# SMO AND LAZY CLASSIFIERS FOR HEART DISEASE PREDICTION

Aung Nway Oo, Thin Naing

*University of Information Technology, Myanmar*

## ABSTRACT

*Data mining and machine learning technique are rapidly developed for many kinds of applications. One of the important applications is in medical diagnosis field. The aim of this paper is to compare the different classifiers for heart disease classification. In this paper, SMO classifier and two lazy classifiers (IBK and KStar) are used for prediction of heart disease. The results of each classifiers are compared and analysis. The experiments results proved that SMO classifier is better than other classifiers for heart disease prediction.*

**Keywords:** *Data mining, Machine Learning, Classification*

## 1. INTRODUCTION

Data mining is used to extract hidden and previously unknown patterns from large datasets. The relationships and knowledge of large data elements and features are difficult to detect with traditional statistical methods [1]. In recent years, several machine learning algorithms are used to perform data mining. Data mining in healthcare is an emerging field of high importance for providing prognosis and a deeper understanding of medical data [2]. Data mining approach can be used healthcare industry research problems. Machine learning algorithms are used in order to undertake the research problem of classification in medical dataset.

According to the WHO report, heart diseases take the lives of 17.9 million people every year, 31% of all global deaths. Heart diseases are disorders of the heart and blood vessels and include coronary heart disease, cerebrovascular disease, rheumatic heart disease and other conditions. Therefore, diagnosis of heart disease is important research of medical field. Medical data mining is the new area for exploring hidden data pattern from huge amount of data.

Classification is the major research topic in data mining. Typically classification represents the data to be categorized based on its features or characteristics. Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of a predefined set of classes or groups.

In machine learning, lazy learning is a learning method in which generalization of the training data is, in theory, delayed until a query is made to the system. Lazy classifiers are most useful for large, continuously changing datasets with few attributes that are commonly queried.

The rest of the paper is organized as follows. Section 2 provides the related work and section 3 presents the overview of classification algortihms. General description of heart disease dataset is described in section 4. The experimental results are discussed in section 5. Finally, conclusion of this study was provided in section 6.

## 2. RELATED WORKS

There are many research papers that represent the prediction of heart disease. Atul Kumar Pandey [3] proposed the decision model based heart disease prediction with j48 decision tree classifier. Paper [4] presented the optimization approach using Genetic algorithm and compared the performance with different classifiers. Dynamic Bayesian

network [5] with temporal abstraction was presented for prognosis of coronary heart disease. evaluate different classification techniques in heart disease diagnosis. Boshra Bahrami and Mirsaeid Hosseini Shirvani [6] compare the classification results of heart disease using J48 Decision Tree, K Nearest Neighbors(KNN), Naive Bayes(NB), and SMO classifier. Comparative study of the various performance of machine learning algorithms was done by Sanjay Kumar Sen [7]. Ms. D. Monica Seles, etal. [8] proposed the Survey on Predicting the Heart Disease Using Data Mining Methods. The performance evaluation of eager (naïve Bayes, ADTree) and lazy (IBk, KStar)classification algorithms are experimented by A. Hency Juliet and R. Padmajavalli [9]. In the work of Sarangam Kodati & Dr. R. Vivekanandam [10], heart disease was predictive using weka tool and orange data mining tool.

In this study, we proposed the comparative study of heart disease classification using SMO, IBK and KStar algorithms.

## 3. CLASSIFICATION ALGORITHMS

The SMO algorithm was proposed by John C. Platt [11] in 1998 and became the fastest quadratic programming optimization algorithm. Sequential Minimal Optimization (SMO) is used for training a support vector classifier using polynomial or RBF kernels. It replaces all missing the values and transforms nominal attributes into binary ones. A single hidden layer neural network uses exactly the same form of model as an SVM.

IBk classifier is a simple instance-based learner that uses the class of the nearest k training instances for the class of the test instances. The number of nearest neighbors can be specified explicitly in the object editor or determined automatically using leave-one-out cross-validation focus to an upper limit given by the specified value [12].

KStar is an instance-based classifier, that is the class of a test instance is based upon the class of those training instances similar to it, as determined by some similarity function. The KStar algorithm can be defined as a method of cluster analysis which mainly aims at the partition of 'n' observation into 'k' clusters in which each observation belongs to the cluster with the nearest mean. We can describe KStar algorithm as an instance based learner which uses entropy as a distance measure. The benefits are that provides a consistent approach to handling of real valued attributes, symbolic attributes and missing values [13].

## 4. DATASET DESCRIPTION

Heart disease dataset contains 13 features and 270 instances. There are two classes: the presence and absence of heart disease. There are 150 patient records without suffer heart disease and 120 records for patient with heart disease. Table-1 shows the features and description of heart disease dataset. The dataset is obtained from UCI machine repository.

Table-1. Features and description

| # | **Features** | **Description** |
|---|---|---|
| 1. | Age | Instance age in years |
| 2. | Sex | Instance gender |
| 3. | Cp | Chest pain type |
| 4. | Trestbps (mmHg) | Resting blood pressure |
| 5. | Chol (mg/dl) | Serum cholesterol |
| 6. | Fbs | Fasting blood sugar |
| 7. | Restecg | Resting electrocardiographic results |
| 8. | Thalach | Maximum heart rate achieved |
| 9. | Exang | Exercise induced angina |

| | | |
|---|---|---|
| 10. | Oldpeak | ST depression induced by exercise relative to rest |
| 11. | Slope | The slope of the peak exercise ST segment |
| 12. | Ca | Number of major vessels (0-3) colored by flourosopy |
| 13. | Thal | 3 = normal; 6 = fixed defect; 7 = reversible defect |

## 5. EXPERIMENTAL RESULTS

Weka data mining tool was used for prediction of heart disease. 66 % of dataset is used for training and remaining 34 % is used for testing. The results of classifiers are showed in the following table.

Table-2. Experimental results of classifiers

| | **SMO** | **IBK** | **KStar** |
|---|---|---|---|
| Correctly Classified Instances | 76 | 70 | 67 |
| Incorrectly Classified Instances | 16 | 22 | 25 |
| Kappa statistic | 0.65 | 0.53 | 0.47 |
| Accuracy | 82.6% | 76.1% | 72.8% |

The visualization results of classifiers are describes in chart-1 and chart-2.

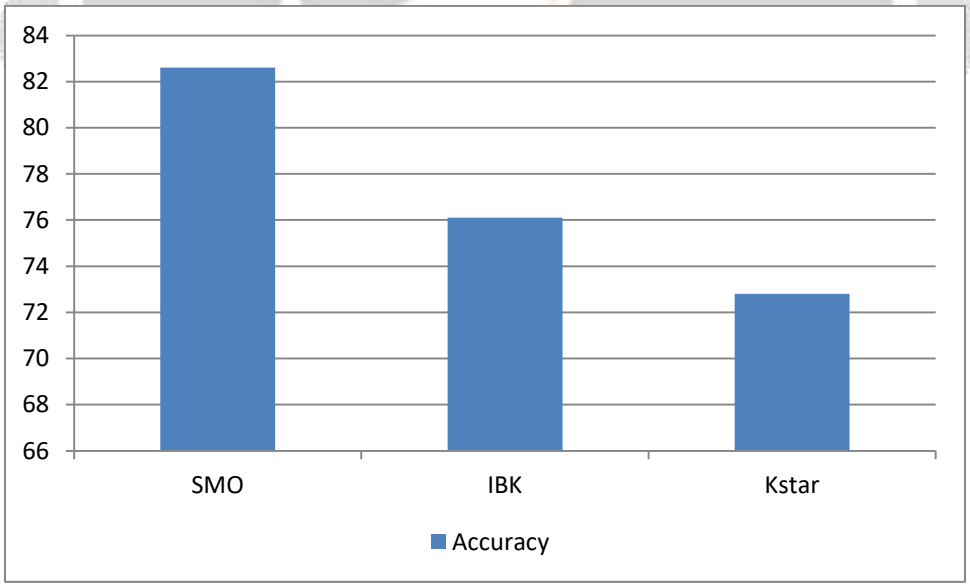


Chart-1 Comparison of accuracy for each classifier

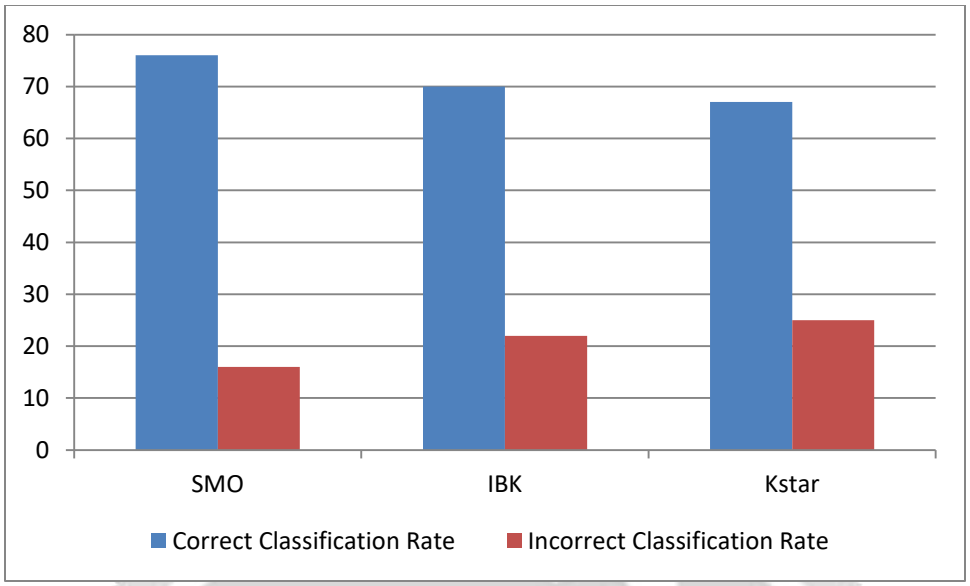


Chart-2 Classification results of classifiers

## 6. CONCLUSION

Classification is important for medical diagnosis. There are several algorithms for classification of data set. In this paper, we carried out the experiment to find the predictive results of SMO and Lazy classifiers. The accuracy results of SMO classifier is 82.6 which is better than other classifiers.

## 7. REFERENCES

[1] Obenshain, M.K., Application of Data Mining Techniques to Healthcare Data. Infection Control and Hospital Epidemiology,2004.

[2] Liao, S.-C. and I.-N. Lee, Appropriate medical data categorization for data mining classification techniques. MED. INFORM.Vol. 27, no. 1, 59–67, 2002.

[3] Atul Kumar Pandey, Prabhat Pandey, K.L. Jaiswal, Ashish Kumar Sen. A Heart Disease Prediction Model using Decision Tree, IOSR Journal of Computer Engineering (IOSR-JCE), Volume 12, Issue 6 (Jul. - Aug. 2013), PP 83-86

[4] P. Suresh, etal Study and Analysis of Prediction Model for Heart Disease: An Optimization Approach using Genetic Algorithm, International Journal of Pure and Applied Mathematics, Volume 119 No. 16 2018, PP 5323-5336

[5] KaliaOrphanou, Athena Stassopoulou, and ElpidaKeravnou,"DBN-Extended: A Dynamic Bayesian Network Model Extended With Temporal Abstractions for Coronary Heart Disease Prognosis", IEEE Journal of Biomedical and Health Informatic, VOL. 20, NO. 3, MAY 2016

[6] Boshra Bahrami, Mirsaeid Hosseini Shirvani: Prediction and Diagnosis of Heart Disease by Data Mining Techniques, Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159-0040 Vol. 2 Issue 2, February – 2015

[7] Sanjay Kumar Sen: Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms, International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 6 June 2017, Page No. 21623-21631

[8] Ms. D. Monica Seles, etal.: A Survey on Predicting the Heart Disease Using Data Mining Methods, International Journal of Science, Engineering and Management (IJSEM) Vol 2, Issue 12, December 2017 , Page No. 47-50

[9] A. Hency Juliet* and R. Padmajavalli: A Comparative Study on Performance Evalutionof Eager versus Lazy Learning Methods, Journal of Computer Science and Information Technology, ol. 3, Issue.3, March 2014, Page No.562–568

[10] Sarangam Kodati & Dr. R. Vivekanandam: Analysis of Heart Disease using in Data Mining Tools Orange and Weka, Global Journal of Computer Science and Technology: CVolume 18 Issue 1 Version 1.0 Year2018, Page No.562–568

[11] Platt, John (1998), Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines

[12] Ms S. Vijayarani etal.: Comparative Analysis of Bayes and Lazy Classification Algorithms, nternational Journal of Advanced Research in Computer and Communication EngineeringVol. 2, Issue 8,August 2013,

[13] Trilok Chand Sharma, Manoj Jain, "WEKA Approach for Comparative Study of Classification Algorithm"