Chapter 8
Diagnosis of Heart Disease Using Improved Genetic Algorithm-Based Naive Bayes Classifier

Kalaivani Karuppiah
Vignana Bharathi Institute of Technology, India

Uma Maheswari N.
PSNA College of Engineering and Technology, India

Balamurugan N.
Sree Vidyanikethan Engineering College, India

Venkatesh R.
PSNA College of Engineering and Technology, India

ABSTRACT

Heart disease is one of the most common diseases all over the world. The primary objective of this investigation is to diagnosis heart disease using hybrid classification based on NaN prediction and ANOVA test (NAN-ANOVA). The anticipated system comprises of two subsets: hybrid accelerated artificial bee colony and chicken swarm optimization algorithm (AABC-CSO) for effectual feature selection, followed by a classification technique with genetic algorithm based naive bayes classifier (GA-NBC). The first system in co-operates three stages: (i) loading the numerical value from the dataset (ii) evaluating the NaN value (iii) performing ANOVA test for efficient selection using AABC-CSO optimization algorithm. In second method, GA-NBC is proposed. The heart data set obtained from UCI machine repository, and was utilized for performing the computation. An accuracy of 61.0777%, sensitivity of 31.5868%, specificity of 67.8467%, precision of 17.9505, F-measure of 23.4050, G-mean of 46.6928 and loss of about 0.4480 was achieved according to the validation scheme.

DOI: 10.4018/978-1-6684-5741-2.ch008

INTRODUCTION

Cardiovascular disease (CVD) is one amongst the primary diseases that causes death every year. An approximation of roughly about 17.5 million people dies due to CVD, signifying about 31% of global deaths. Based on the statistics, heart disease kills one person for every 34 seconds (Karegowda et al., 2010).

Various factors are related to the prediction of heart disease, which complicates the task of physician. In order to assist the physician to take quick decision and to reduce the diagnosis error, classifiers help to quickly examine the medical data with considerable outcome (Rahim et al., 2021). These systems are executed by developing a replica that can classify prevailing records with sample data. Diverse classification algorithms have been generated and utilized as classifiers to support doctors in effectually diagnosing patients with heart disease.

The performance attained while using Statlog (Heart) dataset (Subbulakshmi et al., 2012) from UCI machine learning database are in contrast for evaluating the accuracy. Chicco and Jurman (2021) anticipated a new supervised feature selection technique based on Euclidean distances and bounded sum of weighted fuzzy membership functions (BSWFM) and calculated accuracy of 87.4%. Tomar and Agrawal (2014) cast off Least Square Twin Support Vector Machine (LSTSVM) and F- score features selection approach to recognize heart disease and attained an accuracy of about 85.59%. Li et al. (2019) trained the datasets with Input Selection and Testing (TWIST) procedure to classify the patterns to acquire an accuracy of about 84.14%. Extreme Learning Machine (ELM) classifier reported classification accuracy of about 87.5% (Buscema, Breda, and Lodwick, 2013). Naïve Bayes classifier has shown an accuracy rate of about 85.87% in Mohan, Jain and Agrawal (2021). Srinivas, Rani and Govrdhan (2010) attained classification accuracy of about 83.70%.

In Polat and Güneş (2009) cast off *F*-score and RBF kernel feature selection method to identify heart disease. Here, LS-SVM classifier is utilized to attain an accuracy of about 83.70%. Exley et al. (2022) used GA-AWAIS method to identify and attain heart disease with 87.43% accuracy. Similarly, in Peng et al. (2021), Algebraic Sigmoid Method was used to categorize heart disease and report accuracy as 85.24%. In Mahabadi et al. (2014) and Amin et al. (2013), linear kernel SVM classifiers and three distance criteria was applied to identify heart disease with Statlog (Heart) dataset. The accuracy attained was 83.37% and 83.95% respectively.

Hybrid neural network method was anticipated in Amma (2012) and the accuracy rate was 86.8%. With the use of SVM and ICA classifier, in Venkatalakshmi and Shivsankar (2014), the accuracy was 83.75%. In Dangare and Apte (2012a) artificial immune system termed Attribute Weighted Artificial Immune System (AWAIS) with *k*-fold cross-validation was utilized to get an accuracy of 82.59%. Feature space mapping (FSM), *k*-NN with Manhattan, and separability split value (SSV) procedures were utilized for predicting heart disease with highest accuracy of about 85.6%.

From the above discussed works, it is noted that selecting the features can significantly enhances the classifier performance for diagnosing heart disease [19]. Dependency and the noisy features in heart disease prediction dataset influences diagnosis process. In general, there exist numerous records for the syndromes accomplished with the original datasets along with the huge redundant symptoms. Subsequently, it is essential to diminish the dimensions of original feature set by feature selection technique that can eradicate redundant and irrelevant features.

Heuristic based optimization is one amongst the most popular technique for successful feature selection algorithm. It can accurately predict feature quality with the dependency relationship [20]. There are two major advantages of using the heuristic algorithm: i) it does not facilitate domain specific knowledge