," Complex Intell. Syst., vol. 6, no. 2, pp. 321–334, Jul. 2020, doi: 10.1007/s40747-020-00132-9.

6. B. Arora, N. Bhagat, L. R. Saritha, and S. Arcot, "Rice Grain Classification using Image Processing & Machine Learning Techniques," in 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, Feb. 2020, pp. 205–208. doi: 10.1109/ICICT48043.2020.9112418.

7. W. Wu et al., "Detection and enumeration of wheat grains based on a deep learning method under various scenarios and scales," J. Integr. Agric., vol. 19, no. 8, pp. 1998–2008, Aug. 2020, doi: 10.1016/S2095-3119(19)62803-0.

8. J. Panigrahi, P. Pattnaik, B. B. Dash, and S. Ranjan Dash, "Rice Quality Prediction using Computer Vision," in 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, Mar. 2020, pp. 1–5. doi: 10.1109/ICCSEA49143.2020.9132967.

9. H. Vu, V. N. Duong, and V.-V. Vu, "Exploiting the fine-grained similarity of a large-scale rice species using shape motif discovery," in 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh, Vietnam, Oct. 2020, pp. 1–6. doi: 10.1109/RIVF48685.2020.9140791.

10. K. Robert Singh and S. Chaudhury, "Comparative analysis of texture feature extraction techniques for rice grain classification," IET Image Process., vol. 14, no. 11, pp. 2532–2540, Sep. 2020, doi: 10.1049/iet-ipr.2019.1055.

11. K. Aukkapinyo, S. Sawangwong, P. Pooyoi, and W. Kusakunniran, "Localization and Classification of Rice-grain Images Using Region Proposals-based Convolutional Neural Network," Int. J. Autom. Comput., vol. 17, no. 2, pp. 233–246, Apr. 2020, doi: 10.1007/s11633-019-1207-6.

12. A. Ali et al., "Machine learning approach for the classification of corn seed using hybrid features," Int. J. Food Prop., vol. 23, no. 1, pp. 1110–1124, Jan. 2020, doi: 10.1080/10942912.2020.1778724.

13. Y. Gulzar, Y. Hamid, A. B. Soomro, A. A. Alwan, and L. Journaux, "A Convolution Neural Network-Based Seed Classification System," Symmetry, vol. 12, no. 12, p. 2018, Dec. 2020, doi: 10.3390/sym12122018.

14. A. S. Hamzah and A. Mohamed, "Classification of white rice grain quality using ANN: a review," IAES Int. J. Artif. Intell. IJ-AI, vol. 9, no. 4, p. 600, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp600-608.

15. P. Singh, A. Nayyar, S. Singh, and A. Kaur, "Classification of wheat seeds using image processing and fuzzy clustered random forest," Int. J. Agricultural Resources, Governance and Ecology, Vol. 16, No. 2, 2020

16. J.-P. Ramirez-Paredes and U.-H. Hernandez-Belmonte, "Visual quality assessment of malting barley using color, shape and texture descriptors," Comput. Electron. Agric., vol. 168, p. 105110, Jan. 2020, doi: 10.1016/j.compag.2019.105110.

17. V. Truong Hoang, D. P. Van Hoai, T. Surinwarangkoon, H.-T. Duong, and K. Meethongjan, "A comparative study of rice variety classification based on deep learning and hand-crafted features," ECTI-CIT, vol. 14, no. 1, pp. 1–10, Mar. 2020, doi: 10.37936/ecti-cit.2020141.204170.

18. M. J. Asif, "Rice Grain Identification and Quality Analysis using Image Processing based on Principal Component Analysis," p. 6, 2018.

19. N. Nagoda, "Rice Sample Segmentation and Classification Using Image Processing and Support Vector Machine," Th Int. Conf. Ind. Inf. Syst., no. 978, p. 6, 2018.

20. A. Kaur, N. Saini, R. Kaur, and A. Das, "Automatic classification of turmeric rhizomes using the external morphological characteristics," p. 4, 2016.

21. P. Gupta and M. Bundele, "A New Approach for Rice Quality Analysis and Classification Using Thresholding-Based Classifier," p. 11, 2016.

Paper ID: CS-07

# Game Theory – A Way to Network Security

**Miss.Tejal Kashinath Warake**

*Department of computer science*
*BAS P.O.Nahata college, Bhusawal*
*Email -tejalwarake@gmail.com*

*Abstract - Network security is a complex and challenging the problem. The area of network protection mechanism design is receiving vast attention from the research community for more than two years. However, the network security problem is distant from completely solved. This paper states the current game theoretic solutions which are designed to augment of network security and presents a catalog for classifying the proposed solutions. Thecatalog should provide the reader with a better understanding for game of theoretic solutions to a variety of cyber security problems.*

*Keywords - ONR, IDS, Payoff.*

## I. INTRODUCTION

Current affairs in Internet prove that network attacks result into massive amounts of loss to government, private enterprises, and the general public in terms of money, data confidentiality, and reputation. The research community has been paying consideration to the network security problem for more than two decades. The trouble is far from being fully solved. The key factor which makes this problem hard is that the native network, which wishes to be secured, is characteristically connected to the Internet and most important parts of the Internet are out of the control of network officers. However, the Internet has become an essential component of running the daily business of government, economic institutions, and the general public. As this work is maintained by the Office of Naval Research (ONR) under grant a result, there is a pressing need to design security measures for network attacks.

Usually, web safety solutions service one or the other defensive devices such as firewalls or sensitive devices such as Intrusion Detection Systems (IDSs) and together are used in combination. The interruption detection algorithms are either built on detecting an attack mark or spotting the abnormal performance of the system. When an attack is noticed the employed IDS alerts the network officer who then takes an action to break or diminish the attack. But, now IDSs are not very classy and they trust on ad-hoc arrangements and trial work. The present IDS knowledge may prove enough for caring against unplanned attackers using well known skills, but here is quiet a requirement to policy tools to secure beside refined and organized adversaries. The weakness of the traditional network safety solutions is that they absence a computable decision background. A unusual groups of investigators have started supporting the use of game theoretic methods. As game theory arrangements with problem wherever numerous players by opposing aims participate with each other, it can offer us with a accurate frame for investigation and displaying network security problems. Furthermore, game theory earnings the capability of observing thousands of imaginable situations previously taking the greatest action; later, this one can trendsetter the choice procedure of the network officer to a big scope. Thus, numerous game theoretic methods have proposed to address network security topic. This paper analyses the surviving game theoretic solutions which are intended to improve network security and grants a taxonomy for categorizing them. Underlining the basic game type used in the security mechanisms, though selecting

complete differences, this classification affords the reader with a global view of the problem and solution space. This paper does not promote some specific defense game.

## II. AN OVERVIEW OF GAME THEORY

This section identifies the premise of theory of games to assist the understanding of the games. For an in depth introduction to theory of games refer A Course in theory of games. Theory of Games describes multi-person decision scenarios as games where each player chooses actions which end within the simplest possible rewards for self, while anticipating the rational actions from other players. A player is that the basic entity of a game who makes decisions then performs actions. A game could also be a particular description of the strategic interaction that has the constraints of, and payoffs for, actions that the players can take, but says nothing about what actions they actually take. A key concept is a systematic description of how the game will be played by employing the best possible strategies and what the outcomes might be. A preference relation may be a complete relation on the set of consequences which model the preference of every player within the game. A strategy for a player may be a complete plan of actions altogether possible situations throughout the sport. If the strategy specifies to require a singular action during a situation then it's called a pure strategy. If the plan specifies a probability distribution for all possible actions during a situation then the strategy is mentioned as a mixed strategy. A Nash equilibrium is defines a steady state condition of the game; no player would prefer to change his strategy as that would lower his payoffs given that all other players are adhering to the prescribed strategy. This key concept only specifies the steady state but does not specify how that steady state is reached in the game. This information are going to be wont to define games that have relevant features for representing network security problems.

### a) Definitions

### Game

A description of the strategic interaction between opposing, or co-operating, interests where the constraints and payoff for actions are taken into consideration.

### Player

A crucial entity in a game that is tasked with making choices for actions. A player can signify a person, machine, or group of persons within a game.

### Action

An act constitutes a move in the specified game.

### Payoff

The positive or negative incentive to a player for a given action within the game.

### Strategy

Sketch of action within the game that a given player can take during game play.

### Perfect Information Game

A diversion in which each player is alert of the moves of all other players that have already taken place. Examples of perfect information games are: chess, tic-tac-toe, and go. A diversion where at least one player is not aware of the moves of at least one other player that have taken place is called a defective information game.

### Complete Information Game

A game in which every player knows both the strategies and payoffs of all group of actors in the game, but not essentially the actions. This term is often puzzled with that of perfect information games but is distinct in the fact that it does not take into account the actions each player have already taken. Unfinished information games are those in which at least one player is innocent of the possible strategies and payoffs for at least one of the other players.

### Bayesian Game

The game in which information regarding the strategies and payoff for other players is partial and a player assigns a 'type' to other players at the onset of the game. Such games are labelled Bayesian games owed to the use of Bayesian analysis in predicting the outcome.

### Static/Strategic Game

A one-shot game in which each player chooses his map of action and all players' decisions are made concurrently. This means when choosing a plan of action each player is not up to date of the plan of action chosen by any other player. In this paper, this group of game is referred to as 'static game'.

### Dynamic/Extensive Game

A game with more than one stages in each of which the players can consider their action [32]. It can be considered as a chronological structure of the decision making problems encountered by the players in a static game. The sequences of the game can be also finite, or infinite. In this paper, this group of game is referred to as 'dynamic game'.

**Stochastic Game**

A game that involves probabilistic transitions from side to side several states of the system. A game progresses as a sequence of states. A game begins with a start state; the players choose events and receives a payoff that depend on the current state of the game, and then the game transitions into a new state with a likelihood based upon players' actions and the current state.

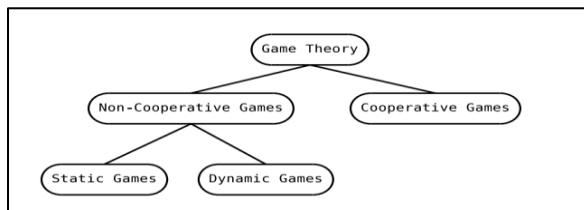## III. TAXONOMY: CLASSIFICATION OF GAME



Figure 1. Classification of games

Figure1 demonstrates the essential classifications of theory of games. An Existing game-theoretic research as applied to network security falls under non-cooperative games. As such, this paper doesn't further expand upon 'cooperative games'. Section 3.1 discusses existing works involving static games while Section 3.2 deals with current works involving dynamic games. Section Finally, Section 3.3 presents some directions for future research. According to the completeness of knowledge, static games are often classified into two sub-classes as listed below. We briefly discuss the prevailing research works which fall under each sub-class of static games.

### a) Complete imperfect information

Jormokka et al. introduced a few of samples of static games with complete information where each example represents an IW scenario. For each scenario the authors found the only strategy of the players during a quantitative form. Especially, they investigated if quite one Nash equilibria exist and if so, then which one is presumably to look because the result given the players' strategies. These examples show that relying on the scenario the players could get the advantage of a bold strategy or a mixed strategy.

### b) Incomplete imperfect information

Liu et al. presented steps to model the interactions between a DDoS attacker and therefore the network administrator. This advance experimental that the power to model and infer attacker intent, objectives, and methods (AIOS) is vital because it can cause effective risk assessment and harm prediction. An incentive-based game-theoretic model to infer AIOS was discussed during this work. a couple of bandwidth parameters were used because the metric to live the impact of the attack and therefore the countermeasure, which successively measures the attacker's, and defender's, incentive. The work also observed that the simplest game model to settle on depends on the degree of accuracy of the employed IDS and therefore the degree of correlation among the attack steps. The simulation results including game plays following the Bayesian model while the simulation experiment was performed on ns-2. A topology measured within the simulation experiment consists of 64 source hosts connected to at least one victim machine via 4 levels of routers. Each router is capable of employing the pushback mechanism as a part of the defence strategy. Group of Nash equilibrium strategies were computed via the simulation.

### c) Complete perfect information

Lye et al. proposed a game model for the safety of a network. During this work, an enterprise network was envisioned as a graph of 4 nodes (web server, digital computer, work station and external world) alongside the traffic state for all the links. A two-player (administrator, attacker), stochastic, general-sum game and therefore the authors focused on 3 attack scenarios namely, defaced website, denial-of-service, and stealing confidential data. The sport was described from the purpose of view of both players. a proper model defined the sport as a 7-tuple— the set of network states, the action set for every player, the state transition function, the reward function and a reduction factor. Especially, this work considered a stochastic game involving 18 network states and three actions for every player at each state.

Xiaolin et al. Planned a Markov theory of games based model for risk assessment of network data system considering the safety status of both present and future. They identified that threats working on vulnerability can induce risk and therefore the risk are going to be larger and bigger by threat spreading. On the opposite hand, the danger are going to be smaller and smaller by the system administrator's repairing the vulnerability.

Essentially, the experiment involves a game of complete and excellent information with two players. Authors formulated a function to capture the damage and used it to assess the danger. Using the damage function the supervisor would choose the repair strategy which minimizes the utmost damage. To estimate the model experts constructed a risk assessment platform with four subsystems which are malicious code Detection Subsystem, Vulnerability Detection Subsystem, Asset Detection Subsystem and Risk Assessment Subsystem. They used Trojan.Mybot6307 as a threat, and three assets to describe states. Their results are similar or better than the normal assessment model like Fault Tree Analysis (FTA) because they effectively incorporated the potential risk also. They claimed that the model also results in the simplest system repair scheme.

Alpcan et al.modeled the interaction between malicious attackers to a system and therefore the IDS employing a stochastic (Markov) game. They captured the operation of the IDS sensor system employing a finite-state Markov chain, and thought of three different information structures:

(a) The players have full information about the sensor system characteristics and therefore the opponents;

(b) The enemy has no information about the sensor system characteristics;

(c) Each player has only information about his own costs, past actions, and past states. a couple of illustrative examples and numerical analysis were presented for these three cases. Tools like value iterations to unravel Markov decision processes (MDP), minimax-Q, and naive Q-learning were wont to find the simplest strategies of the players.

### d) Complete imperfect information

Nguyen et al. as seen the network security problem as a sequence of nonzero-sum games played by an attacker and a defender. The game model, called 'fictitious play (FP)', conservatively considers that the players cannot make perfect observations of every other's previous actions. This effort studied the impact of the error probabilities related to the sensor system on the Nash equilibrium strategies of the players considering two scenarios—

(a) Each player is conscious of these error probabilities;

(b) Neither player knows these error probabilities. Both classical and stochastic FP games are investigated via simulation.

### e) Incomplete perfect information

Chen in his doctoral dissertation used game theoretic model to style the response for the importance-scanning Internet worm attack. The most idea is that defenders can choose the way to deploy an application, that's the group distribution, when it's introduced to Internet to attenuate the worm propagation speed. The attacker can choose the optimal group scanning distribution to maximise the infection speed. Thus a game would be played between the attacker and therefore the defender. The attacker should choose so on maximize the minimum speed of worm propagation, while defender wants to attenuate the utmost speed of worm propagation. By framing the matter this manner it seems to be a zero sum game and a min-max problem. The optimal solution for this problem is that defender should deploy the appliance uniformly within the entire IP-address space or in each enterprise network, in order that the simplest strategy that the attacker exploits is like random scanning strategy.

### f) Incomplete imperfect information

Alpcan et al. modeled the communication of an enemy and therefore the network manager as a repeated game with 'finite steps' or 'infinite steps'. This work understood that the sensor system which is deployed to detect the attacks is imperfect and thought of the sensor system as a 3rd 'fictitious' player almost like the 'nature' player in standard theory of games. You et al. described the means to model the complex safety scenario considering the interaction between the hacker and therefore the defender as a two player, zero sum game. It gave a taxonomy of relevant theory of games and network security terms and suggested a correlation between them. They known at the utility of Nash and Bayesian Equilibria in representing the concepts to predict behaviour and analysed the interaction between the attacker and therefore the defender. They gave an inventory of theory of games terms that are relevant within the network security scenario and explained them. They concluded by suggesting that to unravel this problem linear algorithms would be appropriate.

**g) Scope of future research**

Many of the present game-theoretic security approaches are supported either static game models or games with perfect information or games with complete information. However, actually a network administrator often faces a dynamic game with incomplete and imperfect information against the attacker. A number of the present models involving dynamic game with incomplete and imperfect information are specific to wireless networks while a couple of others don't consider a sensible attack scenario.

Some of the restrictions of this research are:

(a) Current stochastic game models only consider perfect information and assume that the defender is usually ready to detect attacks;

(b) Current stochastic game models assume that the state transition probabilities are fixed before the sport starts and these probabilities are often computed from the domain knowledge and past statistics;

(c) Current game models assume that the players' actions are synchronous, which isn't always realistic;

(d) Most models aren't scalable with the dimensions and complexity of the system into account.

## IV. CONCLUSION

Hacker's activities have significantly increased in cyber space, and are causing damage by exploiting weaknesses in information infrastructure. Theory of games offers promising perspectives, insights, and models to deal with the ever changing security threats in cyber space. This survey highlights important game theoretic approaches and their applications to network security and descriptions possible directions for future research. It's to be noted that classes within the taxonomy might be divided into more detailed levels. It's obvious that new classes may have to be introduced within the taxonomy after new defence mechanisms are proposed within the future.

## REFERENCES

1. T. Alpcan and T. Baser. A game theoretic analysis of intrusion detection in access control systems. Proc. of the 43rd IEEE Conference on Decision and Control, 2004.
2. X. You and Z. Shiyong. A kind of network security behavior model based on game theory. Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies, 2003.
3. K. Lye and J. Wing. Game strategies in network security. Proceedings of the Foundations of Computer Security, 2002.
4. The Trusted Computing Group, http://www.trustedcomputinggroup.org.
5. S. N. Hamilton, W. L. Miller, A. Ott, and O. S. Saydjari. The role of game theory in information warfare. Proceedings of the 4th information survivability workshop (ISW2001/2002), 2002.
6. Z. Chen. Modeling and defending against internet worm attacks. PhD Dissertation at Georgia Institute of Technology, 2007.
7. K. C. Nguyen, T. Alpcan, and T. Basar. Stochastic games for security in networks with interdependent nodes. Proc. of Intl. Conf. on Game Theory for Networks (GameNets), 2009.
8. M. Bloem, T. Alpcan, and T. Basar. Intrusion response as a resource allocation problem. IEEE Conference on Descision and Control, 2006.
9. The National Strategy for Homeland Security, http://www.dhs.gov/interweb/assetlibrary/nat-strat-hls.pdf, 2002.
10. Liang, Xiannuan& Xiao, Yang. (2013). Game Theory for Network Security. Communications Surveys & Tutorials, IEEE. 15. 472-486. 10.1109/SURV.2012.062612.00056.
11. Luo, Yi &Szidarovszky, Ferenc& Al-Nashif, Youssif& Hariri, Salim. (2010). Game Theory Based Network Security. J. Information Security. 1. 41-44. 10.4236/jis.2010.11005.

**Paper ID: CS-08**

# Offline Handwritten MODI Characters Recognition

**Miss. Dipali Murlidhar Nemade**

*Department of computer science*
*BAS P.O.Nahatacollege, Bhusawal*
*Email- dipalinemade106@gmail.com*

**ABSTRACT -** *India is a country where we discover large diversity in traditions, religions and languages. India speaks 780 languages out of which 220 languages have been vanished in last 50 years and another 150 could become extinct in next semi century. The language is a standard of communication between two individuals and it has two forms that are oral and written. Every language has its own temperament set, representation structure and rules, but aim was same and that is 'Communication'. In history, the medium of communication is one of the indication to show the progress of a culture. In this paper the workwas intense on 'MODI' Script. As the contribution data set is contains documents written in MODI script, and since MODI script characters are written in cursive mode, obtain entity characters from the text becomes particularly complicated. Also, there is no disconnection between two nearby words. Although now-a-days, Marathi language uses Devnagari script, quite a few years back MODI Script was used for writing documents in Marathi language*

*Keywords-MODIScript,Hand-writtencharacterrecognitionMODIscript,HistoryofMODIScript, DevnagariAndMODIScript*

## I. INTRODUCTION

MODI script temperament segmentation is one of the most demanding phase, as there is hardly a few disconnection among the contiguous typeset written in these documents. Complete line of these documents is written exclusive of lifting one's hand.

## II. HISTORY OF MODI SCRIPT

MODI script, was use for lettering reason only, which was a cursive form of writing in 'Marathi' (most important language ofMaharashtra state in western India) there is a number of theories in relation to the source of this script. One of them claims with the aim of in 12th Century MODI was developed by 'Hemadpant' or 'Hemadri'. Dr.Rajwade and Dr.Bhandarkar believe to facilitate Hemandpant bring MODI script from Sri Lanka, other than according to Chandorkar, MODI script has evolve on or after Mouryi (Bramhi) script of Ashoka stage. Oldest accessible MODI document is alleged to be of 1429 A.D. according an additional memo Oldest MODI document is alleged to be of 1389 A.D

(Skaka1311) potted in the museum Bharat ItihasSanshodhanMandal (BISM) Pune. It is a well-liked conception that only "Marathi is written in MODI". The chronological evidence, says that The MODI alphabet was imaginary through the 17th century to write down the 'MARATHI' language of Maharashtra and it was commonly and generally used for only writing principle all over the Maharashtra in the period of'Peshwe' (Pune) and 'ChatrapatiShivajiMaharaj'.There are a lot of changes has been complete in writing styles of MODI, most basic in 12th century.

MODI script was famous as 'proto-MODI' or 'Adyakalin', MODI emerge as a discrete script through 13th century known as 'Yadavakalin'. The after that stage of growth is the 'Bahamanikalin' of the 14th–16th century, follow by the 'Shivakalin' of the 17th century. The recognized 'Chitnisi' was residential during 18th century, a range of MODI styles begin to propagate. This period is known as 'Peshvekalin', which last until 1818. The dissimilar style of MODI used from side to side this stage is also

known as 'Chitnisi', 'Bilavalkari', 'Mahadevapanti', and 'Ranadi'. The end period of MODI isconnected with English regulation and is called 'Anglakalin'. This structure of script was second-hand from 1818 until 1952. Most famous chronological forms are Bahamanikalin, Chitnisi, Peshvekalin, and Anglakalin. MODI was new in the primary school books formed throughout the 19th and 20th centuriesas in Figure 1.



Figure 1. Historical forms of MODI script writing

## III. CHARACTER SET OF MODI SCRIPT

The MODI was printed by 'Boru', 'lekhan' (a pen, formed from 'Bambuu', want to elevate too often for reducing in the ink.). The MODI script includes 46 individual letters, of which 36 be consonants in addition to 10 vowels. These typescript are known as basic characters. As evaluate to Devanagari here are 48 individual letters, where 36 consonants and 12 vowels. MODI script do not have lengthy 'i' and long's'. Earlier than the origination of characters in MODI, straight line was pinched diagonally the page. Typeset were in black and white with respect to the straight line. The letters themselves are filled of moulds or curve (cursive script). This is appropriate to the pressure of 'Devanagari' writing and need not to lift'Boru' too often. No punctuation symbols or conjuncts are second-hand in this script. It does not contain any characters to point toward the execution of a condemnation and the vertical stroke is not used. The necessary typescript of MODI are shown in Figure 2.

Range: 11600—1165F    Quantity of characters: 96



Figure 2:-Character Set Of MODI script

## IV. COMPARISION BETWEEN DEVANAGARI AND MODI SCRIPT

A number of point of discrimination between these scripts are given as:

1] A few letters of MODI script are just similar to as devnagari script have only modest spot changes. These are as follows.



ga, gha, bha, sha, cha, ala show extremely fewer disparity. ta have curled to the left. Na have physical note to the right. sha also has bowed shape to the left. Da has all part in curled shape.

2] a number of an additional collection of letters have following differences:



Here, if 'ha' have known greater partly half circle towards exact then it became 'e'. If 'ha' have agreed diminutive circle solid or concave then it became 'dha'. If 'ha' have given greater partly half circle towards left then it became 'kha'.

In the next row it has been originate that if upright bar bowed towards left then it became 'ra'. If the same erect bar bowed to semi cricle then it became 'ta'. If this semi-circle contain point '.' then it became 'tha'. In this approach if 'ja's top left circle become solid then it became 'na'.

## V. STRUCTURAL FACIAL APPEARANCE OF MODI SCRIPT CHARACTERS

The conservative handwritten characters gratitude system is based on the quality taking out from the character figure such as a line preference, comparative position of such line, the thick of the different parts from the line. These quality features can be used for non-cursive handwritten characters or written font. However, these facial appearance are not an adequate amount of the cursive character gratitude. Structural facial appearance are based on the Geometrical and topological property of a character. A character is collected of number of mechanism in the form of stroke. These stroke may well be line, arcs, curves, loops etc. every one character constituent is called as a character primitive. These are extract either from a draft or a curve of a character image. The character is standard on the basis of amount of such primitives, the variety of primitives and the relation among these primitives in some order. Some frequently used topological and geometrical facial appearance in character recognition are end points .The variation in handwritten version of a character appropriate to different writers are so high that this demonstration gives a lot of difference in compute features. In some studies, structural facial appearance have been used to carry out preface categorization where intact alphabet set is separated into groups or subgroups based on topological facial appearance.

## VI. CONCLUSION

A batch of occupation has been done for the expansion of personality acknowledgment system for the a selection of languages of the globe but good ICR of greater part languages are at a standstill not available. Amongst the range of Indian scripts, Like Devnagari, MODI is one of the script. The main motive at the rear minor work done for MODI hand-written nature gratitude may be the intricacy of the script, oldest language and non-availability of a MODI hand-written character database. The gratitude of hand-written script of a language is additional complex as compare to hand-printed or machine-printed. As well, the gratitude of MODI hand-written script is tricky as compare to many other language hand-written script due to the cursive natural history of most of its class members, complex structure of some of its class members, large within class variation due to unusual writing behavior of different persons, and identical structure of some of its component classes. Also it is very difficult to take out geometrical and topological facial appearance from MODI hand-written characters.

## REFERENCES

1. Kulkarni, Sadanand&Borde, Prashant&Manza, Ramesh &Yannawar, Pravin. (2014). Offline Handwritten MODI Character Recognition Using HU, Zernike Moments and Zoning.
2. Besekar, D. &Ramteke, Rakesh. (2013). Study for Theoretical Analysis of Handwritten MODI Script A Recognition Perspective. International Journal of Computer Applications. 64. 45-49. 10.5120/10618-5340.
3. Kulkarni, Sadanand&Borde, Prashant&Manza, Ramesh &Yannawar, Pravin. (2015). Review on Recent Advances in Automatic Handwritten MODI Script Recognition. International Journal of Computer Applications. 115. 5-12. 10.5120/20257-2636.
4. PankajA.patil(2016) Character Recognition System for Modi Script ISSN (e): 2250 – 3005 || Volume, 06 || Issue, 09
5. D.N.Besekar(2012) Special Approach for Recognition of Handwritten MODI Script's Vowels International Journal of ComputerApplications®(IJCA)
6. S. Joseph and J. George, "Handwritten Character Recognition of MODI Script using Convolutional Neural Network Based Feature Extraction Method and Support Vector Machine Classifier," 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), 2020, pp. 32-36, doi: 10.1109/ICSIP49896.2020.9339435.

Paper ID: CS-09

# Covid-19 Detection using Image Processing & Machine Learning with post-Covid infections– A Review

**Sweta G. Phegade**

*Asst. Professor,*
*Computer Department, KCES'S IMR, Jagaon*
*KBC, North Maharashtra University,*
*Jalgaon, India.*
[#1]*sweta384@gmail.com*,

**Charuta Khadke**

*Asst. Professors,*
*Computer Department, KCES'S IMR, Jagaon*
*KBC, North Maharashtra University,*
*Jalgaon, India.*
[*2]*khadke.charuta5@gmail.com*

**ABSTRACT - *The coronavirus disease (COVID-19) eruption has instigated many death cases and affected all sectors of human life. The early finding and diagnosis of COVID-19 at the lowest cost and in the early stages of the disease are among the main challenges in the current COVID-19 pandemic. Accordingly, image processing & machine learning methods are widely used for recognizing the functional disorder in the inner system using different types of image data. Medical and computer researchers incline to use machine-learning models to study medical images. The present systematic review was conducted by taking the consideration of some X-ray & CT-scan images. In this paper, we first focus on succinct the research works related to machine learning applications for COVID-19 medical image processing. According to the findings, machine learning-based models have an unusual capability to offer an exact and effective system for the detection and diagnosis of COVID-19. Use of this processing would lead to improve the understanding and specificity values. The application of machine learning in the field of COVID-19 medical image processing eases false-positive and negative errors in the detection and diagnosis of this disease and offers an irreplaceable chance to provide faster, cheaper, and safer diagnostic services to patients.***

*Keywords -Covid-19, ensemble technique, SVM, image processing, machine learning, x-ray, CT image.*

## I. INTRODUCTION

Covid-19 has upset societies and intensely altered existence across the world. This pandemic has led to dramatic loss of human life. This disease is caused by severe acute respiratory syndrome its most common symptoms are fever, dry cough and tiredness. But some patients are suffering with aches and pains, sore throat, diarrhea, conjunctivitis, headache etc. This disease is started in Wuhan City, China in December 2019. This virus come in some towns in Kerala, India in March 2020 [2]. Daily cases in mid-September with over 90,000 cases reported per-day, dropping to below 15,000 in January 2021 [1]. The increase in patients with COVID-19 is devastating in healthcare and economical systems

around the world. Due to the large number of people affected by this pandemic, the medical departments are facing problems of detecting this virus effectively and fastly. In India vaccination started at January 2021. In India there are three authorized vaccine one is the British Oxford–AstraZeneca vaccine (Covishield), the Indian BBV152 (Covaxin) vaccine, and the Russian Sputnik V vaccine for emergency use. The second wave of Covid-19 beginning in March 2021 was much larger than the first, with shortages of vaccines, hospital beds, oxygen cylinders and other medicines in parts of the country [1]. Diagnosing of covid-19 is somewhat easier by using medical imaging. Medical imaging is the process of making visible images of internal structures of the body for scientific, therapeutic study

and cure as well as a visible view of the function of inner tissues. Disorder identification and management is done by this process. Both organic and radiological imaging which used electromagnetic energies (X-rays and gamma), sonography, magnetic, scopes, and thermal and isotope imaging. The digital images play a necessary role on a daily basis. This processing includes many types of techniques and operations such as image gaining, storage, presentation, and communication. The image is a function that signifies a measure of characteristics such as illumination or color a viewed sight. The digital images have benefits such as quicker and inexpensive processing cost, easy storing and communication, instant quality assessment, multiple copying with reserving the quality, fast and cheap reproduction, and flexible manipulation. The drawbacks of digital images are exploitation copyright incapability to resize with preserving the quality, the need of large-capacity memory, and the need of faster processor for manipulation. An image processing technique is the usage of computer to operate the digital image. This technique has many profits such as elasticity, adaptability, data storing, and communication. X-rays have been used to image the body parts for investigative purposes. This paper was discussed and provide short review of different algorithmic method used to analyze the CT-Scan, X-Ray images of chest to check whether the patient is covid-19 positive. Because of RT-PCR's availability are less and not easily accessible in low financial level people then next best alternatives to check whether the patient is positive or not is Chest X-ray image and CT-Scan images. These images contain sufficient information that finds wide- spread applications in diverse disease diagnosis and decision making to assist the medical experts. Image processing and machine learning are emerging field for detection of Covid-19 in future. These fields use these medical image data and generate the results. In this paper, we consider nearly 30 research papers to study different algorithms are used to analyze different image samples. Also, Post-covid symptoms among patients. Most of the people with COVID-19 get better within weeks of illness, some of them experience post-COVID conditions. Symptoms of COVID conditions are experience more than four weeks after first being infected with the virus that causes COVID-19 [3]. Even some people who did not have symptoms when they were infected can have post-COVID conditions. These conditions can have different types

and groupings of health problems for changed lengths of time.

## II. IMAGE SAMPLES

Following Fig 1. Shows the Chest x-Ray images for various kinds of lungs infection. (a) Normal Lungs (b) The bacterial Pneumonia infection (c) Viral Pneumonia Infection & (d) Covid- 19 Infected Image
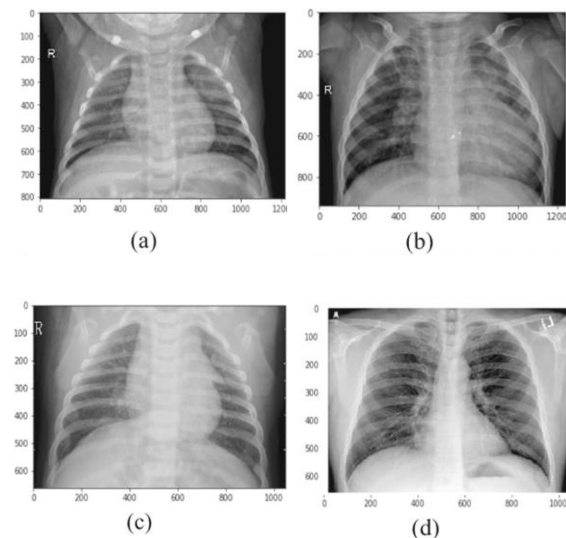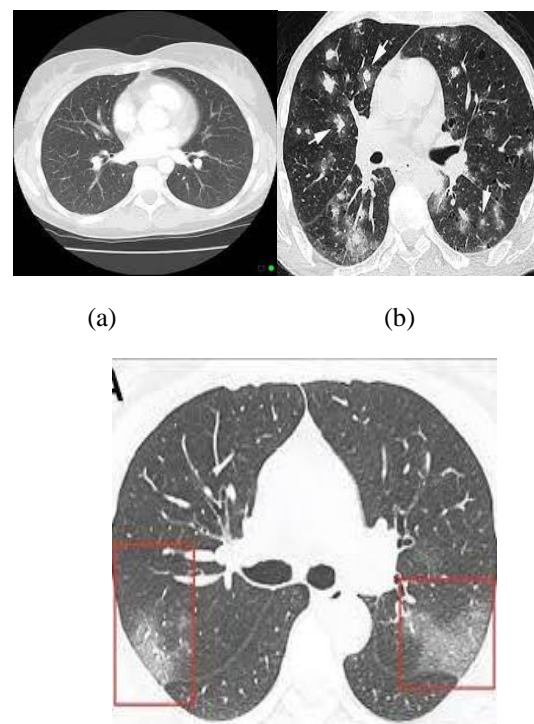


*Fig. 1: Different Lungs X-Ray Images*

Following Fig.2 shows the (Computed tomography)CT-Scan Image samples for various kinds of lungs infection. (a) Normal Lungs (b) Viral Infection Lungs (c) Covid-19 Infection



(a)                    (b)

(c)

*Fig.2: Different Lungs CT-Scan Images*

### III. MACHINE LEARNING

ML has high performance for several image processing applications such as image analysis, image classification, and image segmentation. Image classification achieved by extracting the import features from the images by a descriptor and then these features can be used in the classification task using classifiers such as SVM [5]. To detect COVID-19 infection from chest X-ray images deep learning based method is used. [16]. The classical machine learning method clinical report classified into four different classes. Different techniques like Term frequency or Inverse document frequency, bag of words and report length are used to perform feature engineering. Instead of other ML algorithms Logistic regression and Multinomial Naıve Bayes showed better results [31]. Following is some common Machine Learning algorithms used for detection of covid 19 using either CT images or x-ray images.

1. Logistic regression algorithm predicts the class of numerical variable based on its relationship with the label [32].

2. Multinomial Naıve Bayes algorithm computes class probabilities of a given text by using Bayes rule [33].

3. Support vector machine (SVM) is a supervised machine learning algorithm for classifying text into different categories [34].

4. Decision trees is an alternative approach for classification it partitions the input space into regions and classifies every region independently [35].

5. Bagging is an ensemble machine learning algorithm which improves the performance of other classification and regression algorithms [36].

a. AdaBoost algorithm works with those instances of the dataset, which are weighted [37].

b. Random forest classifier Ensemble machine learning algorithm used for classification and works like a decision tree. This algorithm uses a modified tree learning algorithm which selects and split each learning process by a subset of random features [38].

c. Stochastic gradient boosting this algorithm allows trees to be greedily created from samples of the training dataset [39].

Following Figure 3 shows general model of machine learning. This model is taking input of different x-ray images. Images can be captured and reduces the noise from the images and then perform processing on those images according to their size, numbers of pixel size. These processed images skilled under the training phase using different model validation techniques i.e split, cross validations, McNemar's Test etc. The resultant images identify and analyze using feature extraction, parallel implementation, feature selection etc. and generate result where the result is positive or negative [20].
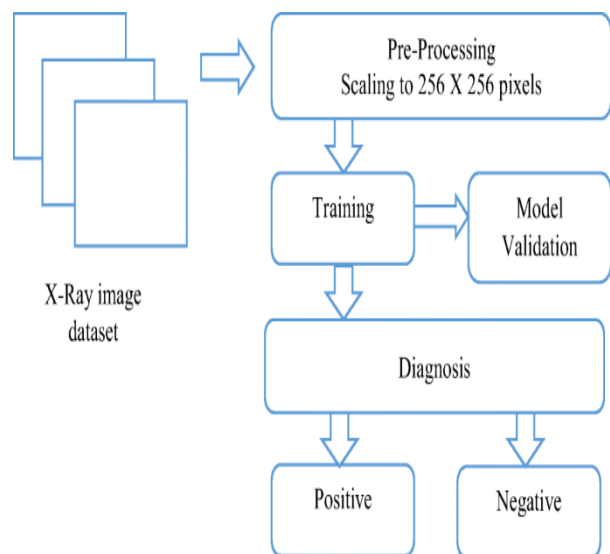


*Figure 3: General Model of Machine Learning [20]*

### IV. RESEARCH PAPERS REVIEW

| Sr. No. | Authors | Data Type | Method | Machine Learning/ Technique | Result | Sensitivity |
|---|---|---|---|---|---|---|
| 1 | Linda Wang et. al. [40] | CXR images | COVID-Net | Deep Neural Network architecture | 93.3 % | 91% |
| 2 | Sara Hosseinzadeh Kassani et. al. [41] | X-ray and CT images. | Desnse Net121 architecture trained by a Bagging tree classifier | Deep learning-based feature extraction | 99 % | |
| 3 | Tahereh Javaheri et. | CT images | CovidCTNet | CNN Model | 90% | 83% |

| No | Author | Image/Data | Method | Model | Accuracy | |
|----|--------|-----------|--------|-------|----------|---|
| | Al [42] | | | | | |
| 4 | Terry Gao [43] | X-ray medical lung images | supervised learning method | Deep CNN | 95% | |
| 5 | medRxiv preprint [44] | chest X-ray | nearest-neighbours algorithm | deep learning models | 90.3% | 90.5% |
| 6 | Aram Ter-Sarkisov [45] | CT scans | COVID-CT-Mask-Net | Torchvision's Mask R-CNN library | 91.66% | 90.80% |
| 7 | Aram Ter-Sarkisov [8] | Chest CT Scans | Detection and Segmentation of Lesion Areas in Chest CT Scans | Mask R-CNN models | 96.33% | 92.68% |
| 8 | Aram Ter-Sarkisov [9] | Chest CT Scans | lightweight Mask R-CNN model | ResNet18+FPN and ResNet34+FPN | 93.95% | 91.35% |
| 9 | AMIT KUMAR DAS et. al. [10] | X-Ray images | Proposed Model = three state-of-the-art deep learning models have been adopted and ensembled. | Deep Convolutional Neural Network | 95.7% | 98% |
| 10 | Padra Silva et. al. [11] | CT images | EfficientCovidNet | Deep Learning Models | 77.34% | 69.39% |
| 11 | Vijayakumar Bhandi et. al. [12] | X-ray Image | CBIR framework | Deep Convolutional Neural Networks | | |
| 12 | Ali Narin 1 [46] | digital x-ray images | Proposed edge detection technique - CEED-Canny | convolutional neural networks | | |
| 13 | Rachna Jain et. al. [14] | Chest x-ray &chest CT | Invent Proposed Model but not compared Comparison of Inception V3, Xception, and ResNeXt models | CNN models | 97.97% | |
| 14 | Younes Akbari et. al. [15] | CT scans images | | | | |
| 15 | Asif Iqbal Khan et. al. [16] | chest X-ray images | CoroNet | Deep Convolutional Neural Network model based on Xception architecture | 89.6% | |
| 16 | Varalakshmi Perumal et. Al [17] | CXR and CT images | transfer learning technique | deep learning techniques | 93% | |
| 17 | Mohamed Abd Elaziz et. al. [18] | x-ray datasets | MRFODE feature selection method | KNN classifier | 96.09% | |
| 18 | Shuang Liang et. al. [19] | X-ray and computed tomography images | | convolutional neural network | | |
| 19 | Michael J Horry et. al. [20] | X-Ray Image | | CNN models | 80% | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 0 | Emtia z Hussa in et. Al [21] | chest X-ray image s | CoroD et | CNN model | 91.2 % | |
| 2 1 | Fei Shan et. Al [22] | from chest CT scans | Quantit ative evaluat ion | DL-based system | 91.6 % | |
| 2 2 | Fátim a A. Saiz et. Al [23] | Chest X-ray Image s | SDD30 0 model | deep learning techniques | 92.00 % | 94.9 2% |
| 2 3 | Steph anie A. Harm on et. Al [24] | Chest CT | | deep learning algorithms | 90.8 % | 84% |

## V.    POST –COVID 19 SYMPTOMS & DISEASE

Covid-19 is affect multiple organs and body systems, including heart and blood vessels, lungs, kidneys, skin, brain and nerves. Due to Covid conditions some cautionary signs like reappearance of fever, brain fog means difficult for people to focus and lead to an incapability to concentrate [2], fatigue means an unexplained weakness, this virus takes a heavy effect on the body's immune system and the whole body, patients may continue experiencing fatigue in the long run [2], painful joints or muscles, continuous shortness of breath indicates low oxygen levels in the body, chest pain, and dizziness on standing. The coronavirus is a horrid microbe that can do a lot of damage to body. If infection was moderate to simple, it is possible that the virus did some amount of damage to your respiratory system.

Mucormycosis, an infrequent but serious fungal infection has been observed in a number of Covid-19 patients which is generally referred as "black fungus". It is caused by a group of mucormycetes present naturally in the environment. Those Patients are admitted to a hospital and looking for treatment for simple or critical COVID-19 carry a higher possibility of emerging mucormycosis. It also affects people who are on medication for health problems that reduces their ability to fight environmental pathogens.[1] The disease often establishes in the skin and also affects the lungs and the brain. Large number of mucormycosis cases detected in Delhi, Maharashtra and Gujarat. Major Symptoms of this disease is Pain and redness around the eyes or nose, with fever, headache, altered mental status, coughing, and shortness of breath, bloody vomits. Treatment of mucormycosis is treated as antifungals treatments, mucormycosis may eventually require surgery. It has importance to control diabetes, reduce steroid use, and discontinue immunomodulation drugs. To maintain adequate systemic hydration, the treatment includes infusion of normal saline before infusion of amphotericin B and antifungal therapy. This treatment is for at least 4-6 weeks.

## VI.    CONCLUSION

The early detection and diagnosis of COVID-19 by machine learning techniques and with the minimum cost and difficulties are the basic steps in preventing the disease and the progression of the pandemic. In future, with the integration of machine learning algorithms in the equipment of medical centers, it will be possible to achieve a faster, cheaper, and safer diagnosis of this disease. The use of these techniques in fast diagnostic decision-making of COVID-19 can be an influential tool to reduce human error and can support them to make decisions in critical conditions. This research supports the idea that machine learning models are a promising way for enhancing healthcare and improving the results of diagnostic and healing procedures. Although machine learning is one of the most powerful computing tools in diagnosis of COVID-19, developers should be careful to avoid over fitting and to maximize the generalizability and usefulness of COVID-19 machine learning diagnostic models.

### REFERENCES

1.    Michael Safi (21 April 2021). "India's shocking surge in Covid cases follows baffling decline". The Guardian. Retrieved 29 April 2021.
2.   https://en.wikipedia.org/wiki/COVID- 19_pandemic_in_India
3.    https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects.html
4.    (Javaheri, et al., 2020). CovidCTNet: An Open-Source Deep Learning Approach to Identify Covid-19 Using CT Image
5.    Terry Gao et al. (2020)."Chest X-ray image analysis and classification for COVID-19 pneumonia detection using Deep CNN".
6.    Keidar, D et al. (2021), COVID-19 classification of X-ray images using deep neural networks. Eur Radiol
7.    AramTer-Sarkisov (2020), "COVID-CT-Mask-Net: Prediction of COVID-19 from CT Scans Using Regional Features".
8.    Aram Ter-Sarkisov (2020), "Detection and Segmentation of Lesion Areas in Chest CT Scans For The Prediction of COVID-19"
9.    Aram Ter-Sarkisov (2020), "Lightweight Model For The Prediction of COVID-19 Through The Detection And Segmentation of Lesions in Chest CT Scans".
10.    Amit Kumar Das et al,"Automatic COVID-19 Detection from

X-Ray images using Ensemble Learning with Convolutional Neural Network"

11. Pedro Silvaa et al, (2020},100427."COVID-19 Detection in CT Images with Deep Learning: A Voting-based Scheme and Cross-Datasets Analysis."

12. Vijayakumar Bhandi,et al. (2020), "COVID-19 X-ray Image Retrieval Using Deep Convolutional Neural Networks.", pp-47-55

13. Hwa, S et al. (2020)." Enhanced Canny edge detection for Covid-19 and pneumonia X-Ray images. IOP Conference Series: Materials Science and Engineering."

14. Jain Rachna et al. (2021). "Deep learning based detection and analysis of COVID-19 on chest X-ray images. Applied Intelligence". 51. 1-11. 10.1007/s10489-020-01902-1.

15. Akbari Y.,Hassen Hanadi et al (2020). "COVID-19 Lesion Segmentation using Lung CT Scan Images: Comparative Study based on Active Contour Models". 10.21203/rs.3.rs-40406/v1.

16. Khan Asif et al.(2020)." CoroNet: A Deep Neural Network for Detection and Diagnosis of Covid-19 from Chest X-ray Images".

17. Varalakshmi P. et al.. (2021). "Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features". Applied Intelligence. 51. 10.1007/s10489-020-01831-z.

18. Elsayed Abd Elaziz Mohamed et al. (2020). "New machine learning method for image- based diagnosis of COVID-19". PLOS ONE. 15. 10.1371/journal.pone.0235187.

19. Liang Shuang et al. (2020). "Fast automated detection of COVID-19 from medical images using convolutional neural networks". 10.21203/rs.3.rs-32957/v1.

20. Michael J Horry et al. (2020)." X-Ray Image based COVID-19 Detection using Pre-trained Deep Learning Models". 10.31224/osf.io/wx89s.

21. Hussain Emtiaz, et al. (2021). "CoroDet: A Deep Learning Based Classification for COVID-19 Detection using Chest X-ray Images". Chaos Solitons & Fractals. 142. 110495. 10.1016/j.chaos.2020.110495.

22. Fei Shan, et al. (2020)." Lung Infection Quantification of COVID-19 in CT Images with Deep Learning".

23. Saiz Fatima, et al. (2020). "COVID-19 Detection in Chest X-ray Images using a Deep Learning Approach".10.9781/ijimai.2020.04.003.

24. Harmon Stephanie et al. (2020)." Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets". Nature Communications. 11. 4080. 10.1038/s41467-020-17971-2.

25. Sabri Nurbaity. Raseeda Hamzah, Shafaf Ibrahim, Khyrina Airin Fariza Abu Samah(2020). "COVID-19 Detection for Chest X-Ray Images using Local Binary Pattern". International Journal of Emerging Trends in Engineering Research. 8. 78-81. 10.30534/ijeter/2020/1281.12020.

26. Talaat Ahmed Yousri, Dalia Ewees, Ahmed Al-qaness, Mohammed A. A., Damasevicius Robertas, Elsayed Abd Elaziz Mohamed. (2020). "COVID-19 image classification using deep features and fractional-order marine predators algorithm". Scientific reports. 10. 15364. 10.1038/s41598-020-71294-2.

27. Mostafiz Rafid, Uddin Mohammad, Munna Nur, Reza Md, Rahman Motiur. (2020). "Covid-19 Detection in Chest X-ray through Random Forest Classifier using a Hybridization of Deep CNN and DWT Optimized Features". Journal of King Saud University - Computer and Information Sciences. 10.1016/j.jksuci.2020.12.010.

28. Purohit Kiran, Kesarwani Abhishek, Kisku Dakshina Ranjan,Dalui Mamata. (2020)." COVID-19 Detection on Chest X-Ray and CT scan Images Using Multi-image Augmented Deep Learning Model". 10.1101/2020.07.15.205567.

29. Luca Brunese,Fabio Martinelli,Francesco Mercaldo, Antonella Santone(2020)."Machine learning for coronavirus covid-19 detection from chest x-rays". Pages 2212-2221

30. Khanday Akib, Rabani Syed, Khan Qamar, Rouf Nusrat, Din Masarat. (2020)." Machine learning based approaches for detecting COVID-19 using clinical text data". International Journal of Information Technology. 1-9. 10.1007/s41870-020-00495-9.

31. Description of Logistic Regression Algorithm.https://machinelearningmastery.com/logistic-regression-for-machine-learning/. Accessed 15 May 2019

32. Description of Multinomial Naïve Bayes Algorithm https://www. 3pillarglobal.com/insights/document-classification-using-multi nomial-naive-bayes-classifier. Accessed 15 May 2019

33. Khanday AMUD, Khan QR, Rabani ST. SVMBPI: support vector machine based propaganda identification. SN Appl. Sci. (accepted)

34. Description of Decision Tree Algorithm: https://dataspirant.com/ 2017/01/30/how_decision_tree_algorithm_works/. Accessed 10 July 2019

35. Description of Boosting Algorithm: https://towardsdatascience. Com/boosting. Accessed 10 July 2019

36. Description of Adaboost Algorithm: https://towardsdatascience. Com/boosting-algorithm-adaboost-b673719ee60c. Accessed 10 July 2019

37. Katuwal R, Suganthan PN (2018) Enhancing Multi-Class Classification of Random Forest using Random Vector Functional Neural Network and Oblique Decision Surfaces, Arxiv:1802.01240v1

38. Friedman JH (2002) stochastic gradient boosting. Comput. Stat. Data Anal. 38(4):367–378. ttps://doi.org/10.1016/S0167-9473(01)00065-2

39. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0235187

40. Linda Wang, Zhong Qiu Lin and Alexander Wong1,2, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images", arXiv:2003.09871v4 [eess.IV] 11 May 2020

41. Sara Hosseinzadeh Kassania et al., "Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach",arXiv:2004.10641v1 [eess.IV] 22 Apr 2020

42. Tahereh Javaheri et al., CovidCTNet: An Open-Source Deep Learning Approach to Identify Covid-19 Using CT Image

43. Terry Gao, "Chest X-ray image analysis and classification for COVID-19 pneumonia detection using Deep CNN" version posted October 12, 2020

44. medRxiv preprint, "COVID-19 Classification of X-ray Images Using Deep Neural Networks",
Version posted January 26, 2021

45. Aram Ter-Sarkisov, "COVID-CT-Mask-Net: Prediction of COVID-19 from CT Scans Using Regional Features", version posted October 14, 2020

46. Ali Narin et al., "Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks", 2020.

Paper ID: CS-10

# Machine Learning In Artificial Intelligence

**Miss. Kalyani V. Patil**

*Department of Computer Science and IT*
*BAS P.O.Nahatacollege, Bhusawal*
*Email- pkalyani362@gmail.com*

*ABSTRACT – Machine learning refers to a group of computer algorithms that learn from examples rather than being clearly programmed to execute a task. It learns to create a universal rule from a set of real examples. ... Machine learning is the support of artificial intelligence. These days, large amount of data is available everywhere. Therefore, it is very significant to analyse this data in order to remove some useful information and to develop an algorithm based on this analysis. This can be achieved throughout data mining and machine learning. Machine learning is an integral part of artificial intelligence, which is used to blueprint algorithms based on the data trends and past relationships between data. Machine learning is used in a variety of fields such as bioinformatics, intrusion finding, Information retrieval, game playing, marketing, malware detection, image deconvolution and so on. This article presents the work done by a variety of authors in the field of machine learning in various claim areas.*
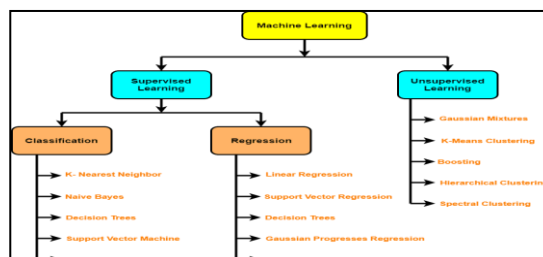
*Keywords -Machine Learning Methods, Supervised learning, unsupervised learning, natural language processing, History of Machine Learning.*

## I. INTRODUCTION

Machine learning is a playing field of artificial intelligence (AI). The aim of machine learning usually is to understand the organization of data and fit that data into models that can be understand and utilized by citizens. Although machine learning is a live field within computer science, it differs from conventional computational approaches. In traditional computing, algorithms are sets of explicitly programmed orders used by computers to estimate or problem solve. Machine learning algorithms as a replacement agree for computers to train on data inputs and use statistical analysis in order to output values that fall within a exacting range. Because of this, machine learning facilitates computers in build models from trial data in order to mechanize decision-making processes based on information inputs. Machine learning is a continuously increasing field. Because of this, there are some considerations to keep in brain as you work with machine learning methodologies, or study the contact of machine learning processes.

## II. MACHINE LEARNING METHODS

In machine learning, tasks are generally classify into large categories. These categories are based on how learning is received or how advice on the learning is given to the system developed.



Two of the majority widely adopt machine learning methods are supervised learning which trains algorithms based on example input and output data that is labelled by humans, and unsupervised learning which provides the algorithm with no labelled data in order to assign it to find arrangement within its input data.

**1]Supervised Learning**

In supervised learning, the computer is provided with example inputs that are labeled with their most wanted outputs. The reason of this method is for the algorithm to be able to "learn" by comparing its real output with the "taught" outputs to find errors, and modify the model consequently. Supervised learning therefore uses patterns to predict label standards on additional unlabeled data.

For example, with supervised learning, an algorithm may be feed data with images of sharks labeled as fish and images of oceans labeled as water. By being trained on this data, the supervised learning algorithm should be able to later on identify unlabeled shark images as fish and unlabeled ocean images as water.

**2] Unsupervised Learning**

In unsupervised learning, data is unlabeled, so the learning algorithm is left to find commonalities surrounded by its input data. As unlabeled data are more abundant than labeled data, machine learning methods that make easy unsupervised learning are particularly valuable.

The aim of unsupervised learning may be as clear-cut as discovering hidden patterns within a dataset, but it may also have a goal of feature learning, which allows the computational machine to without human intervention discover the representations that are needed to classify raw data.

Unsupervised learning is usually used for transactional data. You may have a large dataset of clients and their purchases, but as a human you will likely not be able to make sense of what parallel attributes can be strained from customer profiles and their types of purchases. With this data feed into an unsupervised learning algorithm, it may be single-minded that women of a certain age range who buy unscented soaps are likely to be pregnant, and therefore a marketing promotion related to pregnancy

and baby products can be targeted to this audience order to increase their number of purchases.

**III.** **EARLY HISTORY OF MACHINE LEARNING**

The primary case of neural networks was in 1943, when neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper about neurons, and how they job. They decided to create a model of this using an electrical path, and therefore the neural network was born.In 1950, Alan Turing produced the world-famous Turing Test. This test is literally simple - for a computer to pass, it has to be able to induce a human that it is a human and not a computer.1952 saw the original computer program which could learn as it ran. It was games which play checkers, created by Arthur Samuel.Frank Rosenblatt designed the first synthetic neural network in 1958, called Perception. The main target of this was pattern and figure recognition.

**IV.** **CONCLUSION**

Machine Learning can be a Supervised or Unsupervised. If you have smaller amount of data and plainly labeled data for training, output for Supervised Learning. Unsupervised Learning would generally give superior performance and results for big data sets.

**REFERENCES**

1. Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892. https://en.wikipedia.org › wiki › Machine_learning
2. Artificial Intelligence By Puntambekar , Author: Puntambekar, A. M.Publication: Tech- Max Publication
3. Artificial Intelligence By Rich,Author: Rich, E/ Knight, K.Edition: 2nd ed Publication: Tata McGraw Hill publication (TMH)
4. Introduction To Artificial Intelligence & Expert Systems By PattersonAuthor: Patterson, Dan W.Publication: PHI Education
5. 5. Journal of Machine Learning Research https://www.jmlr.org
6. S B Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques, Informatica, volume 31, p. 249 – 268, Posted: 2007

# The Role of Telephone Banking in Central Bank of India special reference to Jalgaon District

**Prof.Dr. Kailas Chandrakant Deshmukh**

*Associate Professor*
*Nanasaheb Y.N. Chavan Arts, Science and Commerce College,*
*Chalisgaon Dist. Jalgaon, Maharashtra, India*
*kcdmeg@rediffmail.com*

**ABSTRACT -** *Banking business is emerging in both private and public sectors ,also cooperative banking is developed in Maharashtra ,however customers is most important element of any business particularly in today customers are playing most important role in modern banking ,to develop relationship between bank and customer it most important to use management principles or techniques when it implemented through electronic tools then it become electronic –Customer Relationship Management or e-CRM . The paper finds benefits and limitations of implementing e-CRM technique TELEPHONE BANKING in the area of Jalgaon district of Maharashtra.*

*Keywords - Person or organization who has accounts in banks, Relation: Association among customer and bank, Bank: A financial institution providing financial services to the Customer, Management: A set techniques which increases efficiency or productivity CRM-Customer Relationship Management*

## IV.    INTRODUCTION

Jalgaon is one of the districts of Maharashtra, located in western India. It should be noted that like Jalgaon and Nandurbar it is a part of north Maharashtra region, and its boundaries are confined to northern Deccan Plateau. Jalgaon was inhabited by 460,468 people, as per the 2011 census. It also serves as an attraction spot for tourists due to its unique location, at a mere distance of 59 km from the popular Ajanta Caves. Additionally, it connectivity with the railway zone, National Highway No.6 and the presence of an airport make it feasible for the tourists to visit the place. Also the district features well developed roads, furnished shopping malls, and proper transport infrastructure and communication facilities etc. facilities. Major crops produced in the district are cotton, banana, wheat, millet, groundnut, sugarcane etc. However, due to its contribution of half of the states banana production it has been given the title of 'Banana-City'.

The several regional developments such as setup of educational institutes, provision of medical facilities etc initiated in Jalgaon has imparted numerous facilities and even strengthened the local economy. Jalgaon is also known for its excellent gold quality.

**Demographic**

To analyze the efficacy of e-CRM implementation it is of paramount significance to have knowledge about the demographics of Jalgaon. The total population of Jalgaon district was 3,679,936, with 52% males and 48% females (as per the 2001 census). In addition there has also been observed a higher literacy rate of approximately 76%, much higher than the national average of 59.5%, featuring an interesting male literacy percentile of 80% and female percentile of 71%. Interestingly, the 13% of the total population of Jalgaon is below 6 years of age. The common spoken vernacular language of Jalgaon is Marathi.

## Economy

Besides the astonishing demographics of Jalgaon district mentioned in previous section, here an in-depth discussion of its economy has been made. Previously it has been mentioned that Jalgaon district is known as 'banana city' due to its higher production of banana. Noticeably, more than 16% of the India's bananas are produced in Jalgaon, i.e. approximately 3% of the world's total banana production. It has been estimated by an association of banana growers of Maharashtra, called Mahabanana that approximately 66% of the Maharashtra's total land under banana cultivation is occupied by Jalgaon. Additionally, Jalgaon is also one of the world's largest pulse milling centers and a leading cotton producing region.The involvement of Jalgaon in the pharmacies is also remarkable as it possesses more than 500 pharmacies, chemists and drugstores.

Jalgaon Municipal Corporation is the administrative body for the city, which is housed in the 17 floor building that serves as a landmark of the city. Adjacent to is located the Golani complex comprising more than 500 shops and flats. Crossing MahTelephoneBankinga Gandhi road towards the Shastri tower visitors will find Phule market, well known as a retail market for clothing, cosmetics, and almost all consumable items. Further along is another one of Jalgaon's important markets, known as Dana Bazaar. Additionally, the heart of the city, Khandesh Central is the first state of the art shopping mall in Jalgaon and was constructed in December, 2010. The mall covers an area of 5.5 Lac sqft for shopping, comprising Big Bazaar, INOX four screens multiplex and several other stores and wholesale market for food and grains. Establishing itself as one of the biggest jewelry markets in Maharashtra, Jalgaon is known for its excellent gold quality and designs and is thus known as 'Golden City'.

There are several industries set-up in Jalgaon dealing in drip irrigation, fruit processing, pulse milling, pipe manufacturing, bio-fertilizer production, green and solar energy, mattress manufacturing and printing of educational books. Some industries also deal with construction, hardware and paints manufacturing.

Exhibiting a large scale involvement in agriculture, Jalgaon also serves as a distribution center of agricultural goods and agriculture input industry, with a participation in manufacturing bio-fertilizers. Jalgaon also establishes as a center for cotton-textile and vegetable oil mills, especially groundnut oil and hydrogenation plants. To summarize it can be asserted that Jalgaon is nowadays establishing itself as a sound trading centre of Maharashtra and a leading producer of sugarcane. Apart from the other food crops, onion dehydration plants in Jalgaon has annual onion dehydration capacity of 10,000 tons, accounting for a 15% of total exports of dehydrated onions from India. Similarly, pulses one of the staple foods of Indian cuisine has also experienced a large scale export from Jalgaon. Being a basic crop, the pulse industry emphasizes more on its cleaning, processing and packing through different processes. The year 1970 experienced the establishment of numerous pulse milling industries by the companies who pioneered in this field and eventually led to the creation of the world's largest milling center for pulse products in Jalgaon with majority of the plants producing large quantities for India and world in unison with providing opportunity to approximately 10,000 workers to earn their living. The Jalgaon pulse industry was a leading exporter of the pulses and its products to over 25 countries worldwide and thus earned the name of world's best millers. Burma, Canada and Australia are some of the major countries importing raw pulses from Jalgaon.

## Banks in Jalgaon

From the aforementioned agricultural and industrial status of the Jalgaon, it can be affirmed that Jalgaon has a major contribution in the economy of Maharashtra. This eventually has led the financial institutions of India to work on the expansion of their banks and services across the districts. To achieve a successful establishment, public and private banks have been targeting to get higher acquisition of customers for long term. Keeping in mind the public banks, several banks such as State bank of India, Punjab National Bank, Canara Bank, Central Bank of India, Corporation Bank etc are performing the necessary functions across the district. On the other hand, private banks also do not lag behind. HDFC Bank set up its first branch in Jalgaon in August 2002, Axis (UTI) Bank in December, 2003, ICICI Bank in March, 2006, INGVysya in January, 2009 with Kotak still in queue to launch its branch. Axis Bank opened its second branch in Jalgaon in September, 2010.

Apart from these banks a number of other banks such as JalgaonJanta Bank, Jalgaon Co-operative Bank etc are attempting to apply certain approaches in order to successfully achieve higher

customer retention by establishing a better customer relationship management.

The distribution of financial institutions in the district of Jalgaon in North Maharashtra region is illustrated in table 1

*Table 1.Bank in Jalgaon district*

| BANK | JALGAON | MAHARASHTRA | INDIA |
|---|---|---|---|
| NATIONALIZED | 21 | 27 | 27 |
| PRIVATE | 6 | 21 | 21 |
| **Bank's Presence Analysis** | | | |
| **National Bank:** 77.77 % | | | |
| **Private Banks:** 28.57 % | | | |
| * Here the affinity or banking presence has been estimated as per financial institutions available in Maharashtra state | | | |

Source: www.eJalgaon.com,

From the table 1 it can be inferred that approximately 77.77% banks serving or fulfilling the financial support needs are nationalized banks while 28.57% banks are private banks. It can be clearly observed that there is a huge difference in the distribution of national and private banks which can be attributed to confined business possibilities and financial activities. Surprisingly, the statistics pertaining to education, industries and other economical factors of Jalgaon has made Jalgaon a prime region of Maharashtra economy that has grabbed the attention of economical-politico factor to have higher service provisioning and higher living standards. Emphasizing on the implementation of e-CRM, the literacy and economical activities can be considered sufficient strong enough to enhance the quality and foster competitive banking growth in the region. As discussed previously, the literacy rate for men is 80% and for females is 71%, which further provides a robust base for the implementation of e-CRM in the region.

The following Important e-CRM Techniques are available.

1. email
2. SMS
3. Telephone Banking

4. ATM

For the study we have selected Telephone Banking.

## V. OBJECTIVES

To study benefits and limitations of TELEPHONE BANKING as e-CRM(\electronic – Customer Relations Managements) techniques in Central Bank of India of Jalgaon district of Maharashtra [INDIA].

## VI. RESEARCH METHODOLOGY

The overall research methodology encompasses the following phases:

1. *Defining the Research Objectives*
2. *Literature survey and analysis*
3. *Research Hypotheses definition*
4. *Questionnaires preparation*
5. *Primary Data collection*
6. *Sampling and data processing*
7. *Data analysis and Hypothesis test*
8. *Research conclusion and limitations analysis*

## VII. QUALITATIVE RESEARCH PARADIGM

Qualitative research is basically important for obtaining culturally specific information about values, opinions, behaviors, and social contexts of particular populations (respondents from organization at the associate level). The study design in qualitative research consists of data collection and research questions that are adjusted according to what is learned. Unlike quantitative research, qualitative research consists of the secondary resources based investigation which is based upon seeking answers to questions, systematic use of predefined set of procedures to answer the questions, collect evidence, produce findings etc. Qualitative research is essential when a researcher needs to define the problem more specifically and identify any specific objective or data requirements to be addressed through additional research.

Indeed, e-CRM facility is a relatively new phenomenon in North Maharashtra. Although web banking is proliferating, there is little empirical evidence to help bankers fully understand what constitutes customer satisfaction from a Jalgaon, district customer perspective. Therefore, the imperative of an exploratory study is to gain much needed background pertaining to building a long term customer relationship in the banks. The qualitative research model intends to explore fundamental background of the research objectives, key aspects and variable analysis to formulate research hypothesis

and key constructs. In fact, it would put foundation for identifying customer's requirements, measuring customer's acquisition level towards banks and banks early proposal for customer acquisition and strategies for customer's satisfaction enhancement.

In order to define a list of Pre-transaction, During-transaction and Post-transaction e-CRM features, the researcher will use qualitative methods. The first method (which was used as a minor method) is bank documents and websites. The researcher has explored banks websites in North Maharashtra in order to assess what e-CRM features are available for customers. The second qualitative method used was unstructured and semi-structured interviews with bank managers and employees who are responsible for controlling the e-CRM service activities. To perform qualitative analysis, in this thesis, various literatures such as emergence of financial services, e-CRM in the banking sector, benefits of e-CRM facility to bankers and customers and various measures to enhance quality of service in banks have been studied and key constructs have been defined, which have been further examined through quantitative analysis.

**Sample size**

The required sample size depends on factors such as the proposed data analysis techniques, financial and access to sampling frame. In this study, a fairly representative sample has been randomly drawn by selecting 280 customers of Jalgaon district of Maharashtra

## VIII.     DATA COLLECTION

In research, the predominant emphasis has been made on assessing benefits and limitations of e-CRM for customers of Central Bank of IndiaJalgaon Maharashtra. Considering geographical distributions and associated factors such as population, education, banking facilities, reach of technologies etc, the major respondents have been taken from Jalgaon,

Taking into consideration of relatively higher economic development in Jalgaondistrict affirms suitability of the selected samples. On the contrary, the other districts which are primarily based on agriculture have show relatively lower interest towards e-CRM facilities because of lack of awareness, education and other constructs. The diversity of responses from different hierarchies and demographic background affirms optimal research paradigm and ensures best possible information retrieval to enable effective strategic formulation to

provide or enrich e-CRM facilities in the considered districts of the Maharashtra.

Before performing descriptive study and analysis, the demographic assessment of responses can be significant. The following section discusses the demographic analysis of the presented study.

**Demographic Findings in Jalgaon district**

The demographic analysis for the obtained responses is presented as follows:

**Gender Distribution**

Table 5.2 (a) presents the gender distribution of the respondents from Jalgaon districts.

The depiction based on gender is important because in present day scenarios, males as well as females have the equal contribution towards regional as well as national economy, development etc; however based on respective comfort level, interest and expectations towards efficient banking solution, the perception and decision processes might vary between the respondents (male and female). There can be many situation where male can somewhat bear being in queue for financial process at the banks or the financial institution, while women feel uncomfortable under these circumstances. Considering these facts, analyzing gender factor can be of paramount significance.

*Table 1 Gender distribution of the respondents from Jalgoan district*

| Gender | Number of participants |
|--------|------------------------|
| Male | 210 (75 %) |
| Female | 70 (25%) |

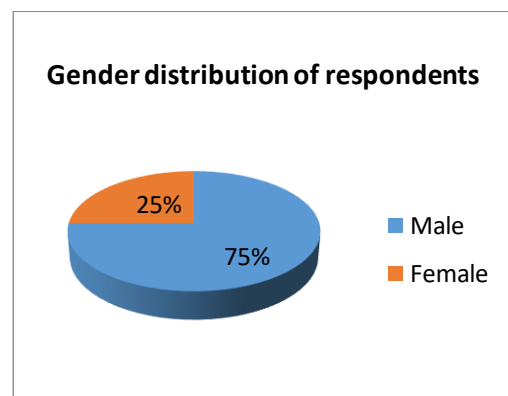**As depicted in Table 1 approximate 75% of respondents were male, while only 25% of respondents were female.**



**Gender distribution of respondents**

It exhibits that there exists higher gap in account holder proportion (gender) in these districts that do affirm that the relatively lower industrial development and conventional business modules confines financial transactions till males. However, the recent government initiatives and various subsidies have motivated both genders to have bank account to avail benefits.

## Age Distribution

This is the fact that the age of a human being is directly related to its maturity to respond and its inherent responsibility. In addition, age and maturity of respondents have higher impact of believe factor and responsible feedback. Thus, considering age as a significant factor to impact ones decision ability and sensitive assessment, in this study age of the respondents have been taken into consideration.
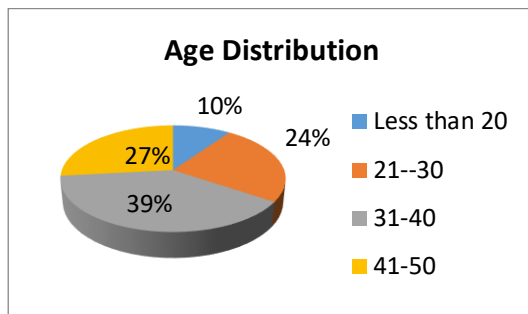


As depicted in Table 4.2(b), it can be seen that the majority of respondents counting 32% (90 out of 280) were in the age range of 31-40 years. Similarly, 20.35% respondents (57 out of 280) were in the age range of 21-30. Since, the banking sector in Jalgaon has shown growth in last 15-20 years and hence only 17.85% of respondents with age range of 50-60 were found having bank accounts. A total of 61 out of 280 respondents were in the range of 41-50 years. Only two respondents, especially belonging to educational institutions counting 22 out of 280 (8%) were of age less than 20 years.

*Table 2) Age distribution of the respondents from Jalgaon districts*

| Age | |
|---|---|
| Less than 20 | 22 (8%) |
| 21-30 | 57 (20.35%) |
| 31-40 | 90(32%) |
| 41-50 | 61 (22%) |
| 50-60 | 50 (17.85%) |

## Educational Background of the respondents

Similar to the age, maturity etc the key factor that impacts one's ability to respond and make decision is the education. In this thesis, the educational background of the respondents has been examined. Education and experience have direct impact on personal (someone's) ability to understand and make proper decision. Therefore, in this thesis, to retrieve optimal data and reliability respondents from varied background, education level and thoughts were considered.

*Table 3 Education of the respondents in Jalgoan district*

| Education | |
|---|---|
| Secondary education | 34(12%) |
| Below graduation | 90 (32%) |
| Graduation | 100 (36%) |
| Post-Graduation | 56(20%) |

Considering the literacy of Jalgaon district, 12% of respondents had minimum secondary education, while 32% of the respondents stated that they have achieved satisfactory education till under-graduation. A significantly higher fraction of respondents (36%) stated that they are graduate and with relatively higher education level approximate 20% respondents affirmed it that they have achieved master's degree or post-graduation. Figure 4.3 depicts that education level of the respondents.
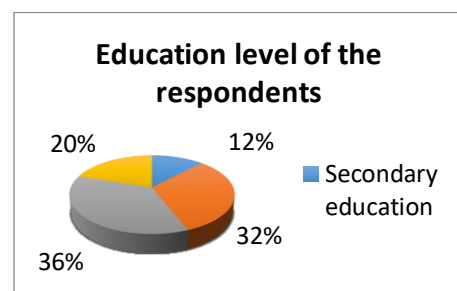


## Income of the respondents

This is the matter of fact that the personal income or buying capacity of a consumer reflects its demands and expectations and since the presented thesis intends to explore the e-CRM facility and its significance, therefore assessing the income of the respondents is vital. In this study approximate 2.7% of respondents stated that their personal income is less than 5000 INR (Indian Rupee), while 14.61% of

respondents agreed that they earn in the range of 5000-15000 INR months. The remaining 29.61% of the respondents stated that their income exist in the range of 15,000-25,000 INR. The remaining of 36.9% respondents were found having better monthly income and their income range was existing somewhat in between 25,000 INR to 35,000 INR. Interestingly, relatively higher fraction (23.84%) of respondents was having monthly salary of more than 35,000 INR. No doubt, majority of the respondents in this category were belonging to certain industry, particularly manufacturing, government sector and personal business or self-employment activities.

*Table 4 Average monthly Income of the respondents*

| Average  Monthly Income | |
|---|---|
| Less than 5000 Rs | 7 (2.69%) |
| 5000-15000  Rs | 38 (14.61%) |
| 15000-25000 Rs | 77 (29.61%) |
| 25000-35000 Rs | 96 (36.92%) |
| More than 35000 Rs | 62 (23.84%) |

Following figure presents the average monthly income of the respondents.
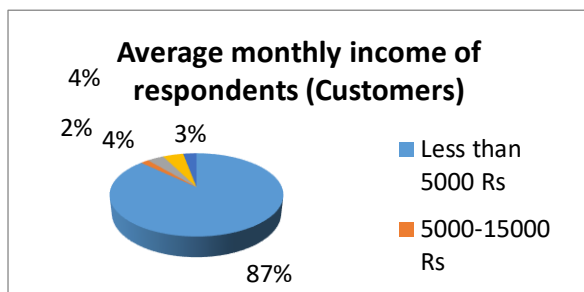


*Figure 4 Average monthly incomes of respondents (customers) from Jalgoan district*

### Experience Level

The personal experience of a human being plays significant role in its respective decision process and thought process. Considering this factor, in this thesis the consumers were requested to state their experience, particularly banking experience or duration of banking process. The study revealed that among the respondents from Jalgaon district approximate 30% of the respondents were having experience more than 5 years of banking (as a regular customer), followed by 40% of respondents (179 out of 280) with more than 5 years of experience.

*Table 5. Education of the respondents in Jalgoan district*

| Experience | |
|---|---|
| Less than 5 years (banking) | 84 (30%) |
| More than 5 years (banking) | 112 (40%) |
| More than 10 years with banking experiences | 84 (30%) |

Among the respondents, a total of 84 (30%) respondents were having more than 10 years of banking experience. Here, it is important to analyze this factor because with a long term experience one better understands what is needed, what is lacked and how to enable.
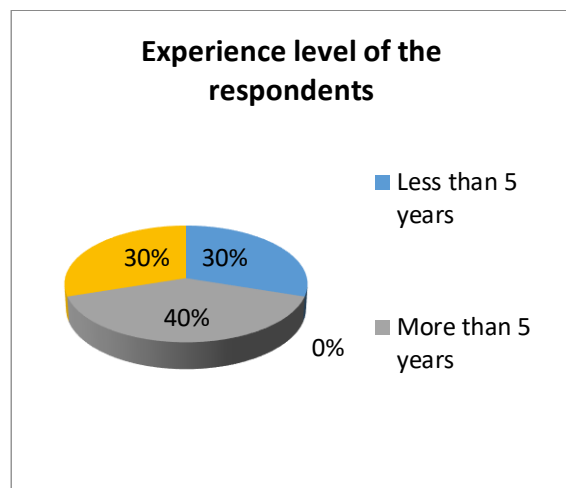


*Figure 6 Experience level of the respondents from Jalgaon district*

Similar to the demographic study of the respondents from Jalgaon, in this thesis the separate demographic assessment has been done. Since, all the considered (three) districts do have different economical, industrial, education etc. background, and therefore assessing respective distinct demographic assessment can play significant role.

### Data Analysis

The following is table shows satisfaction level of customer using TELEPHONE BANKING in Central

Bank of India in Jalgaon district in terms problem of

**Table 7**

| Problem facing use of using e-CRM techniques TELEPHONE BANKING | Responses | No of respondent | % of respondent Satisfied |
|---|---|---|---|
| e-CRM technique TELEPHONE BANKING | | | |
| Lack of knowledge of customer service representation | 210 | 280 | 53% |
| Absence immediate customer to service | 200 | 280 | 71% |
| Lack of prompt service | 190 | 280 | 67% |
| Lack of clear guidance | 210 | 280 | 75% |

technology usage faced by the customer

**Findings**

1. Maximum 75% customers are satisfied with the problem of lack of clear guidance
2. Also 71% of customer has problem of immediate service to customer
3. 67% of customer are not satisfied for lack of prompt
4. 53% of customers reporting problem of lack of knowledge

**Conclusion**

The customers in central bank of India of Jalgaon district lot of problem in Telephone banking mainly regarding lack of clear guidance and not connecting immediately.

**Recommendations:**

1. There should quick response from bank.
2. Bank should provide brief and clear information to customer through telephone.
3. Telephone banking should be 24x7 hours.

**REFERENCES**

1. Kasyap, MM (1969). Planning of survey Library Herald, 2(1&2), 195-199.
2. ALA (1989). American Library Association Presidential Committee on Information Literacy Final Report. Chicago: American Library Association.
3. Fortier,John D.(1998). Wisconsin's Model Academic Standards for Information and Technology Literacy. West Indies.
4. Vishala,B. K. and Bhandli, M. K. (2006). Information Literacy teaching in higher education environment. International convention CALIBER, 4, (pp201-208)
5. Joint,Nicholas (2005) e-Literacy or Information Literacy: Which concept should we prefer? Library Review. 54(9).505-507.
6. Gedam, P. G. and Agashe, A. T. (2009). Information

# A Survey of Cannon-Bard Theory of emotion

**Prof. Mrs. S. G. Waghulade / Mahajan**

*Bhusawal Arts, Science & P. O.Nahata Commerce Jr. College,*

*Bhusawal Dist. Jalgaon (MS), India*

*Email- urmila.waghulade@gmail.com*

**ABSTRACT - The Cannon-Bard theory of emotion, it is also called as the Thalamic theory of emotion, is a physiological explanation of emotion developed by Walter Cannon and Philip Bard. Cannon-Bard theory says that we feel emotions and experience physiological actions such as sweating, trembling, and muscle tension simultaneously. The thalamic region is a major role in Cannon-Bard theory of emotion. This article presents the Cannon-Bard Theory of emotion to detect the emotion in the human body expressing the physiological actions. This article survey of Cannon-Bard Theory of emotion determine the appropriate capturing real situation and predict the emotional responses triggered by action presented in that human body physically.**

*Keywords -thalamic,cortex,viscera,amygdala.*

## I. INTRODUCTION

Walter Bradford Cannon (1871–1945) was a physiologist at Harvard University, who is probably best known for his classic treatise on homeostasis. Philip Bard (1898–1977) was a P.hd student of Cannon's, and they both developed a model of emotion called the Cannon–Bard Theory. Cannon was an experiment on animal physiology and studied that. Under that studies, Cannon and Bard remark the role of the brain ingenerating physiological responses and feelings; this role is important in their explanation of emotion experience and production. [8]

## II. CANNON-BARD THEORY

Cannonsays externalincentive activates receptors and this excitation starts impulses toward the cortex. After the arriving in the cortex, the impulses are associated with conditioned processes that determine the direction of the succeeding response. It is the excite response to the thalamic processes. Once the thalamic processes are activated, they are ready to release. The thalamic neurons fire in a specific combination in a given emotional expression. These neurons then release immediately and intensely. Cannon wrote that inside and near the thalamus, the neurons responsible for an emotional expression lie close to the transfer in the sensory path from the periphery to the cortex, and when these neurons fire in a specific combination they innervate muscles and viscera and excite different paths to the cortex by direct connection.[6]

The key element of the Cannon–Bard theory of emotion is that when the thalamic release is occurs, the bodily changes occur almost at the same time with the emotional experience. The bodily changes and emotional experience take place separately and independently of one another; physiological activation does not have to lead up emotional expression or experience. The theory state that the thalamic region is the brain area responsible for emotional responses to experienced impulses.[6]

Cannon conveys the declaration that serves as the basis for his theory of emotion which state the thalamic region is the coordinating center for emotional response. First, after the discharge of the cerebrum anterior to the thalamus in animal test issue, the animals steady to display rage-like emotional reactions. These reactions cause when the thalamus is then discharge. Secondly, a tumor on one side of the

thalamus can result in unilateral laughter or grimace under the suitable conditions, while cortical and voluntary control of the same muscles is bilateral. Finally, temporary condition of cortical control of lower centers from light amnesia or permanent condition by disease (e.g. tumor or lesion) can cause out of control and prolonged weeping or laughing.[2]

The Cannon-Bard theory tells that the lower part of the brain, also called the thalamus, controls your incident of emotion. Simultaneously, the higher part of the brain, also called the cortex, controls the expression of emotion. It is believed that these two parts of the brain react at that same time.[8]
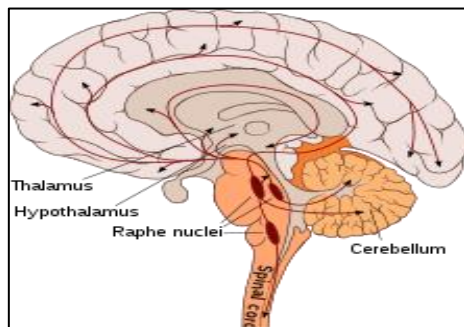


### III. THE CANNON-BARD THEORY WORKS

It is propose that emotions result when the thalamus sends a message to the brain in reation to a impulse, resulting in a physiological response.[10]

For example: I see a snake --> I am scared, and I begin to shake.

By Cannon-Bard theory of emotion, we react to a impulse and incident the associated emotion at an equivalent. The physical responses are not dependent upon the emotional reaction, or in reverse.[3]

When an action occurs, the thalamus transfer a signal to the amygdala. The amygdala is the little, oval-shaped design inside the brain that plays an main role in emotional processing, containing emotions such as fear and rage. The thalamus transfer signals to the autonomic nervous system, resulting in physical reactions such as muscle strain, shaking, and sweating. It's also says that the "fight or flight" response.[1]

### IV. EFFECT OF FIGHT OR FLIGHT ON YOUR HEALTH[5]

Pressure begins in your brain and your senses communicate possible dangers to it. For example, suppose you see a car coming regarding you

or hear a some loud sound. The Cannon-Bard theory present that your amygdala action what you see and hear, transferring possible danger to the hypothalamus.

Your amygdala actions emotions are transfer possible danger to brain, your brain has a tension response to stay and fight the danger or run aside.[4]

When your amygdala initiates a damage signal, your hypothalamus alerts your sympathetic nervous system by transmit signals to your adrenal glands. [2]Adrenaline begins pumping by your veins, quickly resulting in:

- Faster heartbeat
- Higher pulse rate
- Increased blood pressure
- Faster breathing
- Heightened senses
- Alertness

Each of these physiological responses happens so quickly that you are not sure which is first. [8]
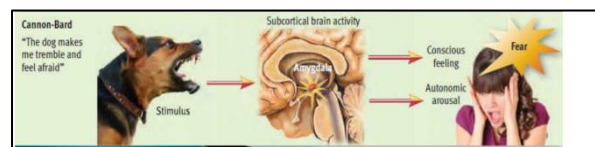


### V. CONCLUSION

Survey of the Cannon-Bard theory of emotion tells that we can detect the emotion of human body by it's physically action taken by human on that incident. In future works on this topic new physical experiment, new concept, new technology and new knowledge is use to improve the extraction of emotion detection process. In future using this theory read the human emotion using the physically action taken by human body and solve the human untold problems and also examine the weakness, illness of human body by its physical action done by human on that incident.

#### REFERENCES

1. Glenn E. Weisfeld and Stefan M. M. Goetz.Department of Psychology, Wayne State University, Detroit, MI 48202, USA. Applying Evolutionary Thinking to the Study of Emotion.Published online 2013 Jul 17. doi: 10.3390/bs3030388
2. Gordan R, Gwathmey JK, Xie LH. Autonomic and endocrine control of cardiovascular function. World J Cardiol. 2015;7(4):204-14. doi:10.4330/wjc.v7.i4.204

3. Duval ER, Javanbakht A, Liberzon I. Neural circuits in anxiety and stress disorders: A focused review. TherClin Risk Manag. 2015;11:115-26. doi:10.2147/TCRM.S48528

5. Goldstein DS. Adrenal responses to stress. Cell MolNeurobiol. 2010;30(8):1433-40. doi:10.1007/s10571-010-9606-9

6. Sullivan LE. Cannon-Bard theory. The SAGE Glossary of the Social and Behavioral Sciences. 2017.

7. https://www.webmd.com/mental-health/what-is-cannon-bard-theory

4. Kollack-Walker S,WatsonSJ,Akil H. Social stress in hamsters:defeat activates specific neurocircuits within the brain.J Neurosci.1997;17(22):8842-8855

8. https://en.wikipedia.org/wiki/Cannon%E2%80%93Bard_theory

9. https://mgkvp.ac.in/Uploads/Lectures/30/4149.pdf

10. https://www.verywellmind.com/what-is-the-cannon-bard-theory-279496

**Paper ID: CS-13**

# Web accessibility issues and challenges for disabled Person

**Mr. Harshal V. Patil**

*Assistant Professor*

*DOCS and IT, Bhusawal Arts, Science &P.O.Nahata Commerce College, Bhusawal*

*harshal4patil@gmail.com*

*ABSTRACT - This paper will discuss how to developed website using web accessibility tools which help disabled. An accessible web site is one which can be used by all its intended visitors. Inaccessible web sites can pose significant barriers to people with disabilities. The challenge of developing web content accessible by everyone has motivated the evolvement of a number of techniques to address web accessibility issues. Unfortunately, web developers often lack sufficient knowledge about these guidelines and techniques to develop accessible web sites. This paper presents a survey of common web accessibility problems. Different studies and reports have been analyzed in order to summarize the most common web accessibility problems.*

*Keywords -Screen reader, JAWS, Dyslexia.*

## I. INTRODUCTION

Currently, the World Wide Web (the Web) is present in all areas of our lives. The Web has firmly entered in our everyday life and has changed our way of making shopping, managing business, accessing to e-government services, having education, enjoying entertainment, and so on. Making websites accessible for all disabled people is critical. The Web has a huge potential to transform their lives, allowing them independent access to information and services for the first time. But, if websites are badly designed and badly supported, this potential is lost and the web becomes just another arena of exclusion.

Web accessibility means that people with disabilities can use the Web. It therefore makes sense to find out whether a website really is accessible by testing it with actual users with disabilities. Accessible websites benefit everyone.

Many disabled people use additional devices (and strategies) to help them use the Web:
– Screen reader
The W3C, WAI, and WCAG is helping developers understand what they should be doing to make their websites accessible.

## II. REVIEW OF LITERATURE

**Terry Thompson Sheryl** in his paper "International Research on Web Accessibility for Persons with Disabilities" identifies countries that may employ promising practices. Given the performance of the "Top Ten" countries relative to other countries, a deeper exploration of each of these countries is a logical next step, although even these countries' results show considerable room for improvement.

**Kanchi V. and Patil H.** "E-participation of Visually Challenged – A Case Study"

**David A. Bradbard** in his paper "Web Accessibility Theory and Practice: An Introduction for University Faculty"

## III. OBJECTIVES

- To explore new areas of career for visually challenged people.
- To explore new possibilities and fields where visually challenged people can make their career and work efficiently as well as independently.
- To improve the lives of people with disabilities (human-centered motivations)

- To capitalize on the a wider audience or consumer base (marketing or economic-centered motivations)
- To avoid lawsuits and/or bad press (public relations and punishment-centered motivations)

## IV. WEB ACCESSIBILITY ISSUES FOR DISABLED

For those unfamiliar with accessibility issues pertaining to Web page design, consider that many users may be operating in contexts very different from your own:

- They may not be able to see, hear, move, or may not be able to process some types of information easily or at all.
- They may have difficulty reading or comprehending text.
- They may not have or be able to use a keyboard or mouse.
- They may have a text-only screen, a small screen, or a slow Internet connection.
- They may not speak or understand fluently the language in which the document is written.
- They may be in a situation where their eyes, ears, or hands are busy or interfered with (e.g., driving to work, working in a loud environment, etc.).
- They may have an early version of a browser, a different browser entirely, a voice browser, or a different operating system.

## V. WEB ACCESSIBILITY TOOLS

Many web designers, developers, and evaluators are introduced to web access through accessibility tools. All accessibility tools perform automated checks of web pages for accessibility issues and all generally have additional features, but each tool targets different audiences. In order to help web designers, developers, and evaluators choose an appropriate tool for their purposes.

However, web accessibility requires more than just accessibility tools; it requires human judgment. It is important to remember that accessibility tools can only partially check accessibility through automation. The real key is to learn and understand the web accessibility standards rather than relying on a tool to determine if a page is accessible or not.

## Screen Reader

Screen readers are applications that visually-impaired users employ to assist them in navigating websites and applications. These applications read out text and navigation items using a synthesized voice. To truly understand the experience you need to install a reader such as JAWS, and attempt to navigate your application with your monitor turned off.

## Using JAWS to Evaluate Web Accessibility

JAWS (Job Access with Speech) is a computer screen reader program for Microsoft Windows that allows blind and visually impaired users to read the screen either with a text-to-speech.

- While working in JAWS, keep the following guidelines in mind:
- While JAWS can be used for accessing Windows and most Windows applications, we will be focusing on accessing web content only.
- Make sure that NumLock is off.
- You will probably want to test JAWS in Internet Explorer, even if it is not your primary browser.
- Maximize the browser window.
- Remember that screen reader users typically do not use a mouse. As you become more comfortable with JAWS, try using only the keyboard.
- Keep in mind that most IE shortcut keys will work when using JAWS.
- The page may not scroll while you read, so you may hear content being read by JAWS that isn't visible on the screen.

## VI. COMMITMENT AND ACCOUNTABILITY OF WEB ACCESSIBILITY

**Awareness:** The foundation of any kind of commitment to web accessibility is awareness of the issues. Most web developers are not opposed to the concept of making the internet accessible to people with disabilities. Most accessibility errors on web sites are the result of lack of awareness, rather than malice or apathy.

**Leadership:** Understanding the issues is an important first step, but it does not solve the problem, especially in large organizations. If the leadership of an organization does not express commitment to web accessibility, chances are low that the organization's web content will be accessible. Oftentimes, a handful of developers make their own content accessible while the majority doesn't bother to, since it is not expected of them.

**Policies and Procedures:** Even when leaders express their commitment to an idea, if the idea is not backed up by policy, the idea tends to get lost among the day-

to-day routines. The best approach for a large organization is to create an internal policy that outlines specific standards, procedures, and methods for monitoring compliance.

## VII. LAWS AND STANDARDS

Applicable laws include ADA, IDEA, and the Rehabilitation Act of 1973 (Sections 504 and Section 508). Many international laws also address accessibility.

The Web Content Accessibility Guidelines provide an international set of guidelines. They are developed by the Worldwide Web Consortium (W3C), the governing body of the web. These guidelines are the basis of most web accessibility law in the world. Version 2.0 of these guidelines, published in December 2008, is based on four principles:

- **Perceivable**: Available to the senses (vision and hearing primarily) either through the browser or through assistive technologies (e.g. screen readers, screen enlargers, etc.)
- **Operable**: Users can interact with all controls and interactive elements using the mouse, keyboard, or an assistive device.
- **Understandable**: Content is clear and limits confusion and ambiguity.
- **Robust**: A wide range of technologies (including old and new user agents and assistive technologies) can access the content.

The disabilities that are most likely to affect your users fall into four major categories and present their own challenges for developers.

### Visual Impairment

Visual impairment includes tunnel vision, color blindness, and, of course, being legally blind. Visually impaired users are often the most affected when it comes to accessibility issues, since everything done on a computer is to some extent visual. These users rely heavily on the keyboard for input and control, and tend to avoid using a mouse. A screen magnifier or screen reader provides a lot of assistance, as do audio events and cues. Some use a Braille display to "read" the text on screen with their fingertips.

### Dyslexia

Dyslexic users are likely to have great difficulty reading a web page or dealing with large amounts of textual controls. In some cases this can extend to difficulty with data entry too, such as using a search function. Strategies such as logical page or screen structure can help, as well as developers adopting design philosophies that make an application as self-explanatory as possible. Dyslexic users may also use screen readers to aid with their comprehension.

### Motor Disabilities

Motor disabilities can range from hand or arm tremors, a loss of limbs, or a lack of control and movement of body parts. In these cases input devices need to be considered, bearing in mind that users may be entirely unable to operate a mouse. Even if the user is unable to type, other assistive technologies exist to help them. Full keyboard navigation and control is essential for these users.

### Cognitive Disabilities

This group is probably the farthest reaching. Disabilities may include faculty impairment such as memory recall, comprehension, and interpretation. In some cases these can be temporary situations; in others they're permanent and there's no easy solution to cater for them all. This group is best served by a combination of techniques used for the other three groups, tied to an overall design philosophy aware that more and more people on the Web are not geeks or programmers, and so approach software differently from the way we expect.

## VIII. KEY PRINCIPLES OF ACCESSIBLE DESIGN

Most accessibility principles can be implemented very easily and will not impact the overall "look and feel" of your web site.
- Provide appropriate alternative text
- ❖ Alternative text is a textual alternative to non-text content, usually images in web pages.
- Make sure that content is well structured and clearly written
- ❖ Write clearly, use clear fonts, and use headings and lists appropriately.
- ❖ Organize your content using true headings (heading 1, heading 2, etc.).
- Help users skip to relevant content
- ❖ In a web page, provide a link that allows the user to skip from navigation to the main content in the page.
- ❖ Provide a table of contents with links to each sub-section

- Provide headers for data tables
- ❖ Tables that are used to organize tabular data should have appropriate table headers. Data cells should be associated with their appropriate headers.
- Ensure users can complete and submit all forms
- ❖ Ensure that every form element (text field, checkbox, dropdown list, etc.) has a label and make sure that label is associated to the correct form element using the <label> tag. Also make sure the user can submit the form and recover from any errors, such as the failure to fill in all required fields
- Ensure links make sense out of context
- ❖ Every link should make sense if the link text is read by itself. Certain phrases like "click here" and "more" must be avoided
- Caption and/or provide transcripts for media
- ❖ Videos and live audio must have captions and a transcript. With archived audio, a transcription may be sufficient.
- Do not rely on color alone to convey meaning
- ❖ The use of color can enhance comprehension, but do not use color alone to convey information. Make sure that color contrast is strong.
- Design to standards
- ❖ HTML compliant and accessible pages are more robust and provide better search engine optimization. Cascading Style Sheets (CSS) allow you to separate content from presentation. This provides more flexibility and accessibility of your content.

## IX. EXPERIMENT OF ACCESSIBLE DESIGN

With respect to enhancing website accessibility for people with various degrees of visual impairments inclusion of website accessibility features such as the following are being identified and implemented. Some of the most common eye disorders such as Retinopathy, Glaucoma, Cataract and Retinitis Pigments in which the strength of the vision gradually deteriorates, the contrast sensitivity of the eye is greatly affected causing insensitivity to different levels of brightness and similar shades. To help such persons with low vision, a second version of the college website with greater contrast is being considered for design.

The webpage's are being recorded with semantically meaningful HTML so that the visually challenged can have easier manoeuvrability and navigability across the webpage's using the text to speech software. The measures include providing textual description for images, easy to understand descriptions to links, etc. Keeping the webpage viewers with low vision in mind, the facility to scale the images and view them enlarged is also being incorporated. The webpage's are being redesigned using alternate style sheets with regular and larger font sizes, ensuring the layout of the pages is not compromised when text-only zoom is enabled in the browser. In case of webpage's where a lot of description is present, text only versions are also being offered so that such text can be manipulated suiting the requirement of the user. For people with color vision deficiency, use of red and green colors is minimized as far as possible. A lot of standard keyboard shortcuts are also incorporated into the webpage's ensuring easier navigation with keyboard and reducing the use of mouse cursor.
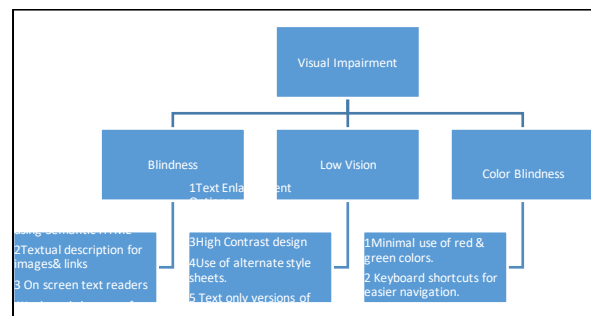


*Figure 2: Different Kinds of Visual Impairments and Web Accessibility Options Incorporated*

It is also realized that the use of college website can be greatly enhanced if the website is made available in other languages. Using the Google's free online language translation service that instantly translates text and web pages, language translation plug-in tool is added to all the webpage's of the college website. Thanks to Google Translate, now the college website is available in over 60 languages, of which six are Indian languages.

## X. CONCLUSION

The web offers so many opportunities to people with disabilities that are unavailable through any other medium. It offers independence and freedom. However, if a web site is not created with web accessibility in mind, it may exclude a segment of the population that stands to gain the most from the internet. Most people do not intend to exclude people with disabilities. As organizations and designers become aware of and implement accessibility, they will ensure that their content can be accessed by a broader population.

This conclusion can be drawn from the review of literature where several groups within academe, that should be aware of accessibility issues, maintained Web sites with low levels of accessibility.

## REFERENCES

1.  La Porte, M., T., Demchak, C., C. &Friis, C. (2001) "Webbing Governance: Global Trends across National Level Public Agencies", Communications of the ACM, vol. 44, no. 1, pp. 63-67.

2.  Napoleon, E. , Asiimwe and Lim., N, " Usability of Government Websites in Uganda", Electronic Journal of e-Government Volume 8 Issue 1 2010, (pp1 - 12) ISSN 1479-439X.

3.  The World Wide Web Consortium (W3C), http://www.w3.org/

4.  S. Lujan-Mora, R. Navarrete, and M. Penafiel, "E-government and web accessibility in South America", in Proceedings of 2014 First International Conference on eDemocracy and eGovernment (ICEDEG), IEEE Computer Society, Washington, DC, pp. 77–82. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press, 2014.

5.  Y. Akgül, and K. Vatansever, "Web accessibility evaluation of government websites for people with disabilities in Turkey", Journal of Advanced Management Science, Vol.4 No. 3, pp. 201-210, 2016.

# Big Data Analytics: Current Scenario The Review

**Atul S. Patil**

*Assistant Professor, Dept. of Computer Science & I.T.
JET's ZulalBhilajiraoPatil College, Dhule
Maharashtra, India.*

atul_patils@yahoo.com

**Dr. Bhojraj Hanumant Barhate**

*Asso. Professor,
Dept. of Computer Science, P. O. Nahata College
Bhusawal*

bhbrama123@gmail.com

**ABSTRACT -** *As we know 'Big data' is a term applied to data sets whose size or type is beyond the capability of traditional RDBMS to hold, manage and process the data. It has characteristics like high volume, high velocity or high variety. Where, Data analytics has allowed several analyst, researchers, and businesses to make decisions accordingly, where various techniques are used to extract new insights from unexplored data sources. In this paper we have emphasised how various businesses are using data analytics to Boost Customer Acquisition and Retention, Solve Advertisers Problem and Offer Marketing Insights, Risk Management, Driver of Innovations and Product Development, Supply Chain Management etc. We have also emphasised which techniques are used now days and previous trends in last couple of years as well as future trends predictions, forecasts and challenges regarding big data analytics.*

*Keywords: Big Data, Data Analytics, IoT, Machine Learning, Artificial Intelligence, Predictive Analysis.*

## I. INTRODUCTION

Several papers have been published on 'Big Data Analytics' with various aspects of it. Here we have tried to compose the term 'Big Data Analytics' with current scenario and future predictions and challenges on one page. "Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. With today's technology, it's possible to analyse your data and get answers from it almost immediately – an effort that's slower and less efficient with more traditional business intelligence solutions."[2].

Even before today's hi-tech world, people were using basic analytics to uncover insights and trends manually. New analytical tools and technologies for data analytics have incorporated speed and efficiency in decision making accordingly.

## II. IMPORTANCE OF BIG DATA ANALYTICS

It helps organizations to work on their huge amount of unstructured, scattered data to extract innovative ideas for decision making accordingly.
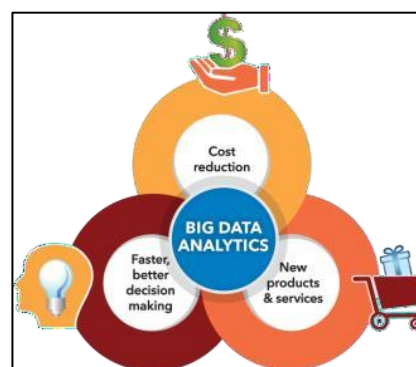


## III. BIG DATA ANALYTICS CAN EXTRACT FOLLOWING VALUES

**Cost reduction.** Big data technologies such as Hadoop and cloud- based analytics bring significant cost advantages when it comes to storing large amounts of data – plus they can identify more efficient ways of doing business.

**Faster, better decision making.** With the speed of Hadoop and in- memory analytics, combined with the ability to analyze new sources of data, businesses are able to analyze information immediately – and make decisions based on what they've learned.

**New products and services.** With the ability to gauge customer needs and satisfaction through analytics comes the power to give customers what they want. Davenport points out that with big data analytics, more companies are creating new products to meet customers' needs. [2].

## IV. REAL WORLD EXAMPLES HOW COMPANIES ARE USING BIG DATA ANALYTICS

- **Big Data Analytics to Boost Customer Acquisition and Retention**.

The customer is obviously the most important asset of any business. Every business focuses on having strong customer base, for that it's obvious to hold customer. But how it can be done? It can be done by observing various customer related patterns and trends. Here big data analytics allows business to trace customer's behaviour, current interest, prediction of his future interest, which in result helps to offer services and products accordingly. A real example of a company that uses big data analytics to drive customer retention is Coca-Cola. In the year 2015, Coca-Cola managed to strengthen its data strategy by building a digital-led loyalty program. Coca-Cola director of data strategy was interviewed by ADMA managing editor. The interview made it clear that big data analytics is strongly behind customer retention at Coca-Cola [3].

- **Big Data Analytics to Solve Advertisers Problem and Offer Marketing Insights.**

Marketing is the backbone of any business. It is the way customer can be attracted towards any service or product without having experience of the service or product. Here big data analytics can gain the insights of customers purchasing behaviour. This leads to more targeted campaign. "Netflix is a good example of a big brand that uses big data analytics for targeted advertising. With over 100 million subscribers, the company collects huge data, which is the key to achieving the industry status Netflix boosts. If you are a subscriber, you are familiar to how they send you suggestions of the next movie you should watch. Basically, this is done using your past search and watch data. This data is used to give them insights on what interests the subscriber most". [3]

- **Big Data Analytics for Risk Management.**

Big Data Analytics tools have emerged largely for providing solutions in Risk Management. Because of increasing availability and diversity of statistical data, big data analytics has a huge potential for increasing the quality of risk management models. This leads smarter risk analysis and management of business strategies accordingly. "UOB bank from Singapore is an example of a brand that uses big data to drive risk management. Being a financial institution, there is huge potential for incurring losses if risk management is not well thought of. UOB bank recently tested a risk management system that is based on big data. The big data risk management system enables the bank to reduce the calculation time of the value at risk. Initially, it took about 18 hours, but with the risk management system that uses big data, it only takes a few minutes. Through this initiative, the bank will possibly be able to carry out real-time risk analysis in the near future". [3]

- **Big Data Analytics as a Driver of Innovations and Product Development.**

By collecting huge amount of feedback data from customer and market and by using data analytics tools, businesses started updating or designing new product, services and re-designing the existing products and services according to insights got from analysis. "Amazon Fresh and Whole Foods. This is a perfect example of how big data can help improve innovation and product development. Amazon leverages big data analytics to move into a large market. The data-driven logistics gives Amazon the required expertise to enable creation and achievement of greater value. Focusing on big data analytics, Amazon whole foods is able to understand how customers buy groceries and how suppliers interact with the grocer. This data gives insights whenever there is need to implement further changes". [3]

- **Big Data analytics in Supply Chain Management.**

In every business, specifically manufacturing sector, agricultural or food sector, supply chain management is extremely important. Mismanagement in supply chain of products may lead to huge loss in context with time and money. Using application of big data analytics, suppliers get relevant insight through supply chains. Also suppliers are able tackle problems faced earlier. "PepsiCo is a consumer packaged goods company that relies on huge volumes of data for an efficient supply chain management. The company is committed to ensuring they replenish the retailers" shelves with appropriate volumes and types of products. The company's clients provide reports that include their warehouse inventory and the POS inventory to the company, and this data is used to reconcile and forecast the production and shipment needs. This way, the company ensures retailers have the right products, in the right volumes and at the right time [3].

## V. SOME OF BIG DATA ANALYTICS TOOLS IN 2020

As we know, huge amount of data is been generated through various sources in several forms every day. But this data is meaningless until it turns up into useful information or relevant intelligence to make decisions accordingly. Several data analytics tools are used for that. Some important open source and paid tools are enlisted here.

### 1. Apache Hadoop



It is the topmost big data analytics tool. Fortune 50 companies use Hadoop. Example: Amazon Web services, Hortonworks, IBM, Intel, Microsoft, Facebook, etc.

Apache Hadoop is a software framework employed for clustered file system and handling of big data. It processes datasets of big data by means of the MapReduce programming model. Hadoop is an open-source framework that is written in Java and it provides cross-platform support [4].

### 2. CDH (Cloudera Distribution for Hadoop)



CDH (Cloudera Distribution Hadoop) is open-source Apache Hadoop distribution provided by ClouderaInc which is a Palo Alto-based American enterprise software company. That also encompasses Apache Hadoop, Apache Spark, Apache Impala, and many more. It is used to collect, process, administer, manage, discover, model, and distribute unlimited data.

### 3. Cassandra.



It is open-source distributed NoSQL data base management system developed to handle huge amount of data spread across several commodity servers. It is the perfect choice when you need scalability and high availability without compromising performance. Some of top companies use it, just like Accenture, American Express, Facebook, General Electric, Honeywell, Yahoo, etc.

### 4. Knime



It is an open source tool that is used for Enterprise reporting, integration, research, CRM, data mining, data analytics, text mining, and business intelligence or predicting new futures. Some top companies using Knime include Comcast, Johnson & Johnson, Canadian Tire, etc. [4].

### 5. Datawrapper
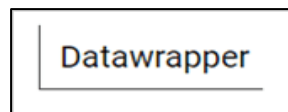


It is an open source platform to generate simple, precise and embeddable charts very quickly. Some companies just as, The Times, Fortune, Mother Jones, Bloomberg, Twitter etc use it. [4]

### 6. MongoDB



MongoDB is an object-oriented, simple, dynamic, and scalable open source NoSQL database. It is based on the NoSQL document store model. The data objects are stored as separate documents inside a collection, instead of storing the data into the columns and rows of a traditional relational database. Some of the major customers using MongoDB include Facebook, eBay, MetLife, Google, etc. [4]

### 7. Lumify



Lumify is a free and open source tool for big data fusion/integration, analytics, and visualization.

Its primary features include full-text search, 2D and 3D graph visualizations, automatic layouts, link analysis between graph entities, integration with mapping systems, geospatial analysis, and multimedia analysis, real-time collaboration through a set of projects or workspaces [4].

### 8. *HPCC*
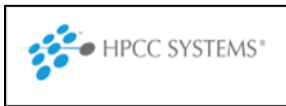


HPCC means High-Performance Computing Cluster. It is called as DAS i.e. Data Analytics supercomputer. It is based on a Thor architecture that supports data parallelism, pipeline parallelism, and system parallelism. It is an open-source tool and is a good substitute for Hadoop and some other big data platforms [4].

### 9. *Apache Storm*



It is a cross-platform, distributed stream processing and fault- tolerant real-time computational framework. Apache Storm is a free and open source distributed real-time computation system. Apache Storm makes it easy to reliably process unbounded streams of data, doing for real-time processing what Hadoop did for batch processing. Apache Storm is simple, can be used with any programming language, and is a lot of fun to use [5].

### 10. *Apache SAMOA*

Scalable Advanced Massive Online Analysis is an open source platform for big data streaming mining and machine learning. „It allows you to create distributed streaming machine learning (ML) algorithms and run them on multiple DSPEs (distributed stream processing engines) [4].

### 11. *Talend*



It is an open source data integration platform. It provides various software and services for data integration, data management, enterprise application integration, data quality, cloud storage and Big Data.

### 12. *Rapidminer*



Rapidminer is a cross-platform tool which offers an integrated environment for data science, machine learning and predictive analytics. Organizations like Hitachi, BMW, Samsung, Airbus, etc have been using RapidMiner [4].

### 13. *Qubole*



It is a multi-cloud data platform for Data Engineering, Analytics, ML, & Analytics for Apache Spark, Hive, and Presto. Qubole data service is an independent and all-inclusive big data platform that manages learns and optimizes on its own from your usage. Warner music group, Adobe, and Gannett are some companies that use it [4].

### 14. *Tableau*



Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It helps in simplifying raw data into the very easily understandable format. Data analysis is very fast with Tableau and the visualizations created are in the form of dashboards and worksheets.

### 15. *R*



R is one of the most comprehensive statistical analysis packages. It is open-source, free, multi-paradigm and dynamic software environment. It is broadly used by statisticians and data miners. Its use cases include data analysis, data manipulation, calculation, and graphical display [4].

## VI. LAST COUPLE OF YEARS WITH BIG DATA ANALYTICS

1. We have seen huge upticks in Big Data use far and wide, with associations rushing to get a handle on the importance of data undertakings and coordination to their business accomplishment.

2. We began seeing this pattern toward the finish of 2018, yet in 2019, the operationalization of big data turned out to be considerably more feasible no matter how you look at it. Already,

organizations could see beginning accomplishment with data operationalization, however scaling data operations and orchestration demonstrated tedious and hard to keep up.

3. Another key data analytics pattern for 2019 was the expanded arrangement between conventional analytics with machine learning (ML) and artificial intelligence (AI) analytics. An ever increasing number of associations are utilizing ML and AI to enlarge ordinary operational analytics pipelines and typical line of business exercises.

## VII. DATA ANALYTICS MARKET, TRENDS, PREDICTIONS FORECASTS AND CHALLENGES

As data sources become increasingly complicated and AI applications grow, coming years are set to be the years of innovation and advancement for big data. And global Data Analytics Market is been continuously growing strongly and it is expected to do so.

1. 'A recent market study shows that the Data Analytics Market is expected to grow at a CAGR of 30.08% from 2020 to 2023, which would equate to $77.6 billion. In 2019, the global analytics market was $49 billion worldwide, which is more than double the value four years ago.' [7]

'The global big data and business analytics market was valued at $168.8 billion in 2018 and is forecasted to grow up to $274.3 billion by 2022 at a CAGR of 13.2%.' [7]
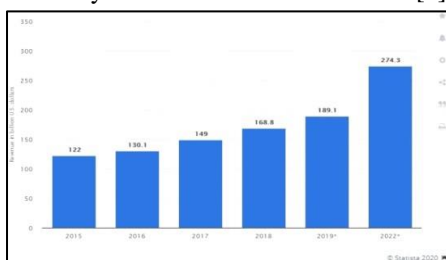


2. Rapid increase in huge data generation by IoT devices has explored an entirely new field of research.

The IoT analytics market has been valued at $17.1 billion in 2019 and is expected to grow at a CAGR of 29.8% between 2020 and 2025 and reach $81.67 billion by 2025. Ericson's mobility report states that there would be 1.5 billion IoT devices with cellular connections in 2022'. [7]
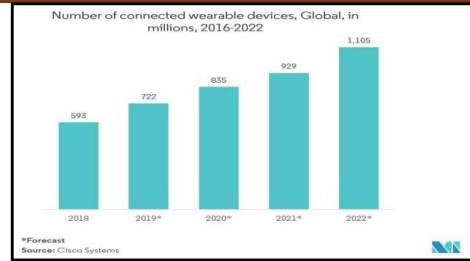


'According to Statista, by 2021, the global cloud data center IP traffic is predicted to reach around 19.5 zettabytes. Connected IoT devices are projected to generate 79.4 zettabytes of data in 2025.' [7]
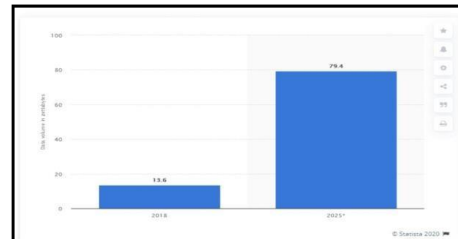


'The size of the business intelligence and analytics software application market is forecasted to reach around$14.5 billion in 2022'.[7]
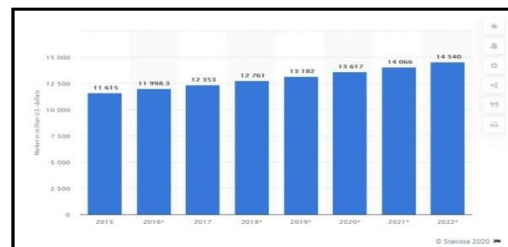


3. 'Machine learning and AI are undoubtedly two of the biggest buzzwords of the last few year'. 'By 2022, the predictive analytics market is expected to reach nearly $11 billion in annual revenue'. [7] As business transactions around the globe has become progressively digitized, huge amount of data generated and can be accessed through predictive analytics tools to give clients a superior comprehension of market element and basic patterns, trends as well.

4. In 2020 onwards, machine learning and AI will be the main players with regards to profiling and focusing on clients, with information assuming an essential role in showcasing and business transactions. Data will be gathered in high volumes, and physical stockpiling will be steadily progressively costly to keep up. Cloud computing will soar in 2020 and in the next years, which will give organizations the capacity to grow so a lot and as quick as they need without being confined by physical infrastructure

## VIII. THOUGHTS OF BIG DATA AND DATA ENGINEERING INDUSTRY VETERANS, RAMESH MENON AND TODD GOLDMAN

- *'The Cloud is the new Data Lake… but Multi-cloud and Hybrid are here to stay'. [6]*

As cloud-based technologies continue to develop, businesses are increasingly likely to desire a spot in the cloud. In 2020 onwards, it is expected to see most of will bring hybrid and multi cloud methodologies in existence.

- *'More Businesses Will Abandon Hadoop for Spark and Databricks, But This Change Won't Solve All Their Problems.' [6]*

Hadoop has been criticized by many in the community for its complexity. Spark and managed Spark solutions like Databricks are the "new and shiny" player and have therefore been gaining traction as data science workers see the platform as an answer to everything they dislike about Hadoop. Spark and Databricks will be especially lauded for their interactive processing capabilities, as well as its internal memory computation for job scheduling and user- friendly interface, which allows data scientists to process stored data via high-level operators." [6]

- *'Digital Transformation Will Be a Key Component of Top-Level Data Strategies' [6]*

Digital transformation is extremely about adopting a data- driven strategy to each part of the business with an end goal to make a competitive benefit.

- *'Machine Learning and Artificial Intelligence Will Continue to Evolve' [6]*

In 2020, automation structures will permit data researchers to make their own data pipelines. This mix of bringing data engineering to data researchers and data science to data analyst will drive an expansion in the quantity of genuine ML and AI algorithms that will go into big business level production.

## IX. CONCLUSION

Big data analytics has been back bone for decision making and making strategies as per the then scenario, near about in each and every business in last couple of years, using ample advanced tools and techniques available with scalability and predictive analysis. In present and future it is going to be the only way to implement business transactions and strategies with heavy use of data analytics using Machine learning and Artificial Intelligence for all sectors to achieve desired outcome in the business.

### REFERENCES

1. https://www.ibm.com/in-en/analytics/hadoop/big-data-analytics
2. https://www.sas.com/en_in/insights/analytics/big-data-analytics.html#nextsteps
3. https://www.mentionlytics.com/blog/5-real-world-examples-of-how-brands-are-using-big-data-analytics/
4. https://www.softwaretestinghelp.com/big-data-tools/
5. https://storm.apache.org/
6. https://www.infoworks.io/big-data-trends/
7. https://www.cognetik.com/blog/data-analytics-market-in- 2020-trends-forecasts-challenges

# Comparative Study of Various Steganography Techniques: A Literature Review

**Yogesh Narayan Chaudahri**

*Assistant Professor*

*KCES's Institute of Management and Research, Jalgaon.*

chaudhariyogn@gmail.com

**Dr. Bhojraj Hanumant Barhate**

*Associate Professor, Vice-Principal, 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 paperexplores digital steganography and its techniques that primarily focus on text steganography. It also present the comparative study of various techniques and its classification. Three types of steganography classifications were discussed, that consist of the type of carrier file, natural key used, and the embedding techniques. Text steganography can then be further separated into three categories: format based methods, random and statistical generation, and linguistics method. Techniques belonging to each category were studied, and comparisons between each technique are introduced by highlighting the findings. This study also confirmed that there are three principal factors that need to be further explored and taken into account in the design of future steganographicsystems, which are the capacity, high transparency, and security.*

*Keywords - Information security, Information hiding, Steganography, Text steganography, Format based method Random and Statistical Method, Linguistics Method.*

## I. INTRODUCTION

In recent trends in the world, the communication is the basic necessity of every growing area. Everyone wants the secure and safecommunication. Information security is a major issue of concern while exchanging a data in an open network, as internet is not only a single network it is worldwide collection of loosely network. The network security is becoming more important as the volume of data being exchanged over the Internet increases day by day.(Arya & Soni, n.d.)

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, protocoletc.

Terminology used in steganography are(Sharma et al., 2016):

1) Cover Object: Text, audio, video, image used for embedding data is known as cover object. 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 resultant output obtained after embedding which is known as stego object.

In general, steganography method masks a secret message in a cover medium. A stegomedium is produced as a result of masking secret message in cover medium. The function that is used to hide secret message in cover medium is called embedding procedure. Extraction procedure retrieves the hidden secret message from stego medium. Important requirements of a steganography method are imperceptibility, robustness and capacity. Imperceptibility is a measure of security of steganography method that determines the level of perception of existence of a secret message in a cover file. Therefore, the primary goals of steganographic approaches are high embedding capacity and low distortion of the cover medium(Rahman et al., 2017).

In order to safely broadcast of confidential data, the multimedia object like audio, video, images are used as a cover sources to hide the data.

## II. TYPES OF STEGANOGRAPHY

- Text Steganography
- Image Steganography
- Audio Steganography
- Video Steganography

The Steganography method uses the text media to skin the data known as text Steganography. There are different methods to embed the secret data in text files.

- Format Based Method
- Random and Statistical Method
- Linguistics Method

Following literature review performed to collect information of text steganography

## III. FORMAT BASED METHOD

(Roy & Venkateswaran, 2013) This paper presents a different method to the English text based steganography with Indian root. In the propose method, no properties of a sentence are used rather features of the English language is used. This gives flexibility and independence from the point view of the sentence construction but it increases computational complexity. Author proposed a technique to hide a 4 letter word, 8 words are required excluding the words that are added to provide

flexibility in sentence creation. So to hide a longer message, this technique requires large no of words and creates a complexity in sentence construction. In combination with other cryptography technique like visual cryptography, the proposed text steganography technique can provide a two layer of authentication and security system in physical and online banking as well as online shopping as shown in section information security.

(Por et al., 2012) A text-based data beating method called UniSpaCh was proposed to embed payload into Microsoft Word document. A simple attack based on the show/hide format- ting mark named DASH was proposed to reveal the presence of embedded information in the conventional methods. Unicode space characters that remained invisible with respect to DASH were utilized together with the normal space character to encode payload using inter-word, inter-sentence, end-of-line and paragraph spacing's. Results confirmed that UniSpaCh is robust with respect to DASH attack while the existence of externally embedded information was instantly exposed in the case of the conventional methods. UniSpaCh offers higher embedding effectiveness when compared to the existing methods. In the best case scenario, UniSpaCh produces output document of size almost 9 times lesser than that of the existing method. Results confirmed that content of the cover document has insignificant effect on embedding efficiency of UniSpaCh.

(Odeh et al., 2012)presents a new text Steganography for Arabic multipoint letters. The new process deals with two bits for each multipoint letter. Author combine our approach with vertical point shifting to improve the amount of hidden data. The retyping process is ainspiring problem for similar algorithms which removes all the hidden data. Author solve this challenge to mitigate any new font format changes by unifying all data which leads a similar file. Finally, the result reported by this implementation has outperformed comparable results reported in literature in terms of the hiding capacity and the possibility of having such steganography mechanism used in hiding information.

(Bhaya et al., 2013) This study proposed a new method of hiding information in Microsoft Word documents. Microsoft Word documents are very much common in normal life of today's digital world. The capability of this method is very high, depending

on the number of Capital Letters in cover document. As show in research paper some fonts take large size when replace it with their comparison such as (Arial) font type and some fonts are not, such as (Lucida Sans, Century) font types. Because the stego document will not modify during compression, copying and paste between computer programs, the data hidden in texts remains intact during these operations.

(Mahato et al., 2014) In this paper, text steganography technique for information hiding is presented. The paper proposes a new approach to provide data safety by hiding secret bit information in Microsoft Word document. As Microsoft Word is very common to all computer users, this can be used as a cover medium to perform text steganography very simply. Secret data are hidden behind the font size of the invisible space character, and hence, document is generated. Similarly, using the reverse process on the stego document, the secret message is easily improved by the intended receiver. The proposed algorithm's embedding capability is very high and can be increased more by putting more space character in between two words. Although this paper is based on word document, it can be expanded to other file formats such as MS-Excel, HTML, XML, CSS, and PDF, etc. Future research can be performed to reduce the time complexity of the algorithm.

(Stojanov et al., 2014) Property Coding is not entirely appropriate for copyright protection applications where robust data-hiding is required, because the attacker can always use Optical Character Recognition (OCR) to completely remove the secrete data. Four new format based methods specially selected for hiding data in MS-Words documents are given. Because they change the properties of some document objects offered by MS-Word, we called the new type of technique Property Coding. These methods are resistant to saving actions, presents very small overhead on the document size, and can embed up to 8 bits per character.

(Mohamed, 2014) This paper presents a new text steganography technique for Arabic texts or any languages written in Arabic letters like Urdu (the official language of Pakistan), Persian (the official language of Iran) and Pashto (the official language of Afghan- istan) based on the features of Arabic text which offers high capacity, more safety, and good robustness (i.e. we introduce a compromised solution

to our problem). Experimental results show the following benefits of the suggested algorithm: The proposed algorithm capacity ratio is nearly equal to 52 bit/kB which is high and the embedding size of our method can be altered flexibly according to the carrier text size. The proposed algorithm skins information with mini- mum change in the cover text hence the stego-texts gen- erated by the proposed algorithm will draw less attention (i.e. It satisfies perceptual transparency).

(Roslan et al., 2014) Therefore, it can be concluded that this hybrid method of ATS can decrease the level of vulnerability while enhancing the robustness of the authentication system. Further, a high level of user confirmation can also be achieved with 0% for both ferrfake and ferrlive of the liveness detection for the fingerprint owner. By taking advantage of the independent credentials of people, namely their fingerprint and heartbeat as a defence mechanism against data spoofing, this method will provide superior text steganography in data sharing. Similarly, the CPM presents a huge potential for hiding secret informations where aextreme of 12 bits of secret information per character can be hidden in each Arabic character. Our experiment compared two distinct techniques for hiding secret information through the random positioning of the basic shape groups and the extreme number of potential hiding places. Both techniques have benefits and difficulties. Since we were more concerned with optimising the embedded secret information capability, our CPM with maximum positioning resulted in 23.5% more optimising compared with the previous method.

Summary of literature review based on Format based method

| Author | Techniques | Finding |
|---|---|---|
| (Roy & Venkate swaran, 2013) | Special Character, Line Shifting, and Word shifting coding techniques. | Provides good beating capacity and no change in stego file, but with low robustness. |
| (Por et al., 2012) | Modifying inter-sentence, inter-word, end of the line and inter-paragraph spacing. | Higher embedding effectiveness. Robust on DASH attack. Also, the contents of the cover file have minimal effect on the embedding efficiency. |

| Author | Techniques | Finding |
|---|---|---|
| (Odeh et al., 2012) | Multi-point letters | Improve the capacity and robustness. |
| (Bhaya et al., 2013) | Capitalizing selected letters in the cover media and changed font type. | The capability is very high and has good transparency. However, stego document increased by nearly 0.766% from the original size. |
| (Mahato et al., 2014) | Changing font size of invisible letter space. | Very high capability, but low robustness when the attacker changing the font size. |
| Stojanov et al., 2014 | Property Coding: character scale or underline and the border of paragraph and sentence. | High capability, but low robustness and slight increase in the size of stego-file of nearly 0.1% |
| (Mohamed, 2014) | Using single letters without any noticeable change in the target word. | High capability; the embedding capacity rate ratio is also high. Resist traditional attacking methods. |
| Roslan et al., 2014 | Primitive structure; Sharp edges, dots, the typographical proportion of the Arabic letter. | Higher capability and higher transparency, but low security. |

## IV.   RANDOM AND STATISTICAL GENERATOR

(Ryabko & Ryabko, 2011) In this work we take an information-theoretic method of steganography, and construct perfectly secure steganographic systems for the case of finite-memory sources of cover text. We also show that some (probabilistic) assumptions on the sources of cover texts are necessary, by demonstrating some principled computational limits on steganography that arise in the absence of such assumptions. Perhaps the first information-theoretic method of steganography was proposed author, who modelled the sequence of cover text by a memory less distribution. Besides laying out basic definitions of steganographic protocols and their security, author has constructed a steganographic protocol, which, relying on the fact that the

probability distribution of cover texts is known, assures that the distributions of cover texts with and without hidden data are statistically close (but, in general, are not equal).

(Dulera et al., 2011) In this paper, author propose newer approaches for text-based steganography for English language texts. In these approaches, we exploit the shapes of the English characters to skin secret bits. We can skin more number of bytes using proposed approaches, while additional time overhead and memory overhead is also minimal compared to feature coding method. These approaches are related to the soft-copy texts as well as hard-copy texts. Our analysis discloses that our approaches impart increased randomness and because of randomness, these approaches are noticeable but it cannot be decoded until a user is not aware about these approaches. In addition, the proposed approaches are also protected to retyping and reformatting of text. However, one of the dimness of the proposed approaches is that once their applicability is known, they can easily be attacked. Hence, it is necessary to keep the application of a particular approach to a particular data set secret.

(Majumder & Changder, 2013) In this paper we have introduced a novel approach for text steganography by creating the text summery by using the reflection symmetry of the alphabets of English language. To skin secret data bits, the proposed method checks the vertical and horizontal reflection symmetry properties of the characters present in each sentence of the text and, if followed, it selects the sentence to generate a summary of the text. The benefit of the proposed algorithm is that, since we are hiding the data without changing the structure of the file and by creating the summery of a publicly available text like newspaper article, therefore it will draw less attention to the unintended recipients and hence more safety is added to the proposed steganographic system. While generating the text summery, as our method provides a much semantically correct output, associated with Microsoft Words auto Summarize Tool, it will draw much less suspicion and this may be considered as a great advantage of the system. Also, as there is no limit to the size of the text file, we can skin a large volume of data using the proposed algorithm. However the drawback of the system lies in the fact that once the applicability of the system is known, it can be simply

attacked. So, it is urgent to keep the application of the method to a specific dataset, secret.

(Satir & Isik, 2012) Researcher aim to explain the benefits and drawbacks of the proposed method. A benefit of the proposed method is not being language specific. The method can be applied to any language by reconstructing the text database and adapting the Latin Square to the concerning language, if necessary (for e.g. Chinese and Arabic languages). Another benefit of the proposed method is keeping the originality of the cover media while communicating. The method does not produce noise in order to skin secret information. It modifies neither meaning nor format of the cover text. In the proposed method, the stego cover is a forward mail platform that comprises two cover medium. One of them is the naturally created cover text. So the text is meaningful, syntactically and grammatically accurate and legitimate.

(Tutuncu & Hassan, 2015) Proposed method is extension of text-based steganography that makes use of email environment to skin the secret text. Capacity ratio and complexity of the proposed method are improvements of the previous e-mail environment based text steganography methods [1-4] in terms of complexity. Reasonable capability was obtained. By using combination of Run Length Encoding (RLE) + Burrows Wheeler Transform (BWT) + Move to Front (MTF) + Run Length Encoding (RLE) + Arithmetic Encoding (ARI) algorithms the size of the secret message was compact. Vigenere cipher added extra complexity or safety to the system for obtaining stego key (K1). The study also presented that the length and the content (variation) of the cover text directly related with the success of the embedding process. Small cover text with not well distributed alphabet content (doesn't match with the frequency of the English alphabet) will result failure in embedding process.

(Malik et al., 2017)In this paper, we have proposed a new technique for text steganography that uses LZW compression and color coding approach for beating the secret data in the forward mail platform. There are several benefits of the proposed method. Firstly, it performs better in term of computational complexity as there is no need to use mapping operation, this saves the time essential to process mapping operation. Secondly, the employment of LZW directly on the secret data reduced the chances of overhead generation thereby increasing the embedding capacity. To further increase the capability, colors are used in the cover text or message of the email to hide some part of the secret data bit stream.

Summary of literature review based on and statistical generator

| Author | Techniques | Finding |
|--------|-----------|---------|
| (Ryabko & Ryabko, 2011) | Carriers (containers) are made via a finite memory source with possibly unknown statistics. | High security, but the rate of transfer of secret message is not optimal and it uses only two (i.i.d) carrier. |
| (Dulera et al., 2011) | Random character sequencing and feature coding methods | Increased randomness, thus aiding in higher security at minor overhead. |
| (Majumder & Changder, 2013) | Generating the summary of a textual file. | Higher capacity and low security. |
| (Satir & Isik, 2012) | LZW data compression algorithm. | Provides a important increment with regard to capacity. |
| (Tutuncu & Hassan, 2015) | A hybrid of Run Length Encoding, Burrows-Wheeler Transform, Move to Front (MTF), Run Length Encoding, and Arithmetic Encoding lossless compression algorithms sequence. | Achieved the hiding capability and also improved the security. |
| (Malik et al., 2017) | Using the number of characters used in the email id to indicate the hidden secret data bits. | Improved the hiding capability compared to some relevant existing methods. |

## V. LINGUISTIC METHOD

(Mir & Hussain, 2011)In this paper, Author have presented Text Steganography combined with Cryptography for beating secret information using XML files. Nine different embedding techniques considered and applied on XML file. System has been implemented using C# language for all nine methods combined with AES which has added additional layer of security. All methods are measured with respect to different standards and it is analysed that white space method, white space replacement method, colour replacement method, line break method synonyms method and acronyms methods are considered stronger and less susceptible. Furthermore, improvements in colour, synonyms and acronyms are needed to make them more practical, effective and robust. Techniques discussed in this paper have therefore been applied on textual information and hence could also be applied on other types of data in XML files, as XML does not only contain text but multimedia based data as well and the idea could be extended toward other parts. These Embedding techniques can also be useful on other web tools like scripting languages. Moreover, a bandwidth or capacity comparison between mark-up languages and scripting languages can also be obtained.

(Shu et al., 2011) In this paper we have proposed a new hiding algorithm model, which is effective for beating information in text, especially in plain text; hidden data could be distributed effectively through the decomposition time after time and then embedded through the multiple selections of the carrier text. In this way, the disguise of the hidden information text could be enhanced effectively. Meanwhile, comparable to the general algorithms based on semantics, this algorithm can resist the re-layout of the carrier text, modification, deletion of the key texts and other operations as well, so with higher robustness. In addition, the algorithm allocates the secret information in space, thus could increase the complexity of information beating. In this way, getting any one part of information is meaningless, so that the security of the algorithm could be further improved. Meanwhile, the new algorithm has the advantage of the usual algorithm based on semantics, that is, battle to the re-layout of the carrier text.

(Vidhya & Paul, 2015) Steganography is the method of beating information such that its presence cannot be detected. A secret message is encrypted in such a manner that the existence of the information is hidden. We have offered a method to text steganography with an Indian local language, Malayalam. The proposed method consists of a custom Unicode based method with embedding based on indexing. After that an embedding algorithm will be designed to mix the encrypted original message with the Malayalam text. The experimental study is done to evaluate the effectiveness of the proposed approach. The comparison study of the proposed method against an existing method exposed that, the proposed steganography methods is more precise in the encoding process and balanced in the decoding process. The proposed method completed a precision rate of .95 and decoding rate of .81.

(Shivani et al., 2015)Transfer confidential data over internet is risky task. Primary concern is to keep data from intruders. As data is transmitted via digital medium, it has certain drawbacks like tampering, easy to access, illegal use, copyright violation etc. Information hiding is a branch of computer science which deals with beating data, object or function details. Steganography is an art of beating message in such a fashion, presence of hidden message cannot be conceived. Several researchers are working in this area to increase the efficiency of steganographic algorithms. Proposed algorithm in this paper improves the data hiding capability. With the help of proposed algorithm users can skin more amount of data without producing any bias in the cover image. It means changes reflected are almost insignificant.

(Lee & Chen, 2013)Author presents a new text steganography scheme in compression domain using a lossless compression coding which called variable Huffman coding. The secret message is embedded into compression codes. The goal of the proposed scheme is convert communication by means of text files, providing high embedding capacity, improving security of the embedded secret message and reducing broadcast cost. Additionally, the original text files can be reconstructed without any bias after the embedded secret message is extracted. In the proposed scheme, each leaf of variable Huffman tree can be used to send a secret bit at least. According to the practical application, the embedding rate for each leaf is to be improved, i.e, the embedding capacity of the proposed scheme is scalability. The secret keys are employed to generate the stego-compression-code for each leaf of variable Huffman tree to keep the embedded message.

(Chang & Clark, 2014) One of the aids of this work is to develop a novel lexical substitution-based stego system using vertex coding that improves the data embedding capacity com- pared to existing systems. The vertex coding method signifies synonym substitution as a synonym graph so the relations between words can be clearly observed. In addition, the NGM method, an automatic system for checking synonym acceptability in context, is integrated in our stego system to ensure data security. In this work, we only evaluated the lexical substitution in terms of the sentence-level naturalness rather than meaning retention and document-level consistency. There- fore, it would be interesting to see to what extent the proposed substitution checkers are useful for the safety of linguistic steganography at the document-level

(Qi et al., 2013) In traditional methods, the redundant synonyms are always wild. Two different coding schedules, which made full use of the wild synonym to improve the security of the steganographic system, are proposed in this paper. The synonyms in a synonym set are dynamically separated into coded words and backup words. One method divides the cover text into blocks to skin the secret information for the first time and then uses a highly efficient coding schedule to embed the secret information for the second and third time. Two novel method are proposed to take benefit of the wild synonym in traditional steganography based on synonym substitution. The first method increases the steganographic capacity by 44% than traditional method and the second method can avoid the steganalysis using the feature derived from synonym pairs.

Summary of literature review based on Linguistic method

| Author | Techniques | Finding |
| --- | --- | --- |
| (Mir & Hussain, 2011) | AES algorithm and synonyms | Increases the security. |
| (Shu et al., 2011) | An algorithm based on multi-text. | Increases the security and higher robustness. |
| (Vidhya & Paul, 2015) | Unicode extraction and diagonal encoding indexing. | Achieved greater security. |

| Author | Techniques | Finding |
| --- | --- | --- |
| (Shivani et al., 2015) | Abbreviation method and Zero Distortion Technique | Increase security and increases data hiding capacity. |
| (Lee & Chen, 2013) | Used a lossless compression coding, which termed variable Huffman coding. | High embedding capacity, reduced broadcast cost and also increasing the security. |
| (Chang &Clark, 2014) | Replacing selected words with the same part of speech synonyms. | Achieved a small capacity within a reasonable level of safety. |
| (Qi et al., 2013) | Employ the benefit of abandoned synonym in traditional steganography based synonym substitution. | Achieved higher capacity, robustness and a least creating syntax error in English text. |

## REFERENCES

1. Arya, A., & Soni, S. (n.d.). A Literature Review on Various Recent Steganography Techniques. International Journal on Future Revolution in Computer Science & Communication Engineering, 1–143. Retrieved July 20, 2021, from http://www.ijfrcsce.org

2. Bhaya, W., Rahma, A. M., & AL-Nasrawi, D. (2013). Text steganography based on font type in MS-word documents. Journal of Computer Science, 9(7), 898–904. https://doi.org/10.3844/jcssp.2013.898.904

3. Chang, C. Y., & Clark, S. (2014). Practical Linguistic Steganography using Contextual Synonym Substitution and a Novel Vertex Coding Method. Computational Linguistics, 40(2), 403–448. https://doi.org/10.1162/COLI_a_00176

4. Dulera, S., Jinwala, D., & Dasgupta, A. (2011). Experimenting with the Novel Approaches in Text Steganography. International Journal of Network Security & Its Applications, 3(6), 213–225. https://doi.org/10.5121/ijnsa.2011.3616

5. Lee, C. F., & Chen, H. L. (2013). Lossless Text Steganography in Compression Coding. Intelligent Systems Reference Library, 40, 155–179. https://doi.org/10.1007/978-3-642-28580-6_8

6. Mahato, S., Yadav, D. K., & Khan, D. A. (2014). A novel approach to text steganography using font size of invisible space characters in microsoft word document. Advances in Intelligent Systems and Computing, 243, 1047–1054. https://doi.org/10.1007/978-81-322-1665-0_107

7. Majumder, A., & Changder, S. (2013). A Novel Approach for Text Steganography: Generating Text Summary Using Reflection Symmetry. Procedia Technology, 10, 112–120. https://doi.org/10.1016/j.protcy.2013.12.343

8. Malik, A., Sikka, G., & Verma, H. K. (2017). A high capacity text steganography scheme based on LZW compression and color coding. Engineering Science and Technology, an International Journal, 20(1), 72–79. https://doi.org/10.1016/j.jestch.2016.06.005

9. Mir, N., & Hussain, S. A. (2011). Secure web-based communication. Procedia Computer Science, 3, 556–562. https://doi.org/10.1016/j.procs.2010.12.092

10. Mohamed, A. A. (2014). An improved algorithm for information hiding based on features of Arabic text: A Unicode approach. Egyptian Informatics Journal, 15(2), 79–87. https://doi.org/10.1016/j.eij.2014.04.002

11. Odeh, A., Alzubi, A., Hani, Q. B., &Elleithy, K. (2012). Steganography by multipoint Arabic letters. 2012 IEEE Long Island Systems, Applications and Technology Conference (LISAT), 1–7. https://doi.org/10.1109/LISAT.2012.6223209

12. Por, L. Y., Wong, K., & Chee, K. O. (2012). UniSpaCh: A text-based data hiding method using Unicode space characters. Journal of Systems and Software, 85(5), 1075–1082. https://doi.org/10.1016/j.jss.2011.12.023

13. Qi, C., Sun, X., & Xiang, L. (2013). A secure text steganography based on synonym substitution. 2013 IEEE Conference Anthology, ANTHOLOGY 2013. https://doi.org/10.1109/ANTHOLOGY.2013.6784896

14. Rahman, M. S., Khalil, I., Yi, X., & Dong, H. (2017). Highly Imperceptible and Reversible Text Steganography Using Invisible Character based Codeword Subjective Evaluation of Market-Driven Cloud Services View project. https://www.researchgate.net/publication/320192946

15. Roslan, N. A., Mahmod, R., Udzir, N. I., & Zurkarnain, Z. A. (2014). Primitive structural method for high capacity text steganography. Journal of Theoretical and Applied Information Technology, 67(2), 373–383.

16. Roy, S., & Venkateswaran, P. (2013). A Text based Steganography Technique with Indian Root. Procedia Technology, 10, 167–171. https://doi.org/10.1016/j.protcy.2013.12.349

17. Ryabko, B., & Ryabko, D. (2011). Constructing perfect steganographic systems. Information and Computation, 209(9), 1223–1230. https://doi.org/10.1016/j.ic.2011.06.004

18. Satir, E., & Isik, H. (2012). A compression-based text steganography method. Journal of Systems and Software, 85(10), 2385–2394. https://doi.org/10.1016/j.jss.2012.05.027

19. Sharma, S., Gupta, A., Trivedi, M. C., & Yadav, V. K. (2016). Analysis of different text steganography techniques: A survey. Proceedings - 2016 2nd International Conference on Computational Intelligence and Communication Technology, CICT 2016, 130–133. https://doi.org/10.1109/CICT.2016.34

20. Shivani, Yadav, V. K., & Batham, S. (2015). A Novel Approach of Bulk Data Hiding using Text Steganography. Procedia Computer Science, 57, 1401–1410. https://doi.org/10.1016/j.procs.2015.07.457

21. Shu, Y., Liu, L., Tian, W., Miao, X., & Wei-na, T. (2011). Algorithm for Information Hiding in Optional multi-Text. Procedia Engineering, 15, 3936–3941. https://doi.org/10.1016/j.proeng.2011.08.737

22. Stojanov, I., Mileva, A., & Stojanović, I. (2014). A new property coding in text steganography of Microsoft Word documents. In SECURWARE 2014 - 8th International Conference on Emerging Security Information, Systems and Technologies.

23. Tutuncu, K., & Hassan, A. A. (2015). New Approach in E-mail Based Text Steganography. International Journal of Intelligent Systems and Applications in Engineering, 3(2), 54. https://doi.org/10.18201/ijisae.05687

24. Vidhya, P. M., & Paul, V. (2015). A method for text steganography using malayalam text. Procedia Computer Science, 46, 524–531. https://doi.org/10.1016/j.procs.2015.02.079

Paper ID: CS-16

# Classifying Fetal Health with Cardiotocography Data Using Machine Learning Techniques

**Mr. Anand J. Maheshwari**

*Assistant professor, R.C.Patel Arts, Commerce, and Science. College, Shirpur, India,*

*a.j.maheshwari39@gmail.com*

**Dr. G. R. Wani**

*Associate professor, BASPONC College, Bhusawal, India*

*gajuwani03@gmail.com*

**ABSTRACT** - *Obstetricians commonly utilize cardiotocography (CTG) to assess the corporal state of an infant during childbirth because it offers information on the fetal heartbeat and vaginal cramps, which takes into account whether or not the fetus is malignant. Based on signal analysis and computer engineering, there are numerous approaches for interpreting conventional cardiotocography data. Machine learning is now commonly used to solve a wide range of problems in branches of science and healthcaredue to its rapid development. For fetal state classification, we apply additive logistic regression (Logiboost), k-NN and random space classifiers to analyze CTG data; Logiboost outperforms k-NN and random subspace in process of reaching accuracy in classifying specific varieties of fetuses up to 92.16 on training and testing set. Both Logiboost and random subspace performs equally and outperform K-NN in feature selection.*

*Keywords -cardiotocography machine learning, splitting. Accuracy,Fetal Health.*

## I. INTRODUCTION

In the field of medicine, novel engineering solutions have shown to be quite useful in assisting doctors in achieving their intended objectives. Without engineering technology assistance, the medical sector cannot function efficiently, live, or improve its existence, as can be seen if one goes back far enough. Clinical research aims to better comprehend the inner workings of the human body[1].

Reduced child mortality is an important indicator of human progress, and it is reflected in several of the United Nations' Development Agenda. The UN expects countries to have eliminated preventable deaths of newborns and kids under the age of five by 2030, with all countries aiming to reduce under-five mortality towards less than 25 per infant mortality rate.

Infant mortality, which accounts for million adult American deaths during and after pregnancy and childbirth, is related to child mortality (as of 2017). The vast majority of these deaths (94%) transpired in low-resource areas, and the overwhelming majority could have been prevented.

Cardiotocograms (CTGs) are a simple and inexpensive way for healthcare providers to examine fetal health and take measures to reduce infant and maternal mortality. The machine sends ultrasound pulses and reads the response, providing information on the fetal heart rate (FHR), fetal movements, uterine contractions, and more.

As per World Federation of Obstetricians and Gynecologists (FIGO) recommendations, a cardiotocograph (CTG) should be described and evaluated as a uncertain or abnormal based on fetal heart, heart rate fluctuations, accelerations and decelerations, Knowledgeable medical professionals can bring out such a perception (e.g., osteopathic physicians) or machine learning software. Because the

trials were of a moderate complexity [2]. Machine intelligence employs a specific methodology and to treat a patient, diverse data observations from the female organism must be collected. Existing algorithms are very good at predicting the pathological state of the unborn child, but not quite so decent at forecasting the potentially malicious state.

The study's purpose was to bring a machine learning model capable of accurately identifying high pregnancies.

## Objectives and problem statement

We apply machine learning to predict fetus health in order to reduce infant and maternal mortality in this study. The intention of this data analysis was to evolve a framework of machine learning that could clearly categorize high-risk pregnancies.

1. Task is to train the classifier and test the efficiency, effectiveness of algorithms

2. Classify fetal health to prevent child and maternal mortality and evaluate performance of classifiers.

## Problem Statement

Using the Weka framework, create a machine learning model that can assess fetal health conditions. Make a model to classify the CTG observations.

## Related Work

1. Zahra Hoodbhoy, Mohammad Noman, 1 Ayesha Shafique, 1 Ali Nasim, 1Devyani Chowdhury2, and Babar Hasan, "Use of Machine Learning Algorithms for Prediction of Fetal Risk Using Cardiotocographic Data" online Use of Machine Learning Algorithms for Prediction of fetus risk (nih.gov), They use the CTG dataset to test 10 algorithms and propose one that is highly accurate on both suspect and pathological cases.

2. J. Jayashree, Harsha T, Anil Kumar C, and J. Vijayashree, "Enhanced Optimal Feature Selection Techniques for Fetal Risk Prediction Using Machine Learning Algorithms", International Journal of Engineering and Advanced Technology (IJEAT), investigates the Fetal Risk Prediction using MRMR

Feature Selection algorithm on four different Classifiers.

3. Gongao Xue's paper "The Application of Machine Learning Models in Fetal State Auto-Classification Based on Cardiotocograms", presented at GBEM2019, Four models are researched all across the entire process.

## Methodology

A classification model that we describe could be a wonderful and efficient complement to already accessible tools like diagnosis, as it monitors specific features of essential tests for pregnant women. The majority of academics are interested in fetal-health data in order to develop a model that will aid health-care professionals. ML-algorithms have a large following and have been used in a variety of health-care-related projects. They offer a variety of useful models to assist health-care providers. [3]

We employ additive logistic regression (Logiboost), k-NN, and random subspace classifiers to deal with the fetal health dataset in this research, and show performance metrics of the strategy to help health care practitioners tackle fetal health difficulties.

The flow of work is defined .Many methodologies has been proposed. Finding and choosing the correct one is important for successful implementation.
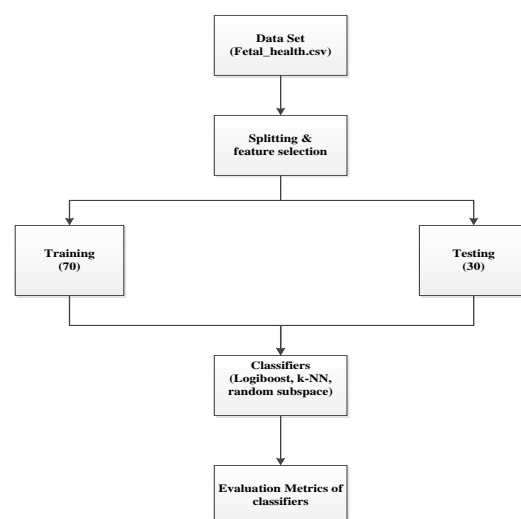


*Fig. #1 Flow of proposed work.*

The important aspect here is to train and test various algorithms and find good features for classification & investigation on various attributes of available data.

Work carried out:

- In the dataset 2126 fetal cardiotocograms (CTGs) were automatically processed and the respective diagnostic features measured.
- Studied various research papers to carryout proposed work.
- Classifiers accessed.

## Data set for proposed work

The dataset for this study was fetal health.csv, which was downloaded from the www.kaggle.com website and sourced from the machine learning repository, which can be found at https://archive.ics.uci.edu/ml/datasets/cardiotocography.

The dataset comprises a collection of twenty one attributes.



*Fig. #2 Snapshot of relation fetal health*

Fetal health is the class label target attribute. Visualization of attributes is shown in Fig. #3



*Fig. #3 Visualization of attributes in Dataset*

The health of the fetus is divided into three categories. (1, 2, 3) To gain a better knowledge of fetal condition, To gain more understanding of class label attribute we convert quantitative data to nominal or categorical values such as 1-Normal, 2-Suspect, and 3-Pathological.

## Pre-processing

Because the majority of the attributes in the dataset are numeric, preprocessing the data by discretizing one or more attributes is required. For this, we use Weka →filter→ supervised→ attribute →discretize to conduct discretization on the baseline value attribute. To discretize a range of numeric values to nominal ones, use the baseline value as the attribute index and apply the instance filter.
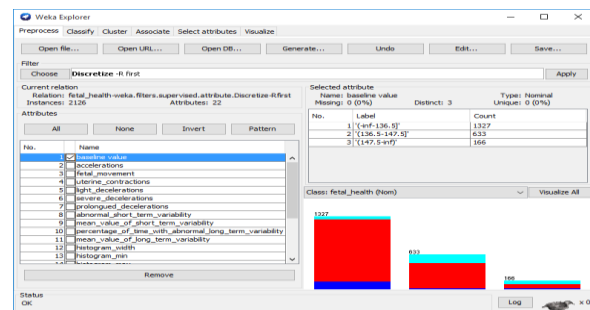


*Fig. #4: Preprocessing of data set.*

## II.     ALGORITHM EVALUATION

The concepts of true positive rate, precision, recall F-measure, and ROC analysis are used to evaluate classification algorithms.

i. Accuracy: It is equal to the total number of occurrences in the dataset divided by the number of valid precisions.

ii. TP-rate: Instances that the class correctly classifies.

iii. Recall: The metric utilized by the algorithm to extract positive interactions.

iv. Precision: The metric that is used to calculate the values that the classifier classifies in order of proximity.

v. F1-Score:  Precision and Recall are weighted averages. As a result, this score takes tangible and intangible false positives and false negatives into consideration. F1, while not as intuitive as accuracy, is widely more beneficial than accuracy, especially when the training set is unequal.

vi. ROC: The primary goal of this analysis is to reach a compromise between profit and expense. It is also possible to graphically compare the results of the tests to determine the efficiency of categorization algorithms. The input file is in Excel and is in.csv format. By introducing columns for desired features as well as rows for mentioning tuples, for each characteristic, each tuple value is taken into account. The class label is represented by the final value.

### III. MACHINE LEARNING TECHNIQUES

Classification is one of the most important machine learning approaches for analyzing data sets by assigning examples to specific classes. It enables model extraction.

1. Additive logistic regression: t is a forecasting technique that takes independent variables to determine the correlation, similar to Linear Regression, but the dependent variable would have to be categorical. Regardless of whether the control variables are numeric or categorical, the dependent variable will always be categorical. Logistic regression is a classification function that provides the Logistic function to model conditional probability. [6]

In probabilistic regression, we quantify the probability distribution of the dependent variable Y given the independent variable X.

$P(Y=1|X)$ or $P(Y=0|X)$ are two ways to write it.

This is the probability distribution of $Y=1$ given X or the linear combination of $Y=0$ given X.

To approximate $P(Y|X)$, a convolution function is applied to a linear collection of input features. The logit function, also known as log-odds, is used in logistic regression to calculate the logarithm of odds. The odds ratio is the proportion of event A's likelihood in the presence of event B to event A's probability in the absence of event B [9].

By plotting the true positive rate using its inherent logistic function, logistic regression analyses the adequacy of the model (our label, what we'd like to anticipate) and one or more target variable (our features). Because this likelihood could have been used to draw conclusions, they must be represented by binary values. This is likely attributable to the logistic function, also widely acknowledged as the sigmoid. A baseline classifier will also be used to modify these values between 0 and 1 either to 0 or 1.

2. 2. k-NN: K-nearest neighbor is a number that indicates how close a person is to another is a straightforward method that treats the training dataset as a classification model in and of itself. The training dataset's records are represented as points in a d-dimensional space (for d characteristics) and a concept of distance between records is developed. [7] the most common distance function is the simple Euclidean distance between the points. The K-nearest neighbor's classifier can use cross-validation to choose a suitable value for K and can also apply distance weighting. Assume that m is the number of training data samples. Assume that p is an unnamed point.
   a. Create an arr[ array of data points] with the training samples. This ensures that each element of the array signifies a tuple(x, y)
   b. for i=0 to m:
   c. Compute the Euclidean distance d (arr[i], p).
   d. Make a list of the shortest S and K distances you've identified. Each of these ranges would seem to match up to a recently described data point.
   e. Replace the standard label with your own.

3. Random subspace classifier: This approach yields a decision tree-based classifier that preserves training information precision and accuracy while attempting to improve generalization veracity as the classifier becomes much more robust. The classifier is composed primarily with several saplings that are constructed thoroughly by pseudo-randomly choosing a subset of similarity measure elements, i.e. trees are constructed in randomly picked subspaces.

### IV. RESULTS AND DISCUSSION

The outcomes of applying the relevant classifiers to fetal health data are shown here. Some classifier parameters for classifier evaluation, such as the threshold value in k-Nearest Neighbor and the seed of a random subspace classifier, may be adjusted while analysis is performed. We use Weka software solution for these evaluations. [8] On a fetal-health dataset, we apply a variety of classifiers in this study.

**Train and Test Split:** To test the data with unknown data points rather than the same ones with which the model was trained. This makes it much easier to

capture model performance. We use 70% and 30% percent splits respectively for our evaluations.

| Classifier used | Accuracy | ROC | MAE | RMSE | Time (Sec) |
|---|---|---|---|---|---|
| Logistic regression (Logiboost) | 92.16 | 0.96 | 0.09 | 0.21 | 0.67 |
| k-Nearest Neighbor(k=1) | 89.96 | 0.86 | 0.06 | 0.25 | 0.02 |
| RandomSubspace | 90.75 | 0.97 | 0.11 | 0.21 | 0.74 |

*Table#1: Accuracy of classifiers*

For accuracy and error performances, visual representations are shown in Fig. #5 and Fig. #6



**Accuracy of classifiers**

- Accuracy: 92.16, 89.96, 90.75
- Logistic…, k-NN(k=1), Random…

*Fig. #5 Accuracy of Classifiers*



**Classification Errors**

- MAE, RMSE
- 0.09, 0.21 (Logi…)
- 0.06, 0.25 (k-…)
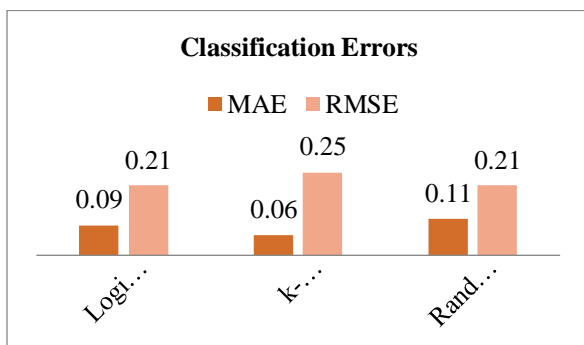- 0.11, 0.21 (Rand…)
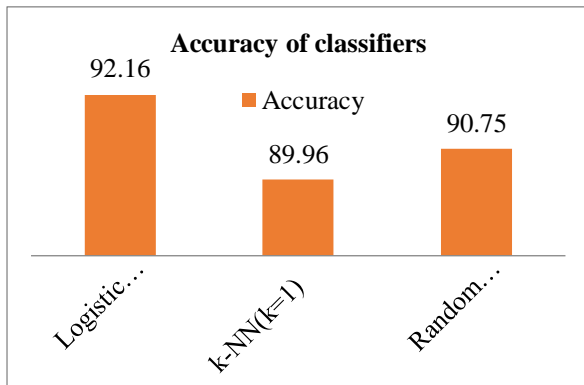
*Fig. #6: Errors reported by classifiers.*

Table#2 shows that the percent split in additive logistic regression (Logi boost) technique outperforms the other two classifiers with accuracy of 92 percent, ROC of 96 percent, MAE of 0.09, and RMSE of 0.21.
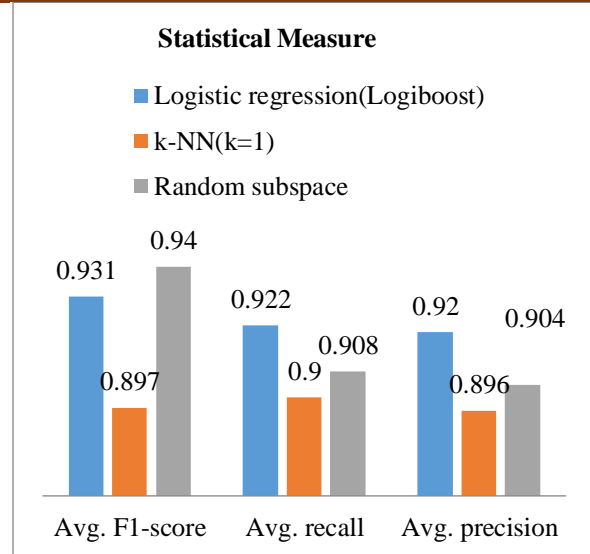


**Statistical Measure**

- Logistic regression(Logiboost)
- k-NN(k=1)
- Random subspace

| | Avg. F1-score | Avg. recall | Avg. precision |
|---|---|---|---|
| Logistic regression(Logiboost) | 0.931 | 0.922 | 0.92 |
| k-NN(k=1) | 0.897 | 0.9 | 0.896 |
| Random subspace | 0.94 | 0.908 | 0.904 |

*Figure #7: Stats on classification algorithms*

We include a set of analytical criteria in the this segment to quantify this same classifier's accuracy in categorizing the fetus health dataset.[4] The mean F1-score is by far the most crucial since it quantifies alteration between specimen means/variations. The increased the F-Score, as shown in table #3, the stronger the categorization. By establishing a link between true positive rate and false positive rate, the ROC AUC measure determines how good a model is. AUC of area 1 will define a flawless model in an ideal dataset. Inequality datasets, on the other hand, have no effect on this value. F1, on the other hand, is an excellent metric for determining the performance of estimators for imbalance datasets because it is ascertained by calculating the weighted of precision and recall. As a result, the proposed system is the one with the highest F1, but if there is a tie, we can also use ROC AUC to split it.

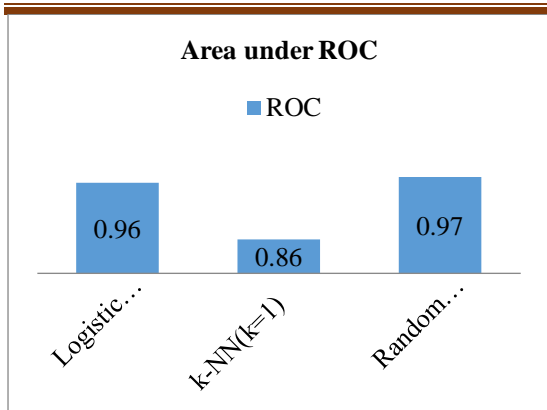| Classifier | Avg. F1-score | Avg. recall | Avg. precision |
|---|---|---|---|
| Logistic regression(Logiboost) | 0.931 | 0.922 | 0.92 |
| k-Nearest Neighbor(k=1) | 0.897 | 0.9 | 0.896 |
| Random subspace | 0.940 | 0.98 | 0.904 |

*Table#2 Statistical measure.*

**Fig. #8 Area under ROC**

**Feature selection:** The fetal health dataset's features were chosen using a variety of subset assessors. It still has a decent level of accuracy when it comes to fetal health classification and diagnosis. For picking important attributes per classification and to assess the quality of spawned subsets, the classifiers subset estimator methodology is being used as the feature selection technique.

| Classifiers | Accuracy | Accuracy in such attributes | attributes chosen |
|---|---|---|---|
| Logistic regression (Logiboost) | 92.16 | 92.16 | 1,2,8,9,10, 12,13,15,18, 19 |
| kNN(k=1) | 89.96 | 89.96 | 2,3,10,12,18 |
| Random Subspace | 90.75 | 92.16 | 1,2,5,7,8,9,1 0, 11,13,14,16, 18,19 |

**Table #3: Accuracy of feature selection with selected attributes.**
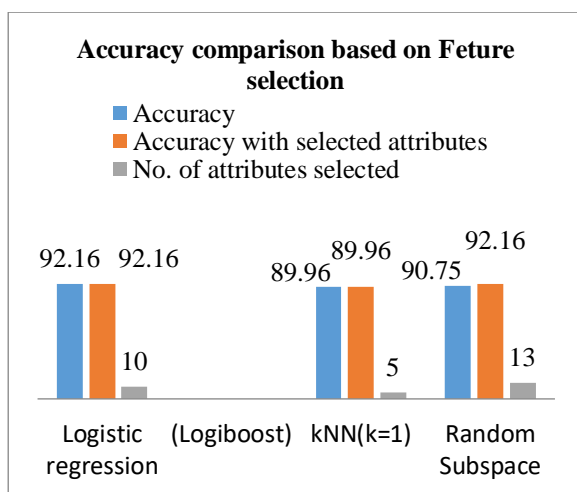


**Fig. #9 Accuracy comparisons**

**Discussion**

In our work data set having CTG records from 2126 expectant mothers, we utilized three different algorithms to assess pessimistic fetus output, i.e. alleged and compulsive circumstances, the Logiboost approach was determined to be the most accurate. CTG perception is completely reliant on the obstetrician's investigation of the parsing, which outcomes throughout rationality in data analysis. [5]The study's strength seemed to be that it would use 3 distinct machine learning algorithms on the CTG data source and advised one of those with the best ability to estimate both suspicious and pathological fetus states. As a result, proper interpretation of CTG is critical for subsequent treatment. The requirement for a uniform machine learning method for analyzing is high. Although there is still much space for development, our work can serve as a viable alternative to artificial analysis, improving diagnostic efficiency and accuracy.

## V. CONCLUSION

Three classifiers were used to classify the fetal health data set in this work. Additive logistic regression (Log boost), k-NN, and random subspace are the algorithms employed. To efficiently categories' fetal health situations, the suggested classifier model is based on training and testing, as well as feature selection. It was discovered that the logi boost classifier produces more accurate results.A classification subset evaluator, as well as best first search, which searches the space of attribute subsets by greedy hill climbing, enhanced with a backtracking facility on the existing dataset and still performs well for classification and diagnosis, can be used for feature selection.Various classifiers can be used to evaluate their performance by modifying some of their parameters, which could be a fascinating feature to add to our work.

**REFERENCES**

1. Zahra Hoodbhoy, Mohammad Noman,1 Ayesha Shafique,1 Ali Nasim,1 Devyani Chowdhury2, and Babar  Hasan, "Use of Machine Learning Algorithms for Prediction of Fetal Risk Using Cardiotocographic Data" ,International Journal of applied basic medical research,2019 Oct-Dec; 9(4): 226–230.
2. J. Jayashree, Harsha T, Anil Kumar C, and J. Vijayashree, "Enhanced Optimal Feature Selection Techniques for Fetal Risk Prediction Using Machine Learning Algorithms",

3. International Journal of Engineering and Advanced Technology (IJEAT)

4. Gongao Xue's paper "The Application of Machine Learning Models in Fetal State Auto-Classification Based on Cardiotocograms", presented at GBEM2019,IOP Conf. Series: Earth and Environmental Science 310 (2019) 052007.

5. Khaled Mohamad Almustafa, "Prediction of chronic kidney disease using different classification algorithms", informatics in medicine unlocked, science direct, ELSVIER

6. Ersen Yılmaz ,Çağlar Kılıkçıer,"Determination of Fetal State from Cardiotocogram Using LS-SVM with Particle Swarm Optimization and Binary Decision Tree", Computational and mathematical methods in medicine, Hindawi

7. Book:Jiawei Han and Micheline Camber, "Data Mining Concepts and Techniques", 2nd Ed., Morgan Kaufmann Publishers, Elsevier, Inc., 2006

8. Vikram Pudi and P. Radha Krishna, "Data Mining", Oxford university press, 5th impression 2012

9. H. Witten, Elbe Frank, Mark A Hall, "Data Mining Practical machine learning tools and techniques", Morgan Kaufmann publisher,2011 Elsevier Inc

10. Christopher M Bishop," Pattern recognition and Machine Learning ", r2006 Springer Science + Business Media, LLC

# International Journal of Computer Research & Technology ( IJCRT )

A Peer Reviewed Half Yearly Research Journal

## Special Issue
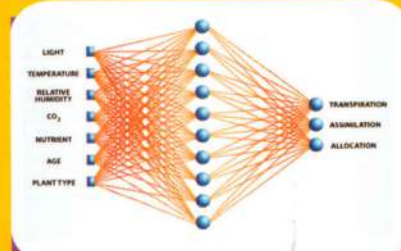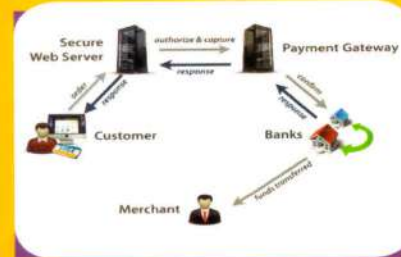
Proceeding of the 7th National Conference on
"Emerging Trends in Computer and Information Technology"
held on 29th May 2021.

**Editor in Chief**
Dr.B.H.Barhate