



# Research Journal of Pharmaceutical, Biological and Chemical Sciences

## A Review on Classification Algorithms of Medical Diagnostics.

Gokulnath BV and Usha Devi G\*.

School of Information Technology and Engineering, VIT University, Vellore, Tamil Nadu, India.

### ABSTRACT

Data mining that is also an emerging technique used in several applications. The diagnosis of disease through patient history can be done using data mining, clustering and pre-processing technique. For this several classification algorithms, tools and technique are used. It is mainly helpful for the doctors in giving treatment to the patient based on the patient history. These algorithms help in extracting useful information from patient data. This paper provides survey of different classification algorithm used for disease diagnosis. The efficiency of each algorithm and their benefits are analyzed. Heart disease and cancer threatens the life of people for the benefits of the patient and as an easy access for the doctor to treat the patient using classification algorithm.

**Keywords:** classification, KNN, medical diagnostics, AKI helper.

*\*Corresponding author*

## INTRODUCTION

The analysis of data can be done in two forms they are classification and prediction.

The prediction of label of the class is done by classification and for predicting valued function prediction is used. For preparing the data the following process are involved and they are data cleaning, relevance analysis, and data transformation and reduction.

The removal of noise and values missed can be done by data cleaning. Smoothing technique and replacement of values help in data cleaning process. Relevance analysis help in attribute are related or not and the transformation process can be done by normalization and generalization. The problems in prediction can be solved using classification tree. Some machine learning algorithm like SVM [18] are used

Machine learning also can be used for solving this problem and it consist of several applications in it. These algorithm helps in fining the relationship among attributes. The process of analyzing the input helps in prediction process. Genetic programming also helps in solving the classification problem. K-Nearest Neighbors [17] classify based on similarity. It's also uses in statistical estimation and pattern recognition.

And it's also a non-parametric method. These technique of classification helps in the process of large amount of data. It also helps in predicting the disease and helps in the better treatment to the patient. It also assign class label for the data set. Methods of classification are strong in communication modeling.

Shameek Ghosh *et al.*[1] proposed pattern based classification in the initial stage they used contrast mining algorithm and then after post processing during the conversion process frequency based feature specification and then order by based on time sequence pattern and SVM and HMM is also used at the final stage and this system mainly helps for better predictive capability with dynamic changes in ICU and improve patient outcome and the difficulty faced in this process is finding sequence pattern.

Rosa altilioet *al.*[2] determines stereophoto grammatical system with genetic algorithm and the system uses medical data useful information and apply gait analysis this analysis mainly used in controlling the moment of neurological disease and the genetic algorithm uses \fitness function as misclassification of error and the technique used in it is useful tool. The results accuracy using six features is very interesting and it mainly focusing on controlling and analyzing the moments of patient.

Kalaiselviet *al.*[3]diagnose the heart disease using Average K-NN algorithm 80% of death causes due to heart disease and there is no proper classification technique for it and here data set for prediction is obtained from UCI machine learning and the main objective is to reduce the number of attribute and K-NN assign weight to each attribute and find and the benefit of the system is class accuracy is better and attribute reduction and the disadvantage of it is more frequent class dominant prediction.

KaliaOrphanouet *al.*[4] proposed combination of Naive Bayes classifier with temporal association rule with the help of temporal association rule a periodic TAR is found and also time series classification is used. Temporal association rule is used for preprocessing and temporal pattern mining detect relation among TAS and validity of it is compare period TAR with simple TAR. It helps in improving the performance of TAR and long term period is significant in it.

Ranganathaet *al.*[5] analyzed heart disease dataset using classification technique and the useful information stored in it are used for diagnosis and treatment and also for the improvement of tele medicine and community medicine. ID3 and naïve Bayesian classification technique are used. ID3 uses decision tree and naïve Bayesian predict heart disease based on condition. These techniques mainly reduce manual work and mare data can be stored as well as smooth work flow.

Imran Ahmedet *al.*[6]detect breast cancer using decision making and autonomous agent. In case of breast cancer early diagnosis will be useful in the treatment of patient and also for long survival. This hybrid model uses datamining and autonomous agent. Data mining uses decision tree and autonomous agent takes decision based on predefined rules from data mining tools and this model works well with team members.

David G Graham *et al.*[7] proposed a novel attribute selection approach for gastro intestinal cancer. They uses attribute selection process for equal weight on all attribute. It mainly uses Minority Interesting Attribute Threshold for finding important attribute values. In traditional all attributes assigned equally but in MIAT based on entropy in each subset it will consider significant and it is helpful in case of heart problem. MadihaGuftaret *al.*[8]proposed framework for the classification of syncope disease and the classification technique used in it is K-Mean distribution and prediction technology is used for data unavailability and lengthy diagnosis for the cause of syncope disease and the validation part is done through case study. K-Mean test, K- Median, X-Mean are used for the prediction accuracy and the results shown that K-Mean is better than K-Median.

Bourouhouet *al.*[9] compares classification method for detecting Parkinson's. Its chronic neurological degenerate disease and three classifiers are used to it they are KNN, Naive Bayes and support machine. From 20 healthy patient and 20 dis patient features extracted like frequency, amplitude, pitch, pulse for this KNN is used. Naive Bayes identify the performance detection of PD. SVM is used for detection of PD and the final result of comparisons is SVM is effective.

IssariyaUbolthamet *al.*[10]proposed AKI helper tool for acute kidney injury. The tool is built based on the guideline of KDIGO. The diagnosis and decision making process is done using simple cart and J48. Three approaches are used in it they are AKI risk which involves J48 and Simple cart and idea to develop AKI helper and for evaluation 2\*2 contingency table. For identifying risk factor decision tree classification are used.

VladanPapicet *al.*[11] proposed Random Forest Algorithm for cardiac arrhythmia detection based on compressive sensing. For the classification purpose random forest algorithm and for the measurement performance ECG compares with MIT BIH arrhythmia database and random forest will randomize decision tree and second randomization is done at node level, each node randomize set of variable. The result shows that 99.3% of accuracy is obtained.

Aravinthet *al.*[12] analyzed detection of disease on brinjalleaves using K-Mean clustering method. Image processing and artificial neural technology is used for finding the cause and loss of production. K-Mean method uses segmentation and neural network classification. Cluster is done by segmentation process and also for the new method for identifying in plant using area, perimeter, centroid, diameter and mean are used.

Muhammad Saqlainet *al.*[13] identified heart failure with the help of unstructured data of cardiac patients. Early detection of heart disease prevents us from expensive treatment. DM techniques provide useful information from medical data set. Machine learning classification algorithmused for extraction of useful information and for multiclass classification and process description Naive Bayesian is used.

Harpreet Kauret *al.*[14]determineliver medical condition Steatosis. And the classification method is implemented that can sort ultrasonic image into normal. In this combination of seven different texture analysis model for extracting features and two fold feature selection method is also used for removing redundancy and reduce in to five features which is used in Naive Bayes classifier and provide accuracy of 92% and sensitivity of 100%. In this individual texture method makes them more efficient.

Franklin vinodet *al.*[15] classification of the data from the big data is identified as a great challenge and it is the process of categorizing data for efficient use and for the analysis of the large scale patient record and the hierarchical approach which is tree structured and train maximum classifier and gives better classification result and also computation efficient mainly Highly Co-related Feature Set Selection for combining hierarchical and to improve performance. It identifies the good quality subset and provides quality features

WarintornPhusosomsaiet *al.* [16]determinethe recognition rate of brain tumors cell image. In this image processing is the first step used as pre-processing technique then histogram orientation gradient is used for future extraction and then Extreme Learning Machine is used for cell classification. It mainly speeds up 3 and 7 times than normal non-parallel scheme and ELM. Along with ELM a comprehensive evaluation against existing and enhancement with parallelism is PH-ELM. Speed of classification is high.

**CONCLUSION**

The classification algorithm helps in early diagnosis of disease which help in the better treatment and early diagnosis for patient. Each technique has its own advantage and disadvantage. The correctness ratio plays a vital role in finding the goodness of algorithm and it is set as the performance metric. The time taken for classifying the data set varies for different technique used. This survey describes different methodology and character used in it. It helps in improvement in medical field of diagnosis and treatment. This opens a wide area of research.

S. NO	METHOD	FEATURES
1	Pattern based classification Contrast data mining SVM,HMM	Predict better with dynamic change in ICU
2	Stereo photo gram etic Genetic algorithm	Focus on controlling and analysing moment on patient
3	Average KNN algorithm	Accuracy of class is good and attribute reduction
4	Naïve Bayes Temporal association rule	Performance of TAR is improved
5	ID3 Naïve Bayes	Reduce manual work Workflow is smoother
6	Decision making Autonomous system	Along with team members perform well
7	Novel attribute selection	It's also useful for heart problem
8	K-mean distribution K-median, X-mean	K-mean betters K-median
9	KNN Naïve Bayes Support Machine	Performance detection of PD SVM is effective
10	AKI Helper Simple cart J48	Helps in kidney injury diagnosis Identify risk factor quickly
11	Random Forest Algorithm	99.3% accuracy
12	K-mean distribution Image processing Artificial neural technology	New method for identifying in plants
13	DM Technique Machine Learning Algorithm	Classification is faster
14	Seven different texture analysis Model	92%accuracy 100%sensitivity
15	Highly correlated feature set selection	Identifying good quality and better features
16	Image processing Pre-processing technique	Speed of classification is high

**REFERENCES**

- [1] Shameek Ghosh and Hung Nguyen, "Predicting short-term ICU outcomes using a sequential contrast motif based classification framework", IEEE pp.5612-5615, 2016.
- [2] Rosa Altilio and Luca Liparulo, "A Genetic Algorithm for Feature Selection in Gait Analysis", Department of Information Engineering, Electronic and Telecommunication, IEEE Congress on Evolutionary Computation (CEC), 2016.
- [3] C. Kalaiselvi "Diagnosis of Heart Disease using Average K-Nearest Neighbors Algorithm of Data Mining", International Conference on Computing for Sustainable Global Development (INDIACom),2016.
- [4] KaliaOrphanou and Arianna Dagliati "Combining Naïve Bayes Classifier with Temporal Association Rules for Coronary Heart Disease Diagnosis", IEEE International Conference on Healthcare Informatics, 2016.

- [5] S.Ranganatha and Pooja Raj "Medical Data Mining and Analysis for Heart Disease Dataset Using Classification Techniques".
- [6] AL-Imran Ahmed "A Hybrid Approach For Decision Making To Detect Breast Cancer Using Data Mining and Autonomous Agent Based On Human Agent Teamwork",17<sup>th</sup> International Conference On Computer And Information Technology (ICIT),2014.
- [7] Avi Rosenfeld and David G. Graham "A Novel Attribute Selection Approach to Better Predict Upper Gastrointestinal Cancer" IEEE International Conference, 2015.
- [8] MadihaGuftar "A Novel Framework for Classification of Syncope Disease Using K-Mean Clustering Algorithm" SA1 Intelligent System Conference, 2015.
- [9] A.Bourouhou "Comparison of Classification Method to Detect the Parkinson Disease", 2nd International Conference on Electrical and Information Technologies ICETT, 2016.
- [10] IssariyaUboltham "AKIHelper: Acute Kidney Injury Diagnostic Tool Using KDIGO Guideline Approach" ICIS, 2016.
- [11] Tea Marasovic and VladanPapic "Cardiac Arrhythmia Detection Using DCT Based Compressive Sensing and Random Forest Algorithm".
- [12] Veni and Aravinth "An Application of Image Processing Techniques for Detection of Disease on Brinjal Leaves Using K-Means Clustering Method" Fifth International Conference on Recent Trends in Information Technology, 2016.
- [13] Muhammad Saqlain and Wahid Hussain "Identification of Heart Failure by Using Unstructured Data of Cardiac Patient" 45<sup>th</sup> International Conference on Parallel Processing Workshops, 2016.
- [14] Harpreet Kaur, Ajay Shiv Sharma, Sukhdeep Kaur "Liver Tissue Classification For Ultrasound Image", March 2015.
- [15] D. Franklin Vinod, V. Vasudevan, "Filter Based Feature Set Selection Approach for Big Data Classification of Patient" February 2016.
- [16] WarintornPhusomsai and WiyadaPunjaruk "Brain Tumor Cell Recognition Schemes Using Image Processing With Parallel ELM Classification on GPU" Communications (ICC), 2016.
- [17] D. Gomez and F. Prieto "Nearest Neighbors by Adaptive Simulated Annealing" IEEE Latin America Transactions, Vol. 13, pp.2398-2404, July 2015.
- [18] Ao Kong and Chinmaya Gupta "Biomarker Signature Discovery from Mass Spectrometry Data" IEEE/ACM Transaction on Computational Biology and Bioinformatics, Vol 11, pp.766-772, July 2014.