IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.7, Special Issue.3, Feb. 2019 E-ISSN: 2347-2693

Machine Learning: An Effective Technique for Health Data Classification

Nishant Behar[1]*, Manish Shrivastava[2],

[1,2]CSE, GGV, Bilaspur India

*Corresponding Author: nishant.itggv@gmail.com, Tel.: +91-98271-79525

Available online at: www.ijcseonline.org

Abstract- Health is precious for every life. But there are many diseases which fall under the category of dangerous or critical due to it mortality rate. Such diseases can be cured or at least prevented if they are identified in their earlier stages. For the proper diagnosis of these diseases, data mining techniques using machine learning methods- k-NN, Naïve Base, Decision trees, Support Vector Machine plays very significant role. In this paper the focus was on finding techniques applied for the common disease classifications, the accuracy of methods reported, dataset used and pros and cons of these methods and concluded with the open challenges and opportunities for further research in health care sector.

***Keywords*—**Machine Learning, Data, classification

I. INTRODUCTION

"**Health is wealth**" this quotation indicates how health is important for a person in his life. But large number of population is suffering from various kinds of health problems. For treatment, patients have to visit healthcare centres, where diagnosis of disease and suitable treatments are prescribed. For effective treatment perfect diagnosis and best treatment is required and this still needs to improve. Nowadays, healthcare devices are used to record, report and monitor patient's health status data. These devices contain lots of healthcare information about patient's health condition, disease characteristics and medical test reports. Healthcare Information is in different sizes and shapes. Main sources of medical data are healthcare equipments, telemetry healthcare systems, E-healthcare, manual prescriptions, Biosensors, Physiological tests, Wearable devices etc. These different sources make the data not only heterogeneous, complex, but sometime produced data are insufficient for proper decision, incomplete and have missing values, as a result it becomes difficult to comprehensively interpret disease analysis [1]. For example for better diagnosis of a patient's disease , health status of the patients, and other health related information like history of patient, his/her food habit, living style, surrounded environment, and its treatment records make significance value. For such deep analysis, healthcare experts require effective and efficient tool. From last few decades data mining and machine learning are being considered as a powerful tool for diagnosis diseases. Data mining techniques like classification, prediction, regression, clustering, anomaly detection etc. are widely used in the area of healthcare. The classification applied on healthcare datasets provides automated diagnosis system, but high featured datasets analysis is currently required.

In this paper we have focused on the various popular classification techniques which are being used for the healthcare related researches for disease diagnosis or predictions. We have mentioned about classification techniques K-Nearest neighbour (K-NN), Naïve Bays (NB), Decision Trees (DT), Support Vector Machine (SVM) and evolutionary technique Artificial Neural Network (ANN) so that the researchers can concentrate on a single method with best accuracy for effective classification.

.

II. RELATED WORK

A. DATA MINING TECHNIQUES

Data mining is the way to gain knowledge from noisy, inconsistent and incomplete raw data [2]. Data mining are used for two purposes (a) for the prediction (b) for the descriptive analysis. Analyzing and extracting useful information from the large medical datasets, data mining techniques are well suited techniques. Various techniques namely association, classification, prediction, sequential patterns, decision trees, K-Nearest Neighbour, Decision tree, Support Vector Machine (SVM), Neural Network (NN), Bayesian Methods [3] data mining techniques are used for pattern extraction for the purpose of disease diagnosis, monitoring and patient related monitoring [4] Data mining have five major steps i.e. (a) Data selection, (b) Pre-processing of data (c) Transformation (d) Interpretation and (e)Evaluation [5] as shown in Fig -1.

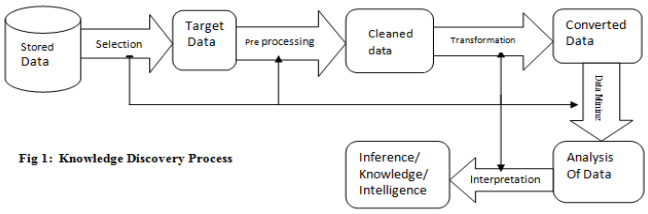

Fig 1: Knowledge Discovery Process

Fig. 1

B. CLASSIFICATION

Classification process selects each unlabeled instance from a test dataset and categorizes it in proper class using a classification model such that classification error decreases [6]. The machine learning methods used for these purpose, has two basic steps as shown in Fig. 2 (step 1) Building classification model on the basis of training dataset Fig (step 2) and using built model for new and test dataset for the proper classification of each instance. Continuous respond attribute for a dataset are used for prediction whereas categorical or discrete respond attribute are known as classification process. Data mining and machine learning techniques are dependent on classification type supervised and unsupervised learning [7]. In this paper we have considered both as classification technique.

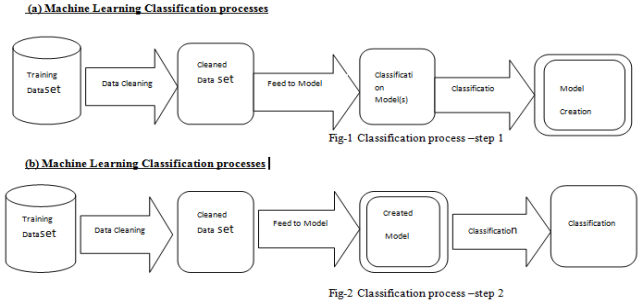

Fig. 2

Classification techniques are very appropriate and fit for the health and medical domain and for disease diagnosis/predictions of large dataset [5]. Classification techniques provides model which helps to classify categorical class data for a particular data set [8][2].

In the research work reported, each sample of classifier has been assigned weights dynamically according to desired parameters [9]. And they developed a model based on dynamic weighted voting system to use multi-classifier and for improvement in its local and global accuracy to detect type II diabetes. Two datasets were used for experimental purpose i.e. Indonesia patient (IP) and Pima Indians Diabetes dataset.

In the paper published recently, authors reviewed the machine learning(ML) techniques for disease diagnostic and provided the comparative analysis of different ML for the diagnosis of different diseases like heart, liver, dengue and hepatitis disease. They found ML techniques a worthy approach for making classy and automatic algorithms. This paper discussed ML tools and provide support for analysis of disease and decision making process accordingly. This review also shows that statistical models are unsuccessful to hold categorical data, deal with missing values and larger data points[10] .

1) *k-Nearest Neighbour*: In the K-nearest neighbour technique, unknown training points are taken and nearest distance (weights) from known sample points are found [6].This distance are calculated by finding Euclidean or Hamming or Manhattan distance. Minimum (Nearest) distance gives the class on which this unknown point belongs. Researchers made a hybrid model having K-NN with GA for prediction of the heart disease complications for the Diabetes –II class patients [11] and its result confirms that proposed hybrid machine learning knowledge model improves accuracy of kNN as better classifier.

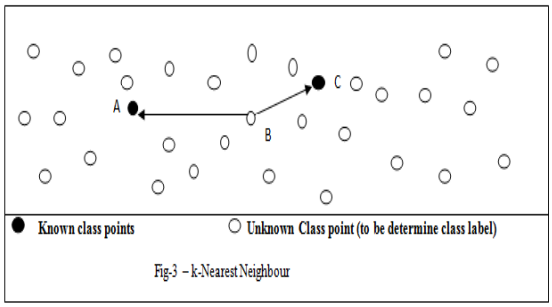

Fig. 3

2) *Decision Tree:* Traversing in decision tree starts from the root node (starting node) and reach to one predetermined leaf node (decision node). During flow from root node to leaf node various intermediate selection on the basis of Yes/No are made. It looks like "nested if-than-else" construct of flow charts. The research work proposed an algorithm which can classify disease over distributed medical datasets. Proposed algorithm enhances traditional decision tree. The advantage of this work was it requires only fraction of the attributes as input instead of all statics of the attributes for the classifier. It reduces the communication overheads and network bandwidth requirement for taking decision [12].

In another experimental work, they have implemented Decision Tree on the filtered and selected DNA dataset by t-GA for the diagnosis of cancer disease. They also found high accuracy of 77.31% [13]. Researchers applied various techniques, Decision Tree, SVM and MLP for the prediction of Chronic Hepatitis on the Single Nucleotide Polymorphism (SNP). In the result decision tree has classified with maximum accuracy of 73.20 % [14]. Authors found that the Classification method Decision tree and naïve bays gave 79.01% accuracy but performance of SVM decreased when PCA was used [15].

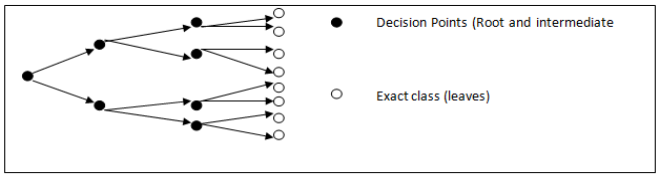

Fig-4 Decision Tree

Fig. 4

3) *Support Vector Machine (SVM):* SVM has two basic properties, (a) to make maximum distance of hyper planes between support vectors of different classes and (b)to separate data into positive and negative class more accurately [16][17]. Also this method is computationally costly on large data set and very difficult to identify the major hyper plane from transformed high dimensional space [18][19]. In the research, a modified Adaptive SVM Model was developed by changing standard SVM bias weights into adaptive SVM structure [20]. The performance reported 100% accuracy for Diabetes and Breast Cancer.

4) *Neural Network (a) Multilayer Perceptron:* **ANN** is a prearranged interconnection of nodes in the form of input, output and hidden layers to imitate human brain. Authors compared in their research various classification techniques Decision Tree, MLP, Naïve Bayes, SMO and instant based for K-NN (IK-NN) on various datasets and found that MLP and J48 hybrid model with PCA performed best among others [7]. (b) **BPNN:** The output values of ANN (errors) are propagated back to its previous hidden layers to readjust the weight and neuron values are updated accordingly.

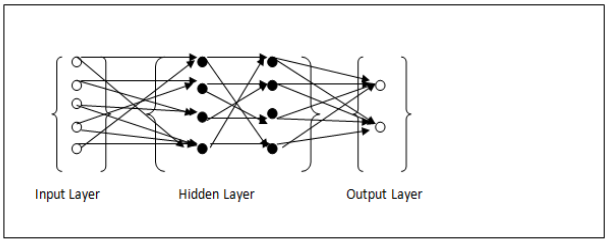

Fig-5 Artificial Neural Network

Fig. 5

Sequential Multilayered perceptron (SMPC) with Back Propagation Neural Network (BPNN) for the prediction of Diabetes on the time associated risk features, researchers found this ANN model performing best as compared to baseline NN and regression methods with respect of gain and sensitivity[21]**.** In another work [22] BPNN used for classification of Heart Disease among eight different types and their experiment showed accuracy of 98.37%**.** A hybrid model of ANN and Fuzzy NN (FNN) for accurate classification of Diabetes on different databases was also used [23]. Authors concluded that hybrid model have accuracy of 84.24% for diabetes and 86.8% for heart disease. A research was carried using SVM, Extreme Learning Machine [25] (ELM), BPNN, were applied for the classification of Arrhythmia (Heart beats). In experimental results, authors [24] found ELM performed best among these and also addressed the slow learning speed of ANN and its unstable performance caused by local minima. In a research work, two types of ANN i.e. MLP and Radial Bases Function (RBF) for the classification between healthy people and people having Parkinson's disease [26]. In these experimental results they have got classification accuracy of 94.72%. A research paper concluded that performance of neural classifiers are strongly depends on domain specific and architecture of neural classifier also affects the accuracy of results [27].

5) *Bayesian Methods:* The Bayesian regression method shows whether given object is similar or not, among the set of predefined classes. Bayesian method is powerful for binary and multiclass and need to exploit more at full extent [28][29]. The main advantage of this network is that it requires less pre-existing knowledge of variable dependencies [30]. A investigation on different classification algorithms such as Decision Tree, Naïve Bays, K-NN and ANN on the Heart disease dataset and found that Naïve Bays with feature selection CFS method gave best results compared to other techniques with highest accuracy 83.70% [31],.

Table 1 : Summary of accuracy as reported by authors

Authors	Classification Method Description	Disease	Accuracy
[32]	Principal component analysis (PCA) and adaptive neuro-fuzzy inference system (ANFIS).	Diabetes	89.47%
[33]	Feature selection with ranking supervised model construction (FSSMC), with optimisation of ReliefF , Naıve Bayes, IB1 and C4.5 methods	Diabetes-II	95%
[34]	Stand-Alone Neural n Network classifier Model	Erythemato-squamous diseases	97.8%
		Lung cancer	99.99%
		.Lymphography, entropy minimization	86.2%

		.Diabetes error rate reduced	From 33% to 24%
[35]	Fuzzy k-NN, and the adaptive network based model	Wisconsin Breast Cancer	97.17%
		Diabetes dataset	77.65%
[36]	Generalized Discriminate Analysis (GDA) and Least Square Support Vector Machine (LS-SVM) with 10 fold	LS-SVM(Diabetes)	78.21%
		GDA–LS-SVM (Diabetes)	82.05%
[23]	Hybrid artificial neural network (ANN) and fuzzy neural network (FNN)	Diabetes	84.24%
		Heart dieses	86.8%
[37]	Classification and regression tree (CART) and case-based reasoning (CBR) techniques	Liver disease	92.94%
[38]	Input partitioning mechanism, generated hierarchical fuzzy rule, multiple consequent fuzzy rules	Breast Cancer	96.08%
		Diabetes	92.26%
		Liver	89.9%
[39]	Simple K-Means Clustering Algorithm and C4.5 algorithm	Diabetes type II	92.38%.
[40]	New feature selection method based on fuzzy entropy measures	Diabetes	75.29%
		Parkinson's	85.03%
		Dermatology	98.28%
		Breast cancer	97.49%
[41]	Rotation Forest (RF) ensemble classifiers of 30 machine learning algorithms	Diabetes	74.47%
		Heart	80.49%
		Parkinson's	87.13%
[42]	Ant Colony Optimization (ACO) with fuzzy classification	Diabetes	84.24%
[29]	Hybrid approach (pair wise comparison, Bayesian regression and the *k*-NN)	Breast cancer	97.74%
[43]	SVM	Physicochemical properties (ROC score)	0.93462
[44]	Artificial Bee Colony	Diabetes	84.21%
[20]	Modified SVM (SVM with adaptive)	Diabetes and Breast cancer	100%
[45]	Hybrid Model : Particle Swarm Optimization for classification with10-fold cross-validation	Diabetes Type II diabetes	92%,
[46]	Modified K means & SVM	Heart	97.87 %,
		Diabetes	96.71 %
[47]	Hybrid model(Modified k-NN and Developed BPNN)	Medical data (heart diseases, breast cancer diabetes)	77.40
[48]	Hybrid prediction model(using K-means clustering with Multilayer Perceptron)	Diabetes	99.82%
		Hepatitis	99.08%
		Breast Cancer	99.39%,
[49]	Hybrid Model using fuzzy ART MAP (FAM) neural network and the classification and regression tree (CART)	Breast Cancer	98.86%
		Diabetes	87.64%
		Hepatitis	96.87%
		Parkinson	91.45%
		Heart disease	100%
		Liver Disorder	94.41%

Table 2: Data mining techniques applied on major disease	
Disease	Technique
Diabetics	Hybrid Adaptive Neuro-fuzzy inference system, NaıveBayes, IB1 and C4.5, fuzzy k-NN ,ANN, Generalized Discriminate Analysis (GDA) and Least Square SVM (LS-SVM), Ant Colony Optimization (ACO) with fuzzy, Hybrid ANN and fuzzy neural network (FNN), Hierarchical fuzzy rule, K-Means Clustering + C4.5 algorithm, Fuzzy entropy Measures, Rotation Forest (RF), Artificial Bee Colony(ABC), SVM with adaptive algorithm, Particle Swarm Optimization, Modified K means & SVM, Hybrid K-means clustering with Multilayer Perceptron, fuzzy ART MAP (FAM) NN and the classification and regression tree (CART)
Heart Diseases	Hybrid ANN and fuzzy NN, Rotation Forest (RF), Modified K means & SVM, Modified k-NN and BPNN, Fuzzy ART MAP (FAM) NN and classification and regression tree (CART)
Parkinson's Diseases	Fuzzy entropy measures, Rotation Forest (RF), Fuzzy ART MAP (FAM) NN and classification and regression tree (CART)
Liver Diseases	Classification and regression tree (CART) and case-based reasoning (CBR) techniques, generated hierarchical fuzzy rule, Fuzzy ART MAP (FAM) NN and CART
Breast Cancer	Fuzzy entropy measures, Hybrid Bayesian regression and *k*-NN, SVM with adaptive, Modified k-NN and BPNN, Hybrid K-means clustering MLP, Fuzzy ART MAP (FAM) NN and classification and regression tree (CART)
Dermatology	Fuzzy Entropy Measures
Hepatitis	Hybrid K-means clustering MLP, Fuzzy ART MAP (FAM) NN and classification and regression tree (CART)
Lung Cancer	ANN

III. CONCLUSION AND FUTURE SCOPE

In this review, we observed that the traditional stastical methods are not suffcient for the disease classissification therfore machine learning techniches are used which provides better accuracy and can handle large amount of categorical and numerical data. Every classification methods have their own capabilities and limitation (Table 3) but model accuracy can be improved by amalgamation of one classification method with other and also using SVM, Ant colony optimization, Fuzzy techniques etc. In the healthcare sector, researchers are working on diabetes, heart diseases, liver, HIV, Parkinson's, Breast Cancers, Hepatitis, ECG analysis etc. so they can be broadly divided in two basic categories of researches i.e. (a) Domain specific (b) Data Specific. In the ANN classifier the domain data based architecture increase the accurecy of classifier. It is also observed that new techniques like deep learning which are based on filter layer, are very prominent for the large dataset and large attributes. This technique is already applied in the Medical image processing. Bagging and boosting techniques also improves the accuracy level. The best part of these techniques is that it offers enough flexibility for the research.

Scientist have to built more effective classifier because many new diseases, (like Ebola), Dengu are affecting a large population in short time span. Also some diseases are mutating themselves so their timely detection is also a challenge for future research work

Table – 3 : Pros and Cons as reported by different Authors in their research work		
Data mining Technique	**Pros**	**Cons**
(a)KNearest Neighbour	1. Fast training. 2.Simple in use 3.Useful when dataset is in few hundreds	1.Less accurate for missing data or redundant data 2.Normalized dataset required
(b)Decision tree	1.Very Simple to understand, popular 2.No complex calculations 3.No specific healthcare knowledge requires 4.Minimum calculation. 5 Applicable on categorical and numeric data types.	1.Highly structured 2.High Memory requirement 3.High degree of tree length for variety of numeric data 4.Lots of decision points
(c)Support Vector Machine	1.Maximize the classification Margin 2.Minimizes the error 3.More accurate for nonlinear classification	1. It requires $O(N^2)$ training set . 2.Difficult to identify major hyper plane among many 3.Difficult to draw hyper plane for nonlinear data and requires a function
(d)Neural Network	1.Variety of NN models available 2.No complex calculation 3.Flexibility in architecture 4.Bias variable and BPN are helpful to readjust weights	1. High degree of links and neurons 2.High number of iteration 3.Gives optimal result if all iterations are not performed
(e)Bayesian Methods	1.No pre-existing knowledge is required. 2. Provides high accuracy. 3. Depends extensively on probability theory. 4.Useful for determining class of unknown component	1. Prior and dependent attributes are not considered 2.Checks the object similarity/dissimilarity only 3.Not provides much accuracy on the dependent attributes.

In the end we can conclude that for researchers working in health care sector there are immense opportunities and at the same time they may face many challenges. Major challenges are - enhancement of the classifiers to solve its tradeoffs, less computational and memory requirement with the performance parameters like accuracy, specificity, ROC curve, unstructured and heterogeneous health records, health information privacy and legal complication, insufficient and redundant data and unavailability of updated datasets. Data sets are available in the various open and closed data repository which gives opportunity to test proposed model. Apart from the entire challenges healthcare is ever flourishing sector which increases proportionately to the increase in population and so there is always wide room for further research.

REFERENCES

[1] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, "*The KDD process for extracting useful knowledge from volumes of data"*. Communications of the ACM, vol. 39(11), pp. 27-34, 1996.

[2] P. Chandrasekar, K. Qian, H. Shahriar, P. Bhattacharya., "*Improving the Prediction Accuracy of Decision Tree Mining with Data Preprocessing*," IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, pp. 481-484, 2017.

[3] A. Dewan, M. Sharma, "*Prediction of heart disease using a hybrid technique in data mining classification*," 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, pp. 704-706,2015.

[4] N.Lavrac, "*Selected Techniques for data mining for medicine*", Artificial Intelligence ,Medical, vol. 16, pp. 3-23, 1999.

[5] K. Chuchra, A. Chhabra, "*Evalauting the performance of tree based classifiers using Ebola virus dataset*," 1st International

Conference on Next Generation Computing Technologies (NGCT), Dehradun, pp. 494-499, 2015.

[6] Sagar S Nikam. *A Comparative study of Classification Techniques in Data Mining Algorithms*.Oriental Journal of Computer Science & Technology.Vol 8(1),(2015),pp 13-19

[7] G. I. Salama , M. B. Abdelhalim, M. A. Zeid, *"Experimental comparison of classifiers for breast cancer diagnosis," Seventh International Conference on Computer Engineering & Systems (ICCES)*, Cairo, pp.180-185, 2012. doi: 10.1109/ICCES.2012.6408508

[8] Mehmed Kantardzic, "*Data Mining: Concepts,Models, Methods, and Algorithms*, ,John Wiley & Sons, 2003. ISBN: 0471228524

[9] J. Zhu, Q. Xie, K. Zheng, "*An improved early detection method of type-2 diabetes mellitus using multiple classifier system"*. Information Sciences, vol. 292, pp. 1-14. 2015.

[10] M. Fatima , M. Pasha, "*Survey of Machine Learning Algorithms for Disease Diagnostic*". Journal of Intelligent Learning Systems and Applications, vol. 9, 2017.

[11] M. Giardina, F. Azuaje, P. McCullagh , R. Harper, "*A Supervised Learning Approach to Predicting Coronary Heart Disease Complications in Type 2 Diabetes Mellitus Patients*," Sixth IEEE Symposium on BioInformatics and BioEngineering (BIBE'06), Arlington, VA, pp. 325-331. 2006,

[12] A. Bar-Or, D. Keren, A. Schuster,R Wolff., "*Hierarchical decision tree induction in distributed genomic databases*," *in* IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 8, pp. 1138-1151, 2005.

[13] Yeh, Jinn-Yi, "*Applying data mining techniques for cancer classification on gene expression data*",in Taylor & Francis, Cybernetics and Systems , vol. 39 (6), pp. 583-602, 2008.

[14] S. Uhmn, D. H. Kim, S. W Cho,. J. Y Cheong, J. Kim, "*Chronic Hepatitis Classification Using SNP Data and Data Mining Techniques*," Frontiers in the Convergence of Bioscience and Information Technologies, Jeju City, pp. 81-86.2007

[15] V. Vijayan V., Anjali C., "*Decision support systems for predicting diabetes mellitus — A Review*," Global Conference on Communication Technologies *(GCCT)*, Thuckalay, pp. 98-103, 2015.

[16] Vinod Chandra S.S, Anand Hareendran S, "*Artificial intelligence and machine learning*",PHI learning Private Limited ,Delhi, 2014.

[17] S Ali, A. Abraham, "*An empirical comparison of kernel selection for support vector machines",* 2nd International Conference on Hybrid Intelligent Systems, Soft Computing systems: Design, Management and Applications, IOS Press,The Netherlands, pp. 321-330, 2002.

[18] B. K Mishra, P. Lakkadwala, N. K Shrivastava, "*Novel Approach to Predict Cardiovascular Disease Using Incremental SVM*," International Conference on Communication Systems and Network Technologies, *Gwalior*, pp. 55-59, 2013.

[19] Wang Wei and Yang Jiong, "*Mining High-Dimensional Data" Data Mining and knowledge Discovery Handbook*, Springer,pp 793-799, 2005.

[20] E. Gürbüz, E. Kılıç," *A new adaptive support vector machine for diagnosis of diseases"*. Expert Systems, vol. 31(5), pp. 389-397, 2014.

[21] J. Park, D. W. Edington." *A sequential neural network model for diabetes prediction"*. ArtifIntellMed , vol. 23, pp. 277– 93, 2001.

[22] S. N. Yu, K. T. Chou, "*Combining Independent Component Analysis and Back propagation Neural Network for ECG Beat Classification*," International Conference of the IEEE Engineering in Medicine and Biology Society, New York,pp. 3090-3093, 2006.

[23] H. Kahramanli, N. Allahverdi, "*Design of a hybrid system for the diabetes and heart diseases"*. Expert systems with applications, vol. 35(1), pp. 82-89, 2008.

[24] J. Kim, H. S. Shin, K. Shin, M. Lee, "*Robust algorithm for arrhythmia classification in ECG using extreme learning machine*" BioMedical Engineering OnLine, vol-8,issue-https://doi.org/10.1186/1475-925X-8-31.

[25] S. Ding, H. Zhao, Y. Zhang, X. Xu, R. Nie, " *Extreme learning machine: algorithm, theory and applications",* The Artificial Intelligence Review, vol. 44(1), pp. 103-15, 2015.

[26] M. F. Caglar, B. Cetisli, İ. B. TOPRAK, " *Automatic recognition of Parkinson's disease from sustained phonation tests using ANN and adaptive neuro-fuzzy classifier"*, Mühendislik Bilimleri ve Tasarım Dergisi,vol. 1(2), pp. 59-64, 2010.

[27] M. S. Shanker, " *Using neural networks to predict the onset of diabetes mellitus"*. J Chem Inf ComputSci. Vol. 36, pp. 35–41.1996.

[28] M. Lopez, J. Ram´ırez, J. M G´orriz., D. Salas-Gonz´alez., I. A´lvarez, F.Segovia, C.G. Puntonet., "*Automatic tool for Alzheimers disease diagnosis using PCA and Bayesian classification rules*", Electronics Letters,vol. 45, pp.389–391, 2009.

[29] F. Calle-Alonso, C.J. Pérez, J. P. Arias-Nicolás., J Martín, "*Computer-aided diagnosis system: A Bayesian hybrid classification method*", Computer Methods and Programs in Biomedicine, Vol. 112, Issue 1, pp. 104-113, 2013.

[30] Divya Tomar and Sonali.Agrawal, "*A survey on Data Mining approaches for Healthcare*", International Journal of BioScience and Bio-Technology, vol. 5, pp. 241-266, 2013.

[31] T. J. Peter, K. Somasundaram, "*An empirical study on prediction of heart disease using classification data mining techniques*," IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), Nagapattinam, Tamil Nadu, pp. 514-518. 2012.

[32] K. Polat, S. Güneş, " *An improved approach to medical data sets classification: artificial immune recognition system with fuzzy resource allocation mechanism*", Expert Systems, vol. 24, pp. 252–270. 2007.

[33] Yue Huang, Paul McCullagh, Norman Black, Roy Harper, "*Feature selection and classification model construction on type 2 diabetic patients data*", Artificial Intelligence in Medicine, Vol. 41, Issue 3, Pp. 251-262, 2007.

[34] Pasi Luukka, "*Similarity classifier using similarity measure derived from Yu's norms in classification of medical data sets"*, Computers in Biology and Medicine, Vol. 37, Issue 8, pp. 1133-1140, 2007.

[35] S. N. Ghazavi, T. W. Liao, "*Medical data mining by fuzzy modeling with selected features"*, Artificial Intelligence in Medicine, vol. 43(3), pp. 195-206. 2008.

[36] Kemal Polat, Salih Güneş Ahmet, Arslan, " *A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine"*, Expert Systems with Applications, Vol. 34, pp. 482-487, 2008.

[37] Rong-Ho Lin, "*An intelligent model for liver disease diagnosis",* Artificial Intelligence in Medicine, Vol. 47, pp. 53-62, 2009.

[38] I. Gadaras, L.Mikhailov, *An interpretable fuzzy rule-based classification methodology for medical diagnosis*. Artificial intelligence in medicine, vol. 47(1), pp. 25-41, 2009.

[39] B.M Patil, R.C Joshi, Durga Toshniwal, "*Hybrid prediction model for Type-2 diabetic patients",* In Expert Systems with Applications, Vol. 37, pp. 8102-8108, 2010,.

[40] Pasi Luukka, "*Feature selection using fuzzy entropy measures with similarity classifier"*, Expert Systems with Applications, Vol. 38, pp. 4600-4607, 2011.

[41] A. Ozcift, Gulten," *Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms",* Computer methods and programs in biomedicine, vol. 104(3), pp. 443-451, 2011.

[42] Mostafa Fathi Ganji, Mohammad Saniee Abadeh, "*A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis"*, Expert Systems with Applications, Vol. 38, pp. 14650-14659, 2011.

[43] Oscar Bedoya, Irene Tischer, *"Remote homology detection incorporating the context of physicochemical properties*", Computers in Biology and Medicine, Vol. 45, pp. 43-50, 2014.

[44] F. Beloufa, M. A. Chikh, " *Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm*. Computer methods and programs in biomedicine, vol. 112(1), pp. 92-103. 2013.

[45] H. R. Marateb, M. Mansourian, E. Faghihimani, M. Amini, D. Farina, "*A hybrid intelligent system for diagnosing microalbuminuria in type 2 diabetes patients without having to measure urinary albumin*", Computers in biology and medicine, vol. 45, pp. 34-42, 2014.

[46] N. Yilmaz, O. Inan, M. S. Uzer, " *A new data preparation method based on clustering algorithms for diagnosis systems of heart and diabetes diseases*", Journal of medical systems, vol. 38(5), pp. 148–59. 2014.

[47] N. Salari, S. Shohaimi, F. Najafi, M. Nallappan, I. Karishnarajah, " *A Novel Hybrid Classification Model of Genetic Algorithms, Modified k-Nearest Neighbor and Developed Backpropagation Neural Network*", PLoS ONE vol. 9(11), 2014.

[48] A. Purwar, S. K. Singh, *Hybrid prediction model with missing value imputation for medical data*. Expert Systems with Applications, vol. 42(13), pp. 5621-5631, 2015.

[49] M. Seera, C.P. Lim, , S.C. Tan, Chu Kiong Loo, "*A hybrid FAM–CART model and its application to medical data classification", Neural Comput & Applic*, Springer London, 2015.

