

# Heart Disease Diagnosis Using Machine Learning Algorithm

Shashikant U. Ghumbre<sup>1</sup> and Ashok A. Ghatol<sup>2</sup>

<sup>1</sup> Computer Engineering Department,  
College of Engineering Pune, Pune, Maharashtra, India  
shashi.ghumbre@gmail.com

<sup>2</sup> Dr. B.A.T. University, Lonere, Maharashtra, India  
vc\_2005@rediffmail.com

**Abstract.** Recent advances in computing and developments in technology have facilitated the routine collection and storage of medical data that can be used to support medical decisions. However, in most countries, there is a first need for collecting and organizing patient's data in digitized form. Then, the collected data are to be analyzed in order for a medical decision to be drawn, whether this involves diagnosis, prediction, course of treatment, or signal and image analysis. In this paper, India centric dataset is used for Heart disease diagnosis. The correct diagnosis performance of the automatic diagnosis system is estimated by using classification accuracy, sensitivity and specificity analysis. The study shows that, the SVM with Sequential Minimization Optimization learning algorithm have better choice for medical disease diagnosis application.

**Keywords:** Decision Support System, Support Vector Machine, Heart Disease, Machine Learning.

## 1 Introduction

The application of machine learning methods in medical field is the subject of considerable ongoing research, which mainly concentrates on modeling some of the human actions or thinking processes and recognizing diseases from a variety of input sources. Other application areas are knowledge discovery [10] and biomedical systems, which include genetics and DNA analysis [1, 12]. The design may also be influenced by the desired performance on one or more specific classes of the problem instead of the overall performance. This is usual in most medical tasks as a different degree of significance may be required for the system's performance on each class. The computer programs or machine learning techniques can be used to reduce the mortality rate, improve the accuracy in disease diagnosis and mainly reduce the diagnosis time. The advancement in computer technology and communication encourages health-care providers to work using the Internet or Telemedicine technology [9,13,14].

In a medical diagnosis problem, what is needed is a set of examples or attributes that are representative of all the variations of the disease. The examples need to be selected very carefully if the system is to perform reliably and efficiently. The fact that there is no need to provide a specific algorithm on how to identify the disease, presents a major advantage over the application of machine learning methods to this

type of problems. However, development of artificial intelligence systems for medical decision making problems is not a trivial task. Difficulties include the acquisition, collection and organization of the data that will be used for training the system. This becomes a major problem especially when the system requires large data sets over long periods of time, which in most cases are not available due to the lack of an efficient recording system. The above mentioned problems or the existing procedures involved in the medical task may not be the only factors affecting the design of a Decision Support System (DSS). The design may also be influenced by the desired performance on one or more specific classes of the problem instead of the overall performance. This is usual in most medical tasks as a different degree of significance may be required for the system's performance on each class. For example, in a heart disease diagnosis task, it is necessary for the accuracy on healthy patients to be as high as possible, as a misclassification in this category may result in a healthy patient going under treatment for no reason. The balance of the system's performance between different classes could vary and is largely dependent on the medical problem itself and the collected data. In addition, in most of the countries, insufficient numbers of medical specialist have increased the mortality of patients suffering from various diseases. Heart diseases have emerged as the number one killer in both urban and rural areas in most of the countries. As of 2010, it is the leading cause of death in the U.S., England and Canada, accounting for 25.4% of the total deaths in the United States. Similar situation is found rest of the countries all over the world. In case of heart disease time is very crucial to get correct diagnosis in early stage[37]. It is observed that, in many cases due to wrong diagnosis or trial/error procedure for diagnosis leads to patient health compromise. The dearth of medical specialists and/or wrong diagnosis procedure will never be overcome within a short period of time [3,4]. Patient having chest pain complaint may undergo unnecessary treatment or admitted in the hospital. In most of the developing countries specialists are not widely available for the diagnosis. Hence, such automated system can help to medical community to assist doctor for the accurate diagnosis well in advance.

The rest of the paper is organized as follows: Section 2 briefly reviews some prior works on machine learning techniques in medical Diagnosis. Section 3 briefly describes the heart disease diagnosis and the proposed Decision Support System (DSS) and its techniques used are discussed. Section 4 details the use of Support Vector Machine in medicine. The experimental results are given in Section 5. Section 6 concludes the paper.

## 2 Literature Review

Zhi-Hua Zhou and Yuan Jiang [4] have proposed an approach named C4.5 Rule-PANE, which gracefully combines the advantages of artificial neural network ensemble and rule induction. A specific rule induction approach, i.e. C4.5 Rule, is used to learn rules from the new training data set. Case studies on diabetes, hepatitis, and breast cancer show that C4.5 Rule-PANE could generate rules with strong generalization ability, which profits from artificial neural network ensemble, and strong comprehensibility, which profits from rule induction.

Leung et al., [32] have presented a data mining framework for biological data sets. And it has been applied to the Hepatitis B Virus DNA data sets which are real world data. Their method has good performance using the fuzzy measure and the nonlinear