International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 9 Issue 1 January 2022

# Heart Diseases Prediction using Classification and Optimization Techniques

Dr. Pankaj Kawadkar[1]
Head & Associate Professor, Department of CSE, School of Engineering, SSSUTMS, Sehore (M.P.), India[1]
kawadkarpankaj@gmail.com[1]

***Abstract:*** *In this paper presents the classification algorithm such as decision tree, support vector machine and proposed work with optimization techniques like particle swarm optimization. Particle swarm optimization methods for the pattern extraction and pattern recognition for the huge amount of dataset, dataset used from the UCI machine learning repository such as heart diseases dataset etc. these dataset are basically belong to medical science dataset.*

***Keywords:*** *Supervised Learning, Decision, Support Vector Machines, Health Care, Medical Science*.

## 1. INTRODUCTION

Clustering is a division of data into groups of similar objects, it's basically examples of unsupervised learning algorithm in which there is no supervisor and no target output for the respective techniques. Its generate and represent the data in the form of group or cluster, here the generation of cluster or grouping of data is done by the various clustering techniques with using pattern recognition and pattern extraction form the large or huge amount of dataset.

Classification is a data mining technique that typically involves three phases, a learning phase, a testing phase and an application phase. A learning model or classifier is built during the learning phase. It may be in the form of classification rules, a decision tree, or a mathematical formula. Since the class label of each training sample is provided, this approach is known as supervised learning. In unsupervised learning (clustering), the class labels are not known in advance. In the testing phase test data are used to assess the accuracy of classifier. If the classifier passes the test phase, it is used for the classification of new, unclassified data tuples. This is the application phase.

A new decision tree learning approach for novel class detection. In this builds a decision tree from data stream which continuously update. Calculate threshold value based on ratio of percentage of data points between each leaf node in a tree and the training dataset and cluster the data points of training data set based on similarity of attributes. If number of data points classify at a leaf node increases than the threshold value increases then novel class arrived.

## 2. CLASSIFICATION

Classification is a data mining technique that typically involves three phases, a learning phase, a testing phase and an application phase. A learning model or classifier is built during the learning phase. It may be in the form of classification rules, a decision tree, or a mathematical formula. Since the class label of each training sample is provided, this approach is known as supervised learning. In unsupervised learning (clustering), the class labels are not known in advance. In the testing phase test data are used to assess the accuracy of classifier. If the classifier passes the test phase, it is used for the classification of new, unclassified data tuples. This is the application phase. The classifier predicts the class label for these new data samples. For classification algorithms, the two major problems on classifying a data stream are the infinite length and the concept drift. The first one makes the traditional multi-pass classification algorithms incapable of classifying a data stream for their requirement of infinite storage and large amount of training time. The second one makes the most static stream classification algorithms incapable of classifying a data stream with concept drifts for the under-lying changes occurred in the stream. For a time changing data stream, an incremental updating manner of the classifier is very important. A temporal model is used to capture the

evolutions of the stream. In general, the classification process is always accompanied by the course of model construction and test. The classification model keeps changing with the progression of the stream. If a static classifier is used to classify an evolving data stream, the accuracy of it will drop greatly.

## 3. RESEARCH GAP

The mining of healthcare database is very critical issue. The healthcare data strode 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. For the extraction of better information used data mining technique for the healthcare system. The diversity and applicability of data mining are increase day to day in the field of medical science for the predication of symptom of disease. The data mining provide lots of technique for mine data in several field, the technique of mining as association rule mining, clustering technique, classification technique and emerging technique such as called ensemble classification technique. The process of ensemble classifier increases the classification rate and improved the majority voting of classification technique for individual classification algorithm such as KNN, Decision tree and support vector machine. The new paradigms of ensemble classifier are cluster oriented ensemble technique for classification of data. Classification of medical data is an important task in the prediction of any disease. It even helps doctors in their diagnosis decisions. Cluster oriented Ensemble classifier is to generate a set of classifiers instead of one classifier for the classification of a new object, hoping that the combination of answers of multiple classification results in better performance. We demonstrate the algorithmic use of the classification technique by extending SVM the most popular binary classification algorithms. From the studies above, the key to improve cluster oriented classifier is to improve binary classification. In the final part of the thesis, we include empirical evaluation that aim at understanding binary classification better in the context of ensemble learning.

## 4. PROPOSED WORK

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 Decision tree, Support vector machines and Proposed method i.e. combination of support vector machines and optimization methods. 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.

Feature optimization is an important area of health care domain. The extraction process gives the better amount of feature for the feature for the processing of feature. But the signal image generate huge amount of feature for the processing of optimization. In this section used feature optimization technique. The feature optimization technique adopted the particle of swarm optimization.

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

In the Step 1 All the particles or object velocity and position are randomly place to within pre-defined ranges.

In the step 2 the updation in velocity of particle at every iteration round, the velocities of all particles are updated based on some expression i.e.

$$v_i = v_i + c_1 R_1 (p_{i,best} - p_i) + c_2 R_2 (g_{i,best} - p_i)$$

In the case of equation 6.4.1 the $p_i$ is the position and $v_i$ are the velocity of particle $i$, $p_{i,best}$ and $g_{i,best}$ is the position with the 'best' objective value found so far by particle $i$ and the entire population respectively;

In the next step the updation in position of particle, and all the particles are updated according to,

$$p_i = p_i + v_i$$

Following updating, $p_i$ should be verified and limited to the allowed range.

In the next step the memory updation update $p_{i,best}$ and $g_{i,best}$ when condition is met,

$$p_{i,best} = p_i \quad if\ f(p_i) > f(p_{i,best})$$
$$g_{i,best} = g_i \quad if\ f(g_i) > f(g_{i,best}) \quad \ldots$$

Where *f(x)* is to be optimized and it is a objective function.

Finally in the last step we update the terminate condition will repeat the algorithm steps 2 to 4 until certain terminating circumstances are met.

**PSO Pseudo Code**

begin
t=0;
initialize particles p (t);
evaluate particles p (t);
while (termination conditions are unsatisfied)

begin
t=t+1;
update weights
select pbest for each particle
select gbest from p (t-1);
calculate particle velocity p(t)
calculate particle position p(t)
evaluate particles p(t)
end
end

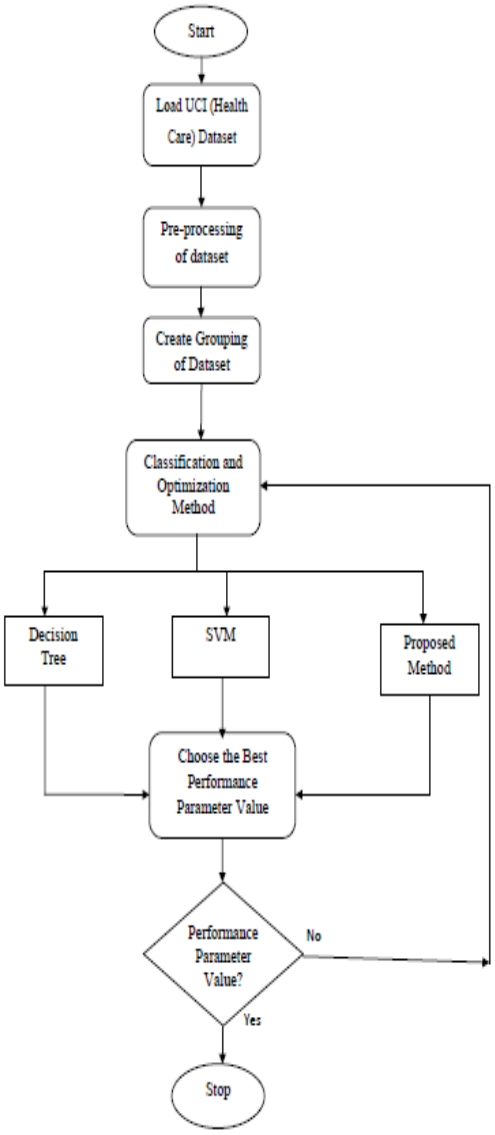


**Figure 1:** Proposed model for the health care disease prediction classification system.

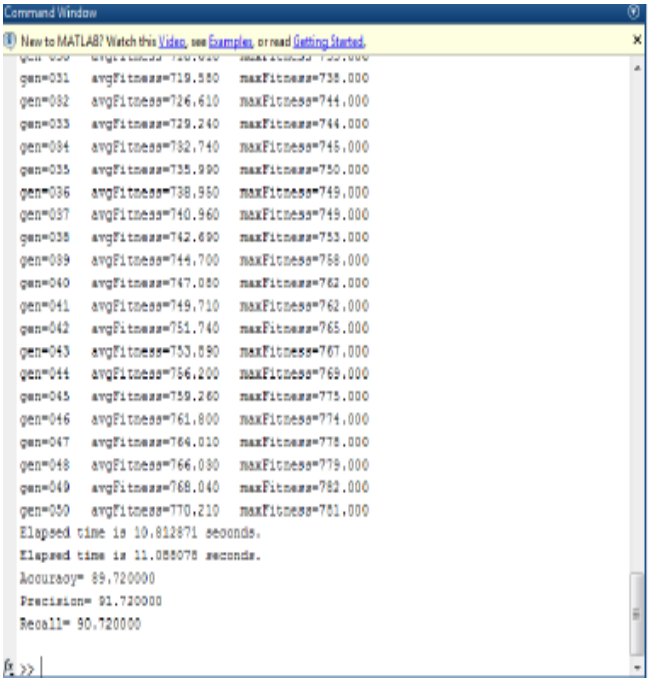


**Figure 2:** This windows show that the result of Proposed methods with accuracy in the experimental process using Heart dataset.

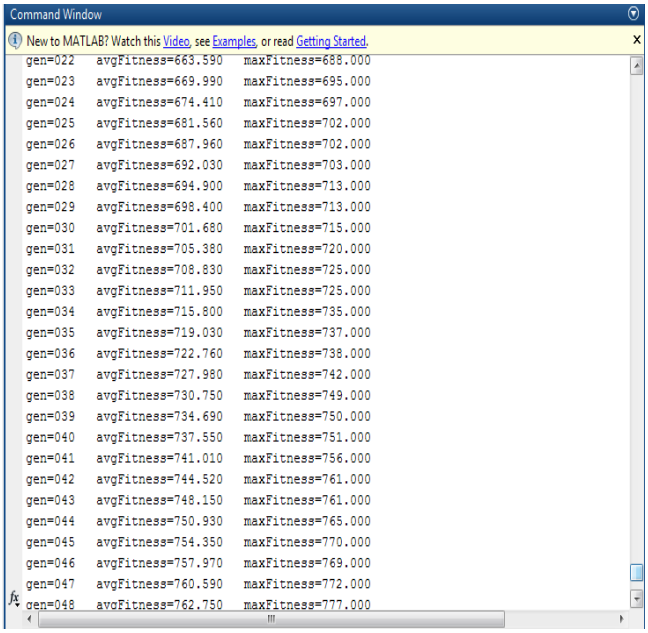


**Figure 3:** This windows show that the result of Decision tree methods with accuracy in the experimental process using Cancer dataset.

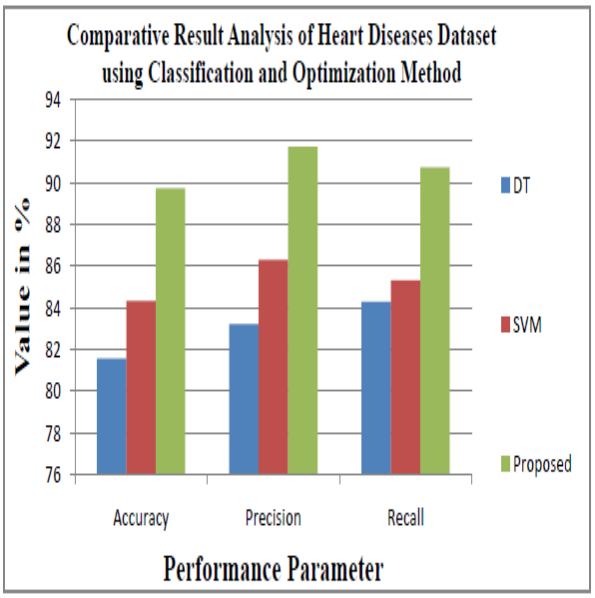


**Figure 3:** Show that the comparative result analysis for the Heart disease diagnosis using various classification and optimization techniques, our empirical result study shows that better accuracy than existing methods.

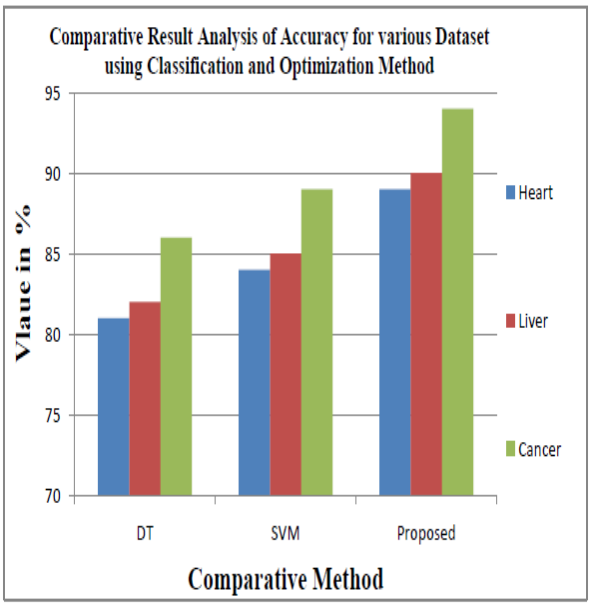


**Figure 4:** Show that the comparative result analysis for the Heart disease, Liver disease and cancer disease dataset diagnosis using SVM and our proposed method, our empirical result study shows that better accuracy than existing methods.

## 5. CONCLUSION

In this paper present a medical science field dataset such as a heart diseases dataset, liver diseases dataset and cancer diseases datasets. The heart rate diseases diagnosis prediction and classification model using the classification method such as decision tree, support vector machine and optimization method such as particle swarm optimization, in which we focus on pattern extraction and pattern analysis of healthcare data environment features which is taken from the heart diseases patients and form various hospitals whose stores the any patients past data or history. The support vector machine is a set of classifier which provide the best classification results in the form of various class we also apply the same dataset using the particle swarm optimization method and getting the results. The particle of swarm optimization gives us better results than classification method.

## REFERENCES

[1] R.Sujatha, E.P.Ephzibah, Sree Dharinya, G. Uma Maheswari, V.Mareeswari Vamsi Pamidimarri " Comparative Study On Dimensionality Reduction For Disease Diagnosis Using Fuzzy Classifier" International Journal Of Engineering & Technology, 2018, Pp 79-84.

[2] Seyed Mohammad Jafar Jalali, Sergio Moro, Mohammad Reza Mahmoudi, Keramat Allah Ghaffary "A Comparative Analysis Of Classifiers In Cancer Prediction Using Multiple Data Mining Techniques" ISCTE-IUL, 2017. Pp 1-12.

[3] S.Leoni Sharmila, C.Dharuman, P.Venkatesan "Disease Classification Using Machine Learning Algorithms - A Comparative Study" International Journal Of Pure And Applied Mathematics, Vol-114, 2017, Pp 1-10.

[4] Meherwar Fatima, Maruf Pasha "Survey Of Machine Learning Algorithms For Disease Diagnostic" Journal Of Intelligent Learning Systems And Applications, 2017, Pp 1-16.

[5] Tabreer T, Hasan, Manal H, Jasim And Ivan A. Hashim " Heart Disease Diagnosis System Based On Multi-Layer Perceptron Neural Network And Support Vector Machine" International Journal Of Current Engineering And Technology, Vol-7, 2017.

[6] 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.

[7] Dong Li, Cheng Tang, Chunlei Xia, Hua Zhang, "Acoustic Mapping And Classification Of Benthic Habitat Using Unsupervised Learning In Artificial Reef Water", Estuarine, Coastal And Shelf Science, 2017. Pp 11-21.

[8] 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.

[9] Arijit Ukil, Soma Bandyopadhyay, Chetanya Puri And Arpan Pal "Heart-Trend: An Affordable Heart Condition Monitoring System Exploiting Morphological Pattern", IEEE, 2016, Pp 6260-6264.

[10] Theresa Princy. R And J. Thomas "Human Heart Disease Prediction System Using Data Mining Techniques", ICCPCT, 2016, Pp 1-5.

[11] Aigerim Altayeva, Suleimenov Zharas And Young Im Cho "Medical Decision Making Diagnosis System Integrating K-Means And Naïve Bayes Algorithms", ICCAS, 2016, Pp 1087-1092.

[12] Mohammed G. Ahamad, Mohammed F. Ahmed And Mohammed Y. Uddin, "Clustering As Data Mining Technique In Risk Factors Analysis Of Diabetes, Hypertension And Obesity", European Journal Of Engineering Research And Science, 2016. Pp 88-94.

[13] Zhenyundeng, Xiaoshuzhu , Debocheng, Mingzong, Shichaozhang, "Efficient Knn Classification Algorithm For Big Data", Elsevier Ltd. 2016, Pp 143-148.

[14] Anna L. Buczak, Erhan Guven, "A Survey Of Data Mining And Machine Learning Methods For Cyber Security Intrusion Detection", IEEE Communications Surveys & Tutorials, 2016. Pp 1153-1176.

[15] Jean Damascene Mazimpaka, Sabine Timpf, "Trajectory Data Mining: A Review Of Methods And Applications", JOURNAL OF SPATIAL INFORMATION SCIENCE, 2016. Pp 61-99.

[16] S M.Inzalkar, Jai Sharma, "A Survey On Text Mining- Techniques And Application", International Journal Of Research In Science & Engineering, 2016. Pp 1-9.

[17] T. Sajana, C. M. Sheela Rani And K. V. Narayana, " A Survey On Clustering Techniques For Big Data Mining", Indian Journal Of Science And Technology, 2016. Pp 1-12.

[18] Ruogu Fang, Samira Pouyanfar∗, Yimin Yang, Shu-Ching Chen, S. S. Iyengar, "Computational Health Informatics In The Big Data Age: A Survey", ACM Computing Surveys, Vol. 49, 2016. Pp 1-36.

[19] Sudha Ram, Wenli Zhang, Max Williams, Yolande Pengetnze, "Predicting Asthma-Related Emergency Department Visits Using Big Data", IEEE Journal Of Biomedical And Health Informatics, VOL. 19, 2015. Pp 1216-1218.

[20] Marco Viceconti, Peter Hunter, Rod Hose, "Big Data, Big Knowledge: Big Data For Personalized Healthcare", IEEE Journal Of Biomedical And Health Informatics, VOL. 19,, 2015. Pp 1209-1215.

[21] Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, And Guang-Zhong Yang," Big Data For Health", IEEE Journal Of Biomedical And Health Informatics, VOL. 19, 2015. Pp 1193-1206.

[22] Yin Zhang, Meikang Qiu, Chun-Wei Tsai, Mohammad Mehedi Hassan, Atif Alamri, "Health-CPS: Healthcare Cyber-Physical System Assisted By Cloud And Big Data", IEEE Systems Journal, 2015. Pp 1-9.

[23] Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, Kayvan Najarian, "Big Data Analytics In Healthcare", Hindawi Publishing Corporation Bio-Medical Research International, 2015. Pp 1-17.

[24] Zoubin Ghahramani, "Probabilistic Machine Learning And Arti_Cial Intelligence", 2015. Pp 1-24.

[25] Ji-Jiang Yang, Jianqiang Li, Jacob Mulder, Yongcai Wang, Shi Chen, Hong Wu, Qing Wang, Hui Pan, "Emerging Information Technologies For Enhanced Healthcare', Elsevier Ltd. 2015. Pp 3-11.