

# MEDICAL DATA CLASSIFICATION USING DIFFERENT OPTIMIZATION TECHNIQUES: A SURVEY

Balasaheb Tarle<sup>1</sup>, Rupali Tajanpure<sup>2</sup>, Sudarson Jena<sup>3</sup>

<sup>1</sup>Research Scholar, GITAM University, Hyderabad

<sup>2</sup>Research Scholar, GITAM University, Hyderabad

<sup>3</sup>Associate Professor, GITAM University, Hyderabad

## Abstract

Classification of health data for perfect opinion is a on the rise field of relevance and investigate in records removal. Medical data classification is an important data mining problem being discussed about for a decade that has attracted several researchers around the world. The sorting techniques provide invaluable information to pathologist for diagnosis and treatment of diseases. By detecting and counting blood cell within the blood smear using sorting techniques, it is quite possible to detect so many diseases. The aspiration of projecting data mining in scientific drug is to take models that can use patient exact information sequence, then system will be predict the conclusion of interest and to thereby support scientific administrative. The various data mining algorithms like Classification, Clustering, Regression techniques, soft computing and many more are used in health data sorting. According paper, we have literature survey some of the methods correlated to health data sorting. The chief heart of survey is to scrutinize DMT like ANN, decision tree, support vector machine, nearest neighbor, Bayesian algorithm required for health statistics removal above all to determine nearby regular diseases such as swine flu, heart disease, lung tumor, dengue, breast growth tumor and so on.

**Keywords:** Medical In Order Classification, Artificial Neural Network, Decision Tree, Support Vector Machine, Nearest Neighbor, Bayesian Network

\*\*\*

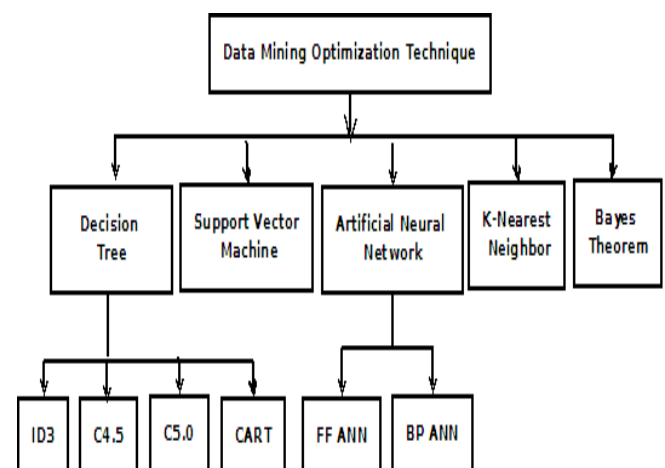
## 1. INTRODUCTION

The process of knowledge discovery in databases includes data mining as an important step where pattern discovery is done by using different techniques of classification [1]. It is a process of discovering the class model according to the values of attribute. In this process the data is analyzed from different perspectives and is summarized into useful information. Data mining finds new patterns for the users and provides meaningful information to the user in order to enhance the decision making process. It is useful to find useful patterns which are important task in medical diagnosis and treatments [2]. Classification of medical data for error free diagnosis is a growing field of research in data mining [7][8]. When these algorithms are used appropriately, they can improve prediction, diagnosis and disease classification [12]. Classification technique segregates data according to some predefined classes. They also predict the class for each data points. The two categories of classification are binary and multilevel classification. Classification model could make predication of categorical label include discrete or unordered variables [3][4].

For medical data different data mining techniques used are decision tree algorithms like ID3, C4.5 [14], C5.0, and CART [16][17][18][19][20], neural networks [9][10], Association rule [11][12], prediction [21], Fuzzy cluster analysis [22], classification [17] [23][24], Bayesian Network [14][25], Naive Bayesian algorithm [15][26], combination

of Naive Bayes, K-means [27], Time series technique [28] [29], combination of SVM, ID3 and ANN [16], cluster and sorting [30][31], k-NN [24], and Bayesian Network [14]. Accordingly the paper review we get the summary of data mining organization algorithms with their application in the health statistics removal.

These algorithms enhance the classification process to be more accurate. Fig.1 shows various approaches in medical data classification [1]. Accordingly the paper review provides the summary of classification algorithms with their usefulness in health statistics classification.



**Fig 1** various data mining optimization techniques useful in classification of medical data

The overall structure of this paper is prepared as follows. Section II has a look on data mining techniques and its merits and demerits as well as challenges of various optimization techniques. Section III discusses about the literature survey of data mining optimization techniques, which are used for medical data classification. Section IV summarizes various algorithms used for classification of different diseases. Section V compares performance of data mining optimization techniques with different data sets. Finally section V wind up the survey paper.

## 2. DATA MINING TECHNIQUES

### 2.1 Decision Trees (DT's)

In decision tree, each non-leaf node represents a test on the data item under consideration. The outcome of the test decides path or particular branch to be selected. We can classify a data item, by starting at root node and follow the path till we reach a leaf. When a terminal node is approached the decision is made. Decision trees uses recursive data partitioning, which is a result of hierarchical organization of rules [14],[15].

#### Merits

- Domain knowledge is not required while constructing decision tree.
- The ambiguity of complicated decision is minimized. Also exact value is assigned to outcomes of various actions.
- The data with high dimension can be easily processed
- It is easy to understand.
- Both numerical and categorical data can be handled by DT.

#### Demerits

- Always as a result one output attribute is created.
- Categorical output is generated.
- The performance of classifier depends upon the type of dataset.
- Complex decision tree is generated for numeric type of dataset.

#### Challenges

- Existing algorithm of decision tree is good for classification of small problems but becomes hard to implement as complexity of problem increases.
- If learning set contains noisy data, it will fail to find a tree as well as affects the accuracy of the output.
- Decision making statements created can become more complex for increase in variable [16], [17], [18], [19], [20].

### 2.2 Artificial Neural Networks

Artificial Neural Network is used as an important tool for classification of medical data. Many researches have worked with this approach. First Neural network is data driven self-adaptive methods while second neural networks are

universal functional approximation. In order to train a neural network properly a large set of training data is needed. Artificial neural networks can be divided into two categories depending on learning phase as supervised and unsupervised [14]. In first category, the network is trained input and respective output patterns. During this, the neural network adjusts the connection weights to match respective output with the actual output in an iterative process till desired results are obtained. An ANN of the later learning type, such as the self-organizing map, only inputs are fed to NN, without known answers. The network calculates the weights and develops its own representation of the input [15]. All these methods are described and reviewed in [11]. These methods are used in many applications like Magnetic Resonance Imaging (MRI)[61].

#### Merits

- Due to the weights, NNs are more robust
- The performance of NN is improved by learning.
- Here low error rate i.e. high degree of accuracy is obtained with appropriate training
- They are more robust in noisy data.

#### Demerits

- Local minima.
- Over-fitting.
- For large neural networks, processing time required is more.

#### Challenges

- Lack of sufficient training data.
- Lack of acceptable validation procedure.

### 2.3 K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) classifier finds unidentified data point using the previously known (nearest neighbor) and classified data points [8]. It classifies the data points using more than one nearest neighbor. There are number of applications of K-NN in different areas like cluster analysis, pattern recognition, medical datasets, images, online marketing *etc.* For classification of chronic disease Jen *et al.*, used K-NN and LDA to generate warning system. Here K-NN is used to analyze the relationship between cardiovascular disease and hypertension[24]. Also the risk factors of various chronic diseases are found in order to reduce the complication while classifying diseases [39]. K-NN classifier is proposed by Shouman *et al.*, for analysis of cardiac patients [40]. The UCI data was used and experimented with or without voting K-NN classifier. The results showed that K-NN performs better without voting as compared to with voting. To diagnosis of thyroid disease using superior Fuzzy K-NN classifier was proposed by Liu *et al.*

#### Merits

- Training is faster.
- Easy to understand and implement

## Demerits

- Large storage space is required.
- Noise sensitivity is more.
- Slow Testing.
- For large samples computational costs are expensive.
- Data is very sensitive for limited structure.

## Challenges

- The accuracy of the KNN to be maintain.
- Maintain closeness graphs to the KNN problem [24].

## 2.4 Support Vector Machine (SVM)

This technique is based on statistics. It uses the class of kernel based methods. SVM attempts to find a linear separator (hyper-plane/ hyper line) between the data samples belonging to two classes in multidimensional area. This hyper plane is called as an optimal hyper plane. A set of instances closest to the optimal hyper plane is known as support vector[18]. Finding the optimal hyper plane is nothing but a linear classification. It is used to finds connections between features and redundant features. By increasing the dimensionality of the feature space, a nonlinear SVM can be created. It is possible to find a separating the hyper plane in a higher dimensions [15]. The type of kernel can be linear kernel or dot-product kernel, polynomial kernel and Gaussian Kernel. The special case of Radial Basis Function (RBF) kernel is Gaussian Kernel. Normally Euclidean distance is used by algorithm [47]. The width of RBF kernel is determined by parameters and distance metric [16].

## Merits

- Accuracy is better as compare to other classifiers.
- Complex nonlinear data can be easily handled.
- As compared to other methods over fitting problem is observed very less.

## Demerits

- It is computationally expensive.
- Selection of right kernel function is important as per application.
- Training process is time consuming.
- It is designed especially for solving binary class problem.
- To solve multi class problem divide it into couple of classes

## Challenges

- It is challenging to find best kernel for particular application
- SVM is very expensive in terms of execution and space

## 2.5 Naive Bayesian

Bayesian classification is the classification based on Bayesian theory. Bayesian Belief Networks (BBN) works on the basis of Bayesian classification. It accepts that all attributes do not dependent on one another. In health area all features such as symptoms and their health condition are interconnected with each other. Still Naive Bayesian classifier shows good presentation in conditions of truthfulness. It is concentrates on prior, posterior and discrete probability distributions of sample data. It has widely used in healthcare field. A decision support system is developed using Bayesian Belief Networks for analyzing risks with medical data [62]. The psychiatric patient data can also be analyzed using Bayesian Belief Networks in making significant decision for psychiatric patients [63].

## Merits

- Computation process is easier.
- For huge datasets, better speed and accuracy is obtained.
- It is well suited to continuous valued attributes.

## Demerits

- It does not guarantee accurate results where dependency among variables exists.
- Training time is more.
- Interpretability is poor.
- It requires number of parameters like topology or structure of network.

## Challenges

- Incorporation of prior knowledge
- Validation and insight
- Learning causal interactions

## 3. LITERATURE REVIEWS

Major research scope is there in classification of medical dataset for accurate diagnosis. This section focuses on some recent developments and necessary information about medical data classification.

According deal with issues of high dimensionality in health data sorting, Yonghong Peng et al. [44] proposed feature selection methods. Feature selection methods are classified in three types: filter, wrapper and hybrid approaches. Filter methods apply an independent test without any learning algorithm, while a pre-determined learning algorithm is used by wrapper method. Filter and wrapper both methods have their advantages and disadvantages. Filter approach is having low computational cost but less reliability in classification while wrapper method has better classification accuracy with more computational cost. The hybrid method combines advantages of first two methods.

A. S. Galathiya et al. [45] have proposed medical data classification using a decision tree. The Comparison was

done between ID3, C4.5 and C5.0. Here feature selection, cross validation, reduced error pruning and model complexity are used. The classification accuracy is increased as compared to the existing system in less time.

M.A. Jabbar et al. [46] used association rules in medical data mining to make accurate decisions. The registrar general of India shows that the coronary heart disease is a major cause of death in our India which causes about 30% death in rural areas.

Mouhamadou Lamine Samb et al. [47] have proposed the modified SVM recursive with feature elimination (SVM-RFE). Here RFE-SVM algorithm is combined with local search operators based on artificial intelligence. The experiments are done on data sets from UCI repository. The result shows the contribution of the local search on the classification. This approach reuses features previously deleted during the RFE-SVM process which improves the quality of the final classifier.

Vipin Kumar et al. [48] proposed the method for categorization and compared existing feature selection methods. Also the guidelines are provided for user to select a feature selection algorithm.

Dimensionality reduction based data analysis is proposed by Ludmila et al. [49]. An adaptive classification method is implemented to check class level of the input data. The proposed PCA based system finds Eigen matrix and Eigen vector for dimensionality reduction. Due to proposed technique, redundant data from original data input reduces dataset size.

A. Sudha et al. [50] have proposed a principle component analysis algorithm which was used for reducing the dimensions and determining the attributes involving more towards the prediction of stroke disease and predicts whether the patient is suffering from stroke disease or not. The stroke was a leading cause of serious, long term disability in US.

Indu Saini et al. [51] have proposed algorithm which was evaluated on two, manually annotated standard databases such as CSE and MIT-BIH Arrhythmia database. The false detection caused by interference in ECG signal is reduced by using digital band-pass filter and the gradient of the signal was used as a feature for QRS-detection. Also KNN based classifier accuracy is dependent on the value of K and type of distance metric.

Sina Khanmohammadi and Mandana Rezaeiahari [52] have proposed a machine learning classification that has been used for developing clinical decision support system (CDSS). They used ten sample medical datasets. Results suggest the SVM model as the most desirable classification algorithm for developing CDSS. The research was not only to identify a classification algorithm that has been performing best in all medical datasets but also reliability of the model.

The major issue in medical data analysis have found by Emanuel Weitschek et al.[53] as missing values, different adopted measure scales and integration of the disparate collection procedures. Hence managing clinical data, discovering patients interactions, and integrating the different data sources are the main challenges. The major challenge is to get relevant information from huge data.

R. Chitra and V. Seenivasagam [54] have proposed work on heart disease prediction technique in category wise. Here Neural Networks have been utilized to make predictions for medical data.

R. Bhuvaneswari and K. Kalaiselvi [55] have proposed the use of Naive Bayes classifier in medical applications for quality service that diagnose patients correctly and administer treatments which are effective. Decision Support in heart disease prediction system was also developed using NB Classification technique. NB classifications have been used as a best decision support system.

A weighted fuzzy rule-based medical decision support system (CDSS) was proposed by P.K. Anooj [56] for heart disease diagnosis for getting record of patient's clinical data.

M. Akhil Jabbar et al. [57] have proposed a combined approach of KNN and genetic algorithm for improving classification accuracy of heart disease data set. Proposed system used genetic search to prune redundant and irrelevant attributes also to find attributes which are important for classification. Least ranked attributes are removed, and classification is done depending on high ranked attributes.

B. Dennis and S. Muthukrishnan [58] have proposed an efficient medical data classification system based on Adaptive Genetic Fuzzy System (AGFS). In proposed method first rules are generated from data and optimized rules selection is performed adapting of genetic algorithm. A simple technique for scheming of membership function, discretization, and a fitness function is designed.

C. V. Subbulakshmi and S. N. Deepa [59] have proposed a machine learning paradigm based hybrid methodology. This paradigm integrates the successful exploration mechanism called self-regulated learning capability of the particle swarm optimization (PSO) algorithm with the extreme learning machine (ELM) classifier. As a recent off-line learning method, ELM is a single-hidden layer feed forward neural network (FFNN), proved to be an excellent classifier with large number of hidden layer neurons. PSO is used to determine the optimum set of parameters for the ELM, thus reducing the number of hidden layer neurons, and it further improves the network generalization performance.

J. Jayanth [60] have proposed swarm intelligence technique in satellite data classification as the traditional statistical classification technique shows limited success in classifying remote sensing data. The traditional statistical classifiers

ignores the spatial distribution of the pixels and only examines the spectral variance and correlation between various bands. The Artificial Bee Colony (ABC) algorithm based upon swarm intelligence is used for extracting information on the basis of object recognition and classification in several areas. The results show that ABC algorithm has 5% improvement over the traditional Maximum Likelihood Classifier (MLC) in accuracy and Artificial Neural Network (ANN) and 3% against SVM.

Accordingly data mining techniques have shown much enhancement in health records sorting with respect to data related to different diseases like swine flu, dengue, tumour, cardio vascular abnormalities, diabetes, and etc. Table 1 summarizes mining algorithms used in medical data, their application areas and the purpose.

**Table 1:** Summary of classification techniques in medical data classification

Sr. No.	Techniques	Purpose	Disease
1	Decision Tree[17][23][24]	Decision Support	Cardio Vascular Diseases
2	ANN [64]	Extracting patterns, Medical data classification	Swine flu, Dengue
3	k-NN [24]	Classification of diseases	Diabetes, Cancer
4	SVM [16][31]	Disease Classification	Diabetes
5	Naive Bayesian [13][26]	Improving classification accuracy	Coronary Heart Disease, Liver disease

#### 4. COMPARATIVE STUDY OF VARIOUS OPTIMIZATION TECHNIQUES

The performance of the selected algorithms was evaluated on the nine publicly available datasets shown in Table 2 which for each dataset shows the estimated classification accuracy of the five algorithms. From Table, we can observe that the classification accuracy of the C5.0 algorithm tends to be better in five datasets out of the nine. Also it can be seen that CART and SVM shows relatively low overall classification accuracy in all the datasets. With the Cars and NHANES dataset, Naive Bayes (NB) has the best classification accuracy among other techniques. For the Vehicle dataset, neural network (NN) performs best. Based on the results, one can argue that the C5.0 algorithm is the best algorithm because of its performance in five out of the nine dataset; similarly NB accuracy is better in two of the nine datasets (NHANES and Cars), but it is inferior among all algorithms in four datasets. Some of the classifiers show good to excellent results on some of the datasets but performs less on most of the datasets. For example all the classifiers show excellent results on the Segment and the

Cars dataset and relatively superior piece on the adult dataset and relatively poor piece is observed on the Red wine, White wine and the NHANES datasets.

**Table 2:** A comparison of the accuracy of NN, NB, SVM, C5.0 and Classification and Regression Tree (CART) on the nine datasets [65]

Datasets	NB	CART	C5.0	NN	SVM
Adult	66.98	84.89	86.24	85.44	84.62
House	66.05	73.38	90.97	75.03	69.92
Credit	62.33	66.67	64.33	74.00	70.33
Segment	76.08	96.19	97.40	96.01	92.59
Cars	93.52	87.04	92.82	91.44	92.59
White Wine	47.47	54.90	66.01	57.35	49.84
Red Wine	56.14	55.14	67.67	57.39	60.40
NHANE S	64.29	61.31	63.39	63.99	63.99
Vehicle	12.41	70.92	75.53	83.33	76.6

From the table 2 we can say that the C5.0 model normally performs well for all the datasets, while the performance of the other classifiers depends on the data used. For example, the NB obtains reasonable results. The Naive Bayes algorithm shows relatively poor results for the Adults and Segment datasets but generates terrible results when using vehicle dataset.

#### 5. CONCLUSION

In this survey we acknowledged and evaluated the most commonly used records removal algorithms. To ensuing as well-performing on health databases, based on recent studies. These algorithms are very useful in health records removal, if it is a big data in this trade. According to the fast growth of health statistics. The use of DMT for decision support and predication systems in the field of healthcare. It has provided the summary of records removal optimization techniques for health records categorization further the diseases they confidential.

#### REFERENCES

- [1] Jawai Han, MichellineKamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, 3<sup>rd</sup>Edition.
- [2] Sharma S, Agrawal J, Agarwal S, Sharma S, "Machine Learning Techniques for Data Mining: A Survey", Computational Intelligence and Computing Research (ICIC), IEEE International Conference on 26-28 Dec. 2013 Page(s):1 - 6, 2013
- [3] Shweta Kharya, "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease", International Journal of Computer Science, Engineering and Information Technology, vol.2, no.2, pp., April 2012.
- [4] Abdülkadir Çakır & Burçin Demirel, "A Software Tool for Determination of Breast Cancer Treatment Methods Using Data Mining Approach", Springer Science and Business Media, J Med Syst ,Published online, pp.1503–1511, 2011.

- [5] K. Rajesh, and V. Sangeetha, "Application of Data Mining Methods and Techniques for Diabetes Diagnosis", *International Journal of Engineering and Innovative Technology*, vol.2no.3, pp.224-229, Sep 2012.
- [6] Qasem A. Al-Radaideh and Eman Al Nagi, "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance", *International Journal of Advanced Computer Science and Applications*, vol.3,no.2, pp.144-151, 2012.
- [7] Bartosz Krawczyk and Gerald Schaefer, "Ensemble fusion methods for medical data and classification", In proceeding of 11th symposium on neural network application in electrical engineering, pp.143-146, 2012
- [8] Markus Brameier and Wolfgang Banzhaf. "A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining", *IEEE*. p1-10. 2001
- [9] Doron Shalvi and Nicholas DeClariss., "An Unsupervised Neural Network Approach to Medical Data Mining Techniques". *IEEE*. 0 (0), p1-6.
- [10] Adepele Olukunle and Sylvanus Ehikioya, "A Fast Algorithm for Mining Association Rules in Medical Image Data". *IEEE*. p1-7.
- [11] Cindy L. Bethel and Lawrence O. Hall and Dmitry Gold, "Mining for Implications in Medical Data", *IEEE*. p1-4.
- [12] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [13] Jeong-Yon Shim, Lei Xu, "MEDICAL DATA MINING MODEL FOR ORIENTAL MEDICINE VIA BYY BINARY INDEPENDENT FACTOR ANALYSIS", *IEEE*. p1-4.
- [14] Jenn-Lung Su, Guo-Zhen Wu, I-Pin Chao, "THE APPROACH OF DATA MINING METHODS FOR MEDICAL DATABASE", *IEEE*. p1-3 2001.
- [15] Safwan Mahmud Khan Md. Rafiqul Islam Morshed U, "Medical Image Classification Using an Efficient Data Mining Technique", *IEEE*, p1-6.
- [16] Yanwei Xing, Jie Wang and Zhihong Zhao, "Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease", *IEEE*. p1-5, 2007.
- [17] Tsang-Hsiang Cheng, Chih-Ping Wei, Vincent S. Tseng, "Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches", *IEEE*. p1-6.
- [18] Mohammad Saraei, George Koundourakis, Babis Theodoulidis, "EASYMINER: DATA MINING IN MEDICAL DATABASES", *IEEE*. p1-3.
- [19] SAM CHAO, FAI WONG, "AN INCREMENTAL DECISION TREE LEARNING METHODOLOGY REGARDING ATTRIBUTES IN MEDICAL DATA MINING". Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.
- [20] My Chau Tu AND Dongil Shin, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms", *IEEE*. p1-5, 2009.
- [21] Vili Podgorelec, Marjan HerikoMaribor "Improving Mining of Medical Data by Outliers Prediction", *IEEE*. p1-6.
- [22] Shuyan Wang Mingquan Zhou Guohua Geng, "Application of Fuzzy Cluster Analysis for Medical Image Data Mining. *IEEE*. p1-6.
- [23] Asha Gowda Karegowda M.A.Jayaram, "Cascading GA & CFS for Feature Subset selection in Medical Data Mining", *IEEE*. p1-4. 2009.
- [24] Graduate Institute of Applied Information Sciences (2009). MEDICAL DATA MINING USING BGA AND RGA FOR WEIGHTING OF FEATURES IN FUZZY K-NN CLASSIFICATION. *IEEE*. p1-6.
- [25] Weimin Xue, Yanan Sun, Yuchang Lu, "Research and Application of Data Mining in Traditional Chinese Medical Clinic Diagnosis", *IEEE*. p1-4.
- [26] Ranjit Abraham, Jay B.Simha, Iyengar, "A comparative analysis of discretization methods for Medical Datamining with Naïve Bayesian classifier", *IEEE*. p1-2.
- [27] Syed Zahid Hassan and Brijesh Verma, "A Hybrid Data Mining Approach for Knowledge Extraction and Classification in Medical Databases", *IEEE*. p1-6.
- [28] Michele Berlingerio, "Mining Clinical Data with a Temporal Dimension: a Case Study", *IEEE*. p1-8.
- [29] Wojciech Froelich, Alicja Wakulicz-Deja, "Mining Temporal Medical Data Using Adaptive Fuzzy Cognitive Maps". *IEEE*. P1-8. 2009
- [30] Michael Barnathan, Jingjing Zhang, Vasileios, "A WEB-ACCESSIBLE FRAMEWORK FOR THE AUTOMATED STORAGE AND TEXTURE ANALYSIS OF BIOMEDICAL IMAGES", *IEEE*. p1-3. Khaleel et al., *International Journal of Advanced Research in Computer Science and Software Engineering* 3(8), August - 2013, pp. 149-153 © 2013, IJARCSSE All Rights Reserved Page / 153
- [31] Sarojini Balakrishnan, "SVM Ranking with Backward Search for Feature Selection in Type II Diabetes Databases", *IEEE*. p1-6.
- [32] Sofianita Mutalib, et al, "Intelligent Classification in Medical Data", in proceeding of IEEE EMBS International Conference on Biomedical Engineering and Sciences, pp.120-124, Dec 2012.
- [33] Vikas Chaurasia and Saurabh Pal, "Data Mining Approach to Detect Heart Dieses", *International Journal of Advanced Computer Science and Information Technology*, vol.2, no.4, pp.56-66, 2013
- [34] Hanaa Ismail Elshazly, Abeer Mohamed Elkorany, Aboul Ella Hassanien and Ahmad Taher Azar, "Ensemble classifiers for biomedical data: performance evaluation", in proceeding of 8th International Conference on Computer Engineering & Systems (ICCES), pp.184 - 189, Nov. 2013.
- [35] Sivapriya TR and Nadira Banu Kamal AR, "Hybrid Feature Reduction and Selection for Enhanced Classification of High Dimensional Medical Data", in proceeding of IEEE International Conference on

- Computational Intelligence and Computing Research (ICCIC), pp.1-4, Dec 2013.
- [36] Pushpalata Pujari and Jyoti Bala Gupta, "Improving Classification Accuracy by Using Feature Selection and Ensemble Model", International Journal of Soft Computing and Engineering, vol.2, no.2, pp.380-386, May 2012.
- [37] Aqueel Ahmed and Shaikh Abdul Hannan, "Data Mining Techniques to Find Out Heart Diseases: An Overview", International Journal of Innovative Technology and Exploring Engineering, vol.1, no.4, pp.18-23, Sep 2012,
- [38] Prof.K.Rajeswari, Dr.V.Vaithyanathan and Shailaja V.Pede "Feature Selection for Classification in Medical Data Mining", international conference of emerging Trent and technology in computer science, vol.2, no.2pp. April 2013.
- [39] V. Sneha Latha, P.Y.L.Swetha, M.Bhavya, G. Geetha and D. K.Suhasini, "Combined Methodology of the Classification Rules For medical Data-Sets", International Journal of Engineering Trends and Technology, vol.3, no.1, pp.32-36, 2012.
- [40] Mohammed Abdul Khaleel, Sateesh Kumar Pradham and G.N. Dash, A "Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases", International Journal of Advanced Research in Computer Science and Software Engineering, vol.3, no.8pp., August 2013.
- [41] Divya Tomar and Sonali Agarwal, "A survey on Data Mining approaches for Healthcare" International Journal of Bio-Science and Bio-Technology, vol.5, no.5, pp.241-266, 2013.
- [42] Lamia AbedNoor Muhammed , "Using Data Mining technique to diagnosis heart disease", in proceeding of statistic in science, business and engineering, pp.1-3, Sep 2012.
- [43] Patil D.V, Bichkar R.S, "Issues in Optimization of Decision Tree Learning: A Survey", International Journal of Applied Information Systems (IJAIS) Foundation of Computer Science FCS, New York, USA Volume 3 – No.5, July 2012.
- [44] Yonghong Peng, Zhiqing Wu, Jianmin Jiang, "A Novel Feature Selection Approach for Biomedical Data Classification", School of Informatics, University of Bradford, UK Journal of Biomedical Informatics 43 (2010) 15–23.
- [45] A.S. Galathiya, A. P. Ganatra and C. K. Bhensdadia, "Improved Decision Tree Induction Algorithm with Feature Selection, Cross Validation, Model Complexity and Reduced Error Pruning", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012.
- [46] M.A.Jabbar, Dr Priti Chandra and Dr.B.L. Deekshatulu, "Knowledge Discovery From Mining Association Rules For Heart Disease Prediction", Journal of Theoretical and Applied Information Technology, vol.41, no.2, pp., July 2012.
- [47] Mouhamadou Lamine Samb, Fode Camara, Samba Ndiaye, Yahya Slimani and Mohamed Amir Esseghir, " A Novel RFE-SVM-based Feature Selection Approach for Classification", International Journal of Advanced Science and Technology Vol. 43, June, 2012.
- [48] Vipin Kumar and Sonajharia Minz, "Feature Selection: A literature Review", Smart Computing Review, vol. 4, no. 3, June 2014.
- [49] Ludmila I. Kuncheva, Member, IEEE, and William J. Faithfull, "PCA Feature Extraction for Change Detection in Multidimensional Unlabeled Data", IEEE Transactions on Neural Networks and Learning Systems, Vol. 25, No. 1, January 2014.
- [50] A.Sudha, P.Gayathri and N.Jaisankar, "Effective Analysis and Predictive Model of Stroke Disease using Classification Methods", IEEE International Journal of Computer Applications, vol.43, no.14, pp.26-31, April 2012.
- [51] Indu Saini, Dilbag Singh and Arun Khosla, "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases", Elsevier Journal of Advanced Research, vol.4, no.4, pp.331-344, July 2013.
- [52] Sina Khanmohammadi and mandana rezaeiahari, "AHP based classification algorithm selection for clinical decision support development", Elsevier Procedia Computer Science, Vol. 36, pp. 328-334, 2014.
- [53] Emanuel Weitschek, Giovanni Felici, and Paola Bertolazzi, "Clinical data mining: problems, pitfalls and solutions", IEEE 24th International Workshop on Database and Expert Systems Applications ,Pages 90-94, 2013
- [54] R. Chitra and V. Seenivasagam, "Review Of Heart Disease Prediction System Using Data Mining And Hybrid Intelligent Techniques", ICTACT Journal On Soft Computing, , vol. 03, no. 04, pp.605-609, July 2013.
- [55] R. Bhuvaneshwari and K. Kalaiselvi, "Naive Bayesian Classification Approach in Healthcare Applications", international Journal of Computer Science and Telecommunications, Vol. 3, no. 1, pp.106-112, January 2012.
- [56] P K. Anooj. "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," Elsevier Computer and Information Sciences, Vol. 24, no. 1, pp. 27-40, January 2012.
- [57] M. Akhil Jabbar, B.L. Deekshatulu and Priti Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm" Elsevier Procedia Technology, Vol. 10, pp. 85-94, 2013.
- [58] B. Dennis and S. Muthukrishnan, "AGFS: Adaptive Genetic Fuzzy System for medical data classification", Applied Soft computing, Vol. 24, pp. 242-252, 2014.
- [59] C. V. Subbulakshmi and S. N. Deepa, "Medical Dataset Classification: A Machine Learning Paradigm Integrating Particle Swarm Optimization with Extreme Learning Machine Classifier", The Scientific World Journal March 2015.
- [60] J. Jayanth , Shivaprakash Koliwad, Ashok Kumar T , "Classification of remote sensed data using Artificial

- Bee Colony algorithm” ,National Authority for Remote Sensing and Space Sciences. Production and hosting by Elsevier, Available online 7 April 2015.
- [61] Beatriz A. Garro, Humberto sossa and Roberto A. Vazquez, “Artificial neural network synthesis by means of artificial bee colony algorithm” IEEE-2011.
- [62] Yu, J., Xiand, L., Wang, S.: An Improved Particle Swarm Optimization for Evolving Feedforward Artificial Neural Networks. Neural Processing, vol. 26, pp. 217–231, 2007.
- [63] Celal Ozturk and Dervis Karaboga ”Back Propagation Algorithm for Neural Network Training”, IEEE 2012.
- [64] Saroj Kr. Biswas, “An ANN Based Pattern Classification Algorithm for Diagnosis of Swine Flu”, Journal of Intelligent Computing Volume 5 Number 1 March 2014.
- [65] Niranjan J. Chatap Ashish Kr. Shrivastava “ A Survey on Various Classification Techniques for Medical Image Data” International Journal of Computer Applications (0975 – 8887) Volume 97– No.15, July 2014