ISSN: 2581-3404 (Online)
International Journal of Innovative Research in Technology and Management, Vol-5, Issue-6, 2021.

Enhanced Diseases Recognition Classification Through Support Vector Machine & Optimization Method

Sangeeta Kumari[1], Ritesh Kumar Yadav[2], Varsha Namdeo[3]
CSE Department, SRK University, Bhopal, India[1, 2, 3]

Abstract- *Medical science has grown tremendously in the last decade. However, our country's healthcare system is unable to keep up with the demand of an increasing population. An increasing number of people prefer private hospitals over government hospitals given the availability of physicians, equipment, diagnostics, and services, among other things. Today's health-care industry creates and produces vast or massive amounts of datasets, necessitating the development of specialized tools for the effective and accurate extraction of knowledge or information from stored data. Data mining provides a variety of methodologies and tools for data diagnosis in the health care business. Some data is based on clustering, while others are based on categorization, association rule analysis, regression, and so on. Data classification is one of the most extensively utilized data mining approaches in healthcare organizations. It is an example of supervised classification methods. Classification algorithms are also used to anticipate the therapy cost of healthcare services since data is separated into calls and recognized according to their provided aim or output. Current data mining and machine learning approaches are particularly effective for diagnosing patients and storing massive amounts of data. Some optimization strategies are also employed to increase the detection rate and diagnosis findings of illnesses. Modern medicine is a one-of-a-kind combination of a patient, healthcare experts, and technology. In this dissertation, we proposed a new model based on classification methods including support vector machines, neural networks, and optimization methods, which is a bi-logically inspired method for improving classifier results in terms of some performance characteristics such as accuracy, precision, recall, and so on. We measure and improve for all performance parameters for various datasets such as heart patients, liver patients, and cancer patients. The dataset for all patients is collected from the UCI machine learning repository, which provides the legitimate dataset for the study activity, and the simulation program is MATLAB. Our experimental results in this dissertation demonstrate that our approach has a higher detection rate of categorization for performance characteristics than other current techniques.*

Keywords:- Disease Classification, SVM, Optimization Method, Machine Learning, MATLAB Introduction.

Introduction

The mining of healthcare database is very critical issue. The healthcare data stored the information about medical diseases and patient's information. For the estimation of patients and disease used some intelligent software for the predication of disease. The prediction of disease and medicine is fundamental issue in health care environment [1]. For the extraction of better information used data mining technique for the healthcare system.

In today's lifestyle and diagnosis system, predicting and classifying heart disease is a serious concern. Various classification algorithms, such as support vector machines, neural networks, and other techniques, are utilized for prediction and classification. Another classification approach based on association rule mining is associative classification. The association rule for related data is generated using the association rule mining approach [2]. These categorization rules establish a certain number of classes. Other approaches, like neural networks, will be utilized to categories the data into a number of categories, with back propagation neural networks and feed-forward neural networks being employed in the process. We may also utilize the superset

of classifiers, i.e., support vector machines, to get the best classification results and evaluate other performance criteria.

From the past years there is medical science interesting domain for researcher, due to large number of population are affected from its. In our country most of the peoples are not able to get medical treatment on time at everywhere, or most of the peoples suffer from the various types of diseases but not getting any prevention to recover from diseases [3]. To cover large number of peoples we need rich infrastructure such as large number of hospitals, medical lab, equipment, diagnostic tools, machinery, huge memory to stored large number of information, available data everywhere at each time. There are also the tools and techniques for the medical diseases diagnosis are playing very vital role in this sector. There is various techniques such as data mining tools which further classified the data in the supervised and unsupervised manner, some evolutionary techniques such as genetic algorithm mostly used for the big data analytics, neural network classifier, support vector machine, decision tree classifier, rule based classifier and some swarm intelligence family methods for the optimization such as ant colony optimization, particle swarm optimization and honey bee classification [4].

This study deals with the application of data mining techniques in medical science diseases diagnosis system. The main objective of this study is to detect for any diseases at early stage for a patients therefore we can increase the human life. Here we improve the detection rate of patients diseases classification rate in the terms of some performance parameter for each dataset, therefore we can enhance the result which are generated from previous or existing techniques in the medical science domain. To evaluate the performance comparative study we used some classification techniques such as neural network classifier and support vector machine classifier and optimization methods for medical database, Build and test models using MATLAB software and used reputed dataset form UCI machine learning Repository.

II. Related Work

[5] In the base paper author provides the comparative study between various tools and methods for the disease diagnosis specially for the heart diseases and shows the accuracy, here author compare their own methods which is a combination of feed forward neural network and particle swarm optimization shows better results than all other existing methods. They used here heart diseases data set for the classification and diagnosis of patients which is extracted from the UCI machine learning repository.

[6] In This paper author presents the comparative study for the various dataset using Artificial Bee colony optimization and Feed forward neural network, where Artificial Bee colony is used to initialize and optimize the connection weights of Feed forward neural network. They proved with their results that, Artificial Bee colony optimized Feed forward neural network approach has outperformed the back propagation neural network approach.

[7] In this paper author proposed a new intelligent classifier by using hybrid Three Term back propagation based on a Multi-objective Genetic Algorithm was introduced and successfully applied for the medical disease diagnosis classification problem; it is effective as a classifier with good performance and a high accuracy rate. The results indicate that our proposed method demonstrated effectiveness in dealing with the medical disease diagnosis classification problems.

[8] In this paper author proposed medical diseases diagnosis for the heart diseases dataset using neural network approach. Their computing power, their ability to make predictions, and the impact they have on medical life brought them on the top of artificial intelligence tools that are used in this domain.

[9] Describe the value of data mining algorithms and systems for medical institutions, with a particular emphasis on creating and implementing telemedicine solutions in order to improve the quality of healthcare services delivered to patients is the topic of this study. They give a framework for integrating data mining algorithms into telemedicine systems, as well as an outline of how to evaluate and improve the established solution through the use of Business Process Management methodologies. They are mostly utilized in the

processes of gathering medical information from patients, giving the necessary medical information to physicians, patients, and other members of the healthcare team society, real-time patient monitoring, and diagnosing patients with the help of expert systems.

[10] Using a network-based approach, the authors of this research devised and validated a method to priorities genes for human illnesses on the basis of their phenotypic profile. It has been developed into a strong and transparent tool for the identification and ranking of candidate genes. It is a new resource for the molecular interpretation of the evidence from genome-wide linkage maps, and it will aid in the knowledge of how genetic variations linked with illness or quantitative phenotypes impact disease or statistical phenotypes.

[11] Author may assess various patterns in this work using data mining techniques, which will be used in the future to develop intelligent systems and make judgments. It is defined as many techniques of recognizing information or adopting solutions based on understanding and data extraction from these data in order for them to be employed in different areas such as judgments, the prediction worth for the predicted and computation.

[12] This paper shows the comparative study of hybrid and non-hybrid techniques and it is obviously seen that fuzzy system performs well with genetic algorithm than with neural networks. Since GA reduces the attributes, the times taken to solve the problem also get reduced. Fuzzy grid partition system along with GA gives promising prediction of the disease. In future, more studies with other datasets are needed to establish the higher performance of genetic algorithm based hybrid models. This paper proposes a comparison of fuzzy and hybrid techniques like fuzzy genetic and neuro-genetic for the diagnosis of diabetes on a Pima Indian Diabetes dataset from UCI machine learning repository. Diabetes is one of the major challenging diseases to the world. We also extend the hybrid algorithm of combining genetic with Neuro-Fuzzy.

III. Proposed Work

In this section we proposed the new model for the health care system to compute some performance parameters such as accuracy using classification and optimization techniques. Here we improve the efficiency rate in the terms of accuracy for the proposed system compare than existing system which is provide better results in the medical science domain.

Particle swarm optimization method is types of swarm intelligence family methods which is particularly used for the optimization of results or improve the rate of any classification dataset which is used for various numbers of applications. Particle swarm optimization mostly used with four improves the results for any area mostly used with data mining methods and some other classification methods.

Here we describe some number of step for the particular proposed heart based and other patients' dataset and they are following:-

1. Generate random population of n solution (particles)
2. For each individual i€N: calculate fitness (i)
3. Initialize the value of the weight factor, w
4. For i=1 to n particles
   Set pBest as the best position of particle i
   If fitness (i) is better than pBest then
   Set pBest (i) =fitness (i)
   End for i
5. Set gBest as the best fitness of all particles
6. For i=1 to n particles
7. Update the value factor of the weight w
8. Check if termination is true.
9. End

IV. Proposed Model

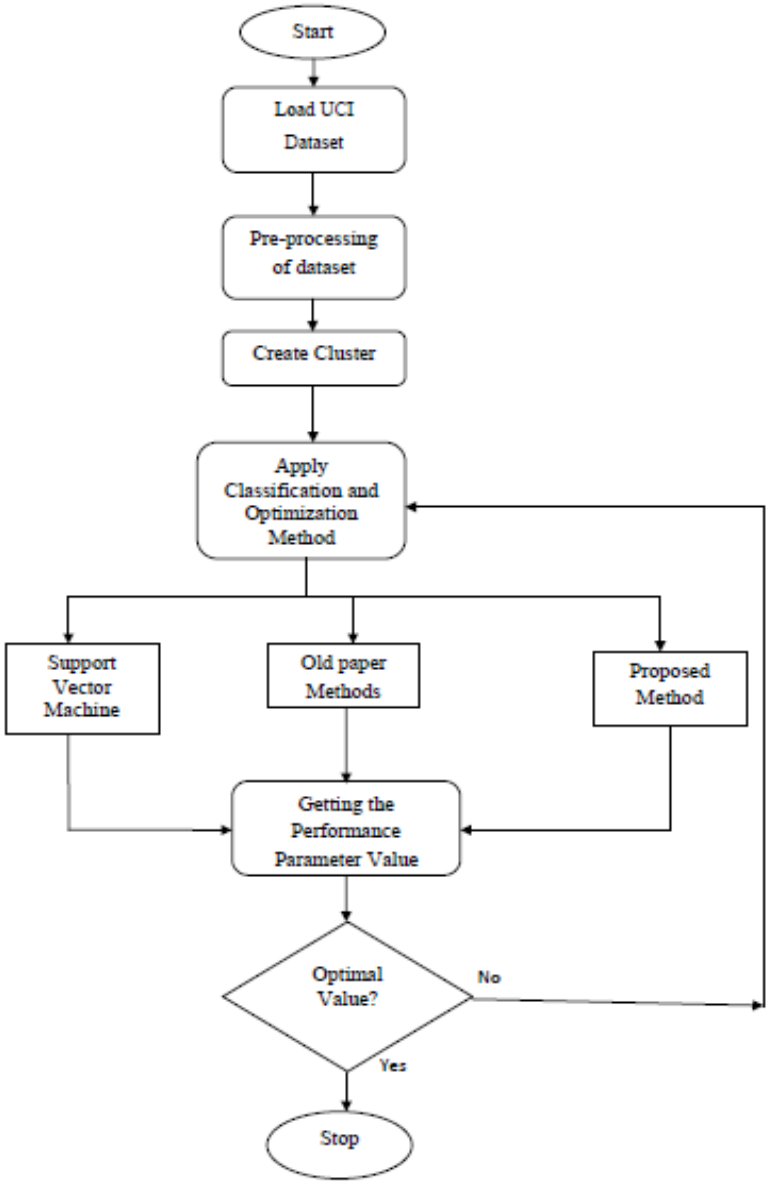

Figure 1: Classification and optimization proposed model**.**

V. Result Analysis

In this section, experimental process of we show that the comparative result analysis study for the Health care sector with disease diagnosis of various dataset such as Heart, Liver, Cancer etc. are performed. This process of disease diagnosis of various dataset is done by using Three methods that are support vector machine, the base paper method and Proposed method. For the evaluation of performance parameter we used MATLAB software for the authenticity and effectiveness of results.

In this section we discuss about the dataset which we used for the diseases detection in the field of health care. There are all these dataset types will be fetched from the UCI machine learning repository for the research purpose. In future we implement the diseases detection and improve the accuracy and other performance evaluation parameter with the help of all these dataset using data mining, optimization techniques, the used dataset is here Heart dataset, liver dataset and Cancer dataset.

Result Parameters:

Precision- Precision is defined as the proportion of projected positives/negatives that are really positives/negatives. Precision is measured in percentages.

Recall - It is the ratio of the total positives and negatives that are projected to be positive and negative

Accuracy-It is defined as the function of number of predictions that were right, or the percentage of cases that were properly categorized in a certain time period.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP}$$

We will demonstrate how to determine these parameters using the appropriate formulae in the section below. In addition, we have included a graph for that specific data set in the next section.

Comparative Result Analysis:

Dataset name	**Method**	**Elapsed Time (Sec)**	**Accuracy in (%)**	**Precision in (%)**	**Recall in (%)**
Heart	SVM	9.3	78.23	75.49	76.62
	OLD METHOD	8.66	80.25	77.24	74.56
	PROPOSED	10.48	93.59	87.47	88.68

Table 1: Comparative result analysis for the Heart dataset.

Dataset name	Method	Elapsed Time (Sec)	Accuracy in (%)	Precision in (%)	Recall in (%)
Liver	SVM	9.22	78.56	77.56	75.62
	OLD METHOD	10.24	80.78	78.24	74.38
	PROPOSED	11.36	94.56	88.56	89.62

Table 2: Comparative result analysis for the Liver dataset.

Dataset name	Method	Elapsed Time (Sec)	Accuracy in (%)	Precision in (%)	Recall in (%)
Cancer	SVM	21.33	83.78	80.26	79.45
	OLD METHOD	22.56	85.47	82.49	78.69
	PROPOSED	23.56	97.45	82.59	93.46

Table 3: Comparative result analysis for the Cancer dataset.

VI. Comparative Result Graph

COMPARATIVE RESULT GRAPH

Figure 2: Show that the comparative result analysis for the Heart disease diagnosis using various classification and optimization techniques, our empirical result study shows that better accuracy than existing methods.

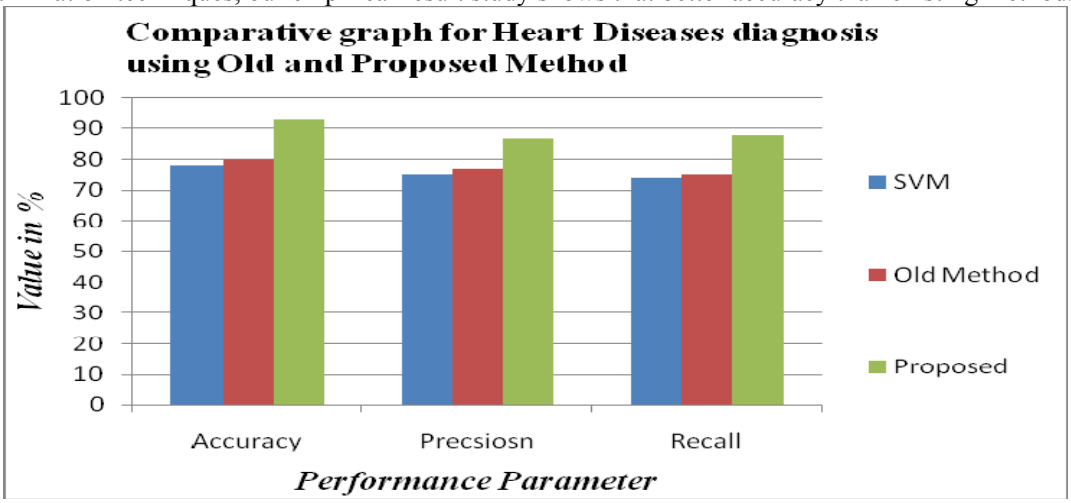

Figure 2: Comparative result analysis for the Heart disease diagnosis.

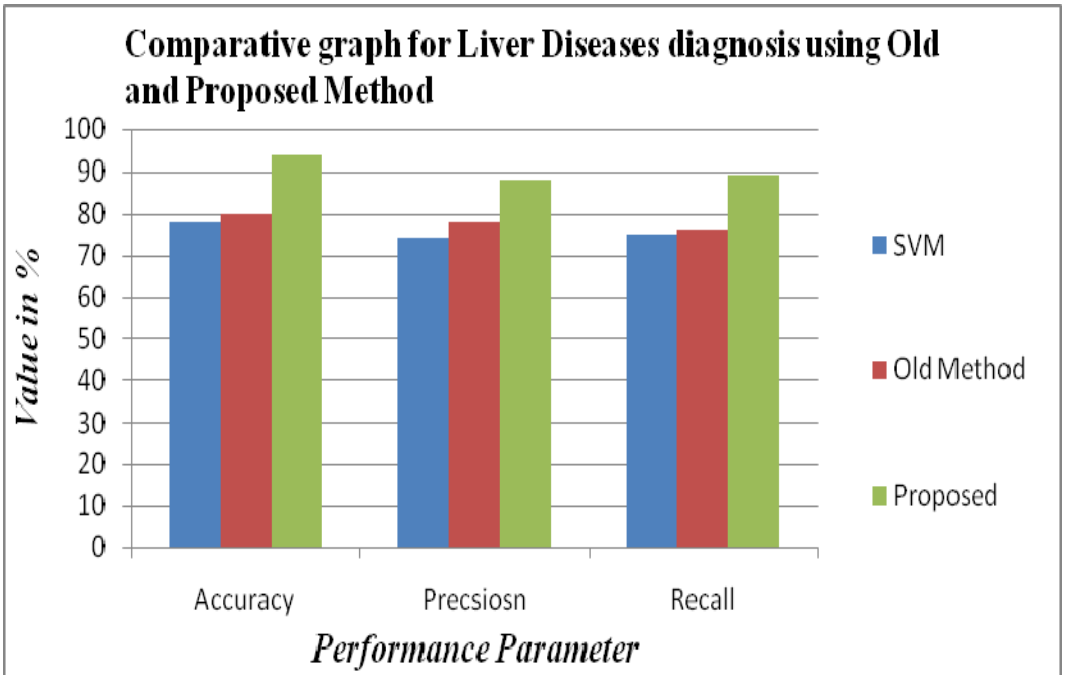

Figure 3: Comparative result analysis for the Liver disease diagnosis.

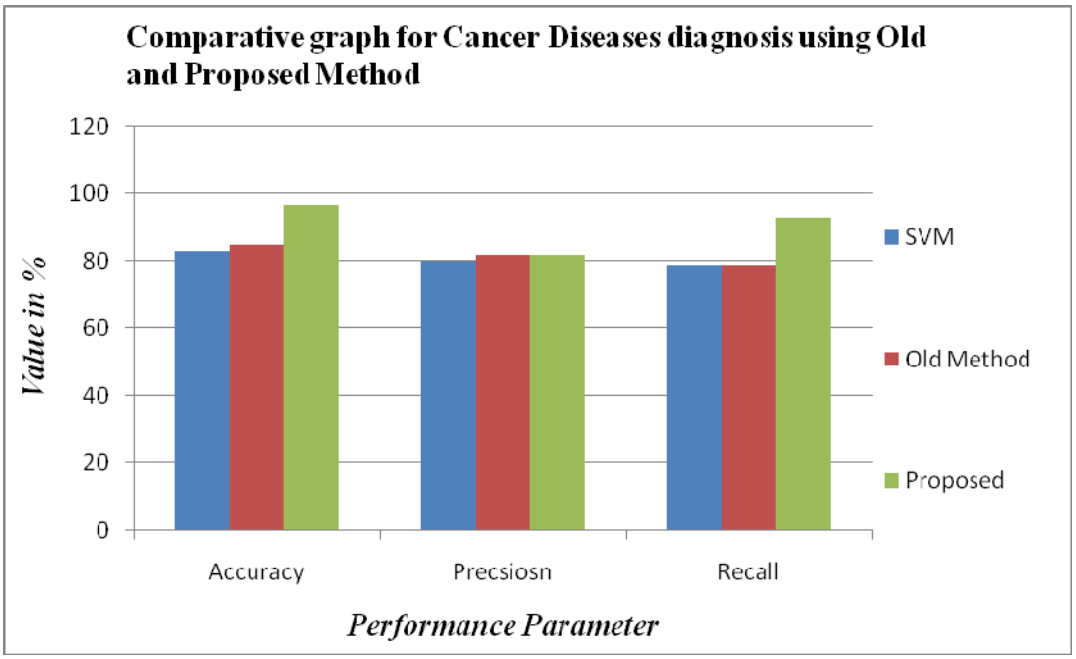

Figure 4: Comparative result analysis for the Cancer disease.

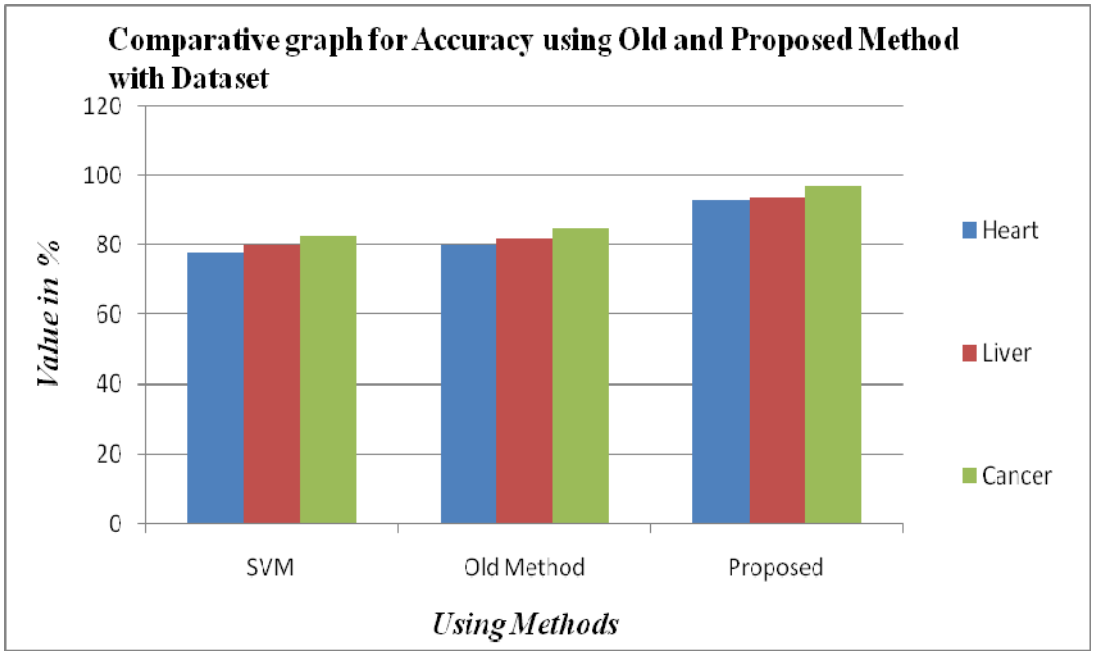

Figure 5: Comparative result analysis of Accuracy for the Heart disease, Liver disease and cancer disease dataset.

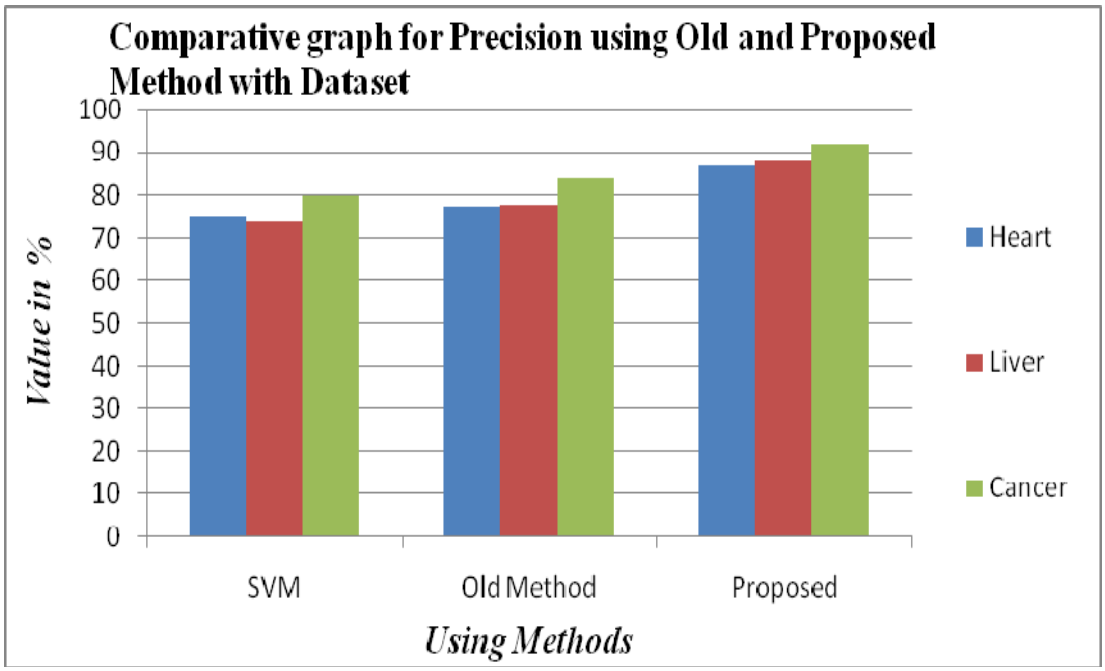

Figure 6: Comparative result analysis of Precision for the Heart disease, Liver disease and cancer disease dataset.

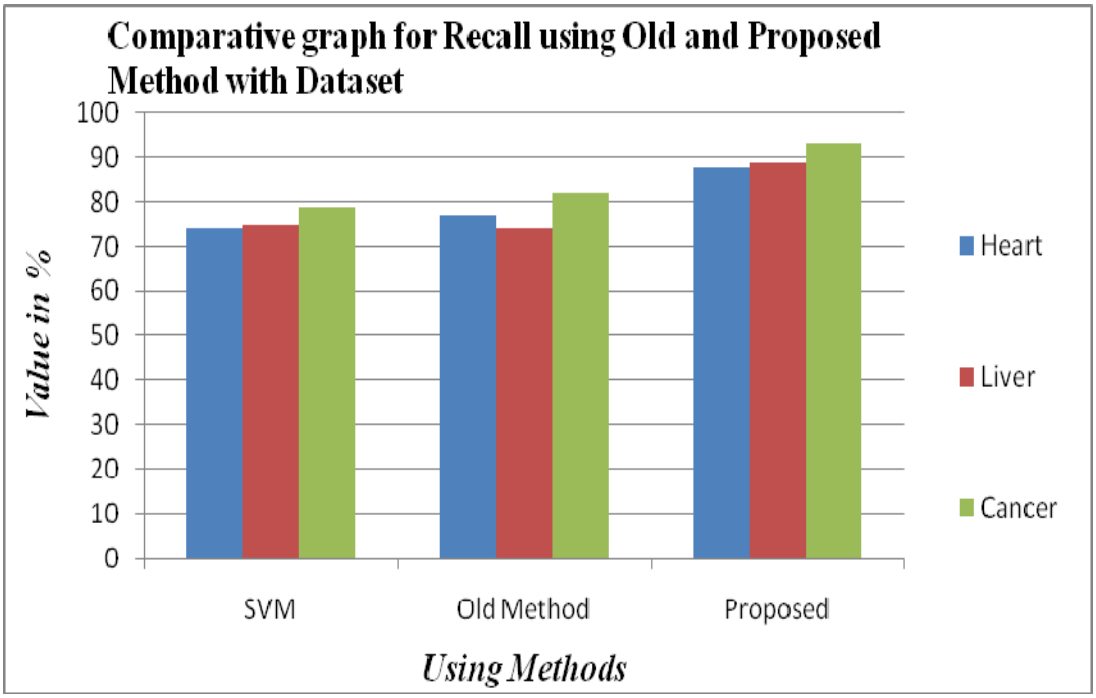

Figure 7: Comparative result analysis of Recall for the Heart disease, Liver disease and cancer disease dataset.

VII. Conclusion

The major purpose of every government or agency is to deliver effective and timely healthcare services; nevertheless, the process of illness diagnosis is a difficult field for medical science to master. In the contemporary day, we employ a variety of technologies to diagnose medical disorders. The diagnostic system is based on a penitent's treatment, historical history, and a large amount of medical data, among other things. In modern civilization, the healthcare business is the fastest-growing sector of many countries' economies. There are several disorders to diagnose, including heart disease, cancer, diabetes, and liver disease. In today's lifestyle and diagnosis system, predicting and classifying heart disease is a serious concern. Various clustering and classification algorithms, such as support vector machine, neural network, and other optimization techniques, are utilized for prediction and classification. To enhance the outcomes of classifier and optimization approaches, optimal cluster and classifier value selection is applied. Numerous studies have aided in the discovery of consensus clustering. On the same datasets, we compared the classification accuracy of traditional and suggested techniques. Medical science is a critical aspect for every country's development; if a country achieves the greatest outcomes and makes the most advancement in the medical industry or health-care system, it will undoubtedly boost the country's growth in all sectors. There are a number of strategies we employ for medical treatment of patients, such as illness diagnosis, but they are limited by the time domain and the accuracy rate of this categorization. Finally, we view computing efficiency as a significant concern in the near future when researchers need real large-scale data for clustering and classification. The variety of medical diagnostic and illness data is growing every day, making it a difficult task to store and use vast amounts of data. In the future, we may be able to apply big data analytics in the health services system to improve the outcomes even more.

References:-

[1] M.Akhil jabbar , Priti Chandra and B.L Deekshatulu "Prediction of Risk Score for Heart Disease using Associative Classification and Hybrid Feature Subset Selection", IEEE, 2012, Pp 628-634.

[2] Basheer Mohamad Al-Maqaleh, Ahmed Mohamad Gasem Abdullah "Intelligent Predictive System Using Classification Techniques for Heart Disease Diagnosis" International Journal of Computer Science Engineering, 2017. Pp 145-151.

[3] Atif Khan, John A. Doucette, Changjiu Jin, Lijie Fu, Robin Cohen "An Ontological Approach To Data Mining For Emergency Medicine" University Avenue West Waterloo, 2010.

[4] Saurabh Pal, Vikas Chaurasia "Data Mining Approach to Detect Heart Diseases" IJACSIT 2013 PP 56-66.

[5] Majid Ghonji Feshki and Omid Sojoodi Shijani "Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network", IEEE, 2016, Pp 48-53.

[6] Asha Gowda Karegowda and Darshan M "Optimizing Feed Forward Neural Network Connection Weights Using Artificial Bee Colony Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, 2013, Pp 452-454.

[7] Sultan Noman Qasem, Siti Mariyam Shamsuddin, Siti Zaiton Mohd Hashim Maslina Darus and Eiman Al-Shammari "Memetic multiobjective particle swarm optimization-basedradial basis function network for classificationproblems", Elsevier, 2013, Pp 165-190.

[8] Ashraf Osman Ibrahim, Siti Mariyam Shamsuddin, Nor Bahiah Ahmad and Sultan Noman Qasem "Three-Term Backpropagation Network Based On Elitist Multi-objective Genetic Algorithm for Medical Diseases Diagnosis Classification", Life Science Journal, 2013, 1815-1822.

[9]Mr. Deepak Dhanwani and Prof. Avinash Wadhe "Study Of Hybrid Genetic Algorithm Using Artificial Neural Network In Data Mining For The Diagnosis Of Stroke Disease", International Journal of Computational Engineering Research, 2013, Pp 95-100.

[10] Delia-Maria FILIMON and Adriana ALBU "Skin Diseases Diagnosis using Artificial Neural Networks", IEEE, 2013, Pp 1-6.

[11] Mihaela Gheorghe, Ruxandra Petre "Integrating Data Mining Techniques into Telemedicine Systems" Informatica Economică 2014 PP 120- 130.

[12] Li Jiang, Stefan M Edwards, Bo Thomsen, Christopher T Workman, Bernt Guldbrandtsen , Peter Sørensen "A random set scoring model for prioritization of disease candidate genes using protein complexes and data-mining of GeneRIF, OMIM and PubMed records" BMC Bioinformatics 2014 PP 1-13.

[13] Temesgen Dileba Kale "Application of Data Mining Techniques to Discover Cause of Under- Five Children Admission to Pediatric Ward: The Case of Nigist Eleni Mohammed Memorial Zonal Hospital" Health & Medical Informatics 2015 PP 1-14.

[14] Kendra J. Ahmed, Mahbub K. Ahmed, Scott McKay "A Brief Review of Alternative Uses of Data Mining: Education, Engineering, & Others" ASEE 2015.

[15] Wu-Chen Su "A Preliminary Survey of Knowledge Discovery on Smartphone Applications (apps): Principles, Techniques and Research Directions for E-health" ICME 2014 PP 369-374.

[16] S.Stanley young, Robert L,Christophe Lambert, "Bias and response heterogeneity in an air quality data set" Center for Global Health, USA 2015 PP 1-15.

[17] Peng Cao, Xiaoli Liua,, Hang Bao, Jinzhu Yang, Dazhe Zhao "Restricted Boltzmann machines base oversampling and semi-supervised learning for false positive reduction in breast CAD Restricted Boltzmann machines based oversampling and semi-supervised learning" 2015 PP 1541–1547.