Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques:
Heart Disease Prediction

Dhyan Chandra Yadav, Veer Bahadur Singh Purvanchal University, Jaunpur, India

Saurabh Pal, Veer Bahadur Singh Purvanchal University, Jaunpur, India

https://orcid.org/0000-0001-9545-7481

ABSTRACT

This paper has organized a heart disease-related dataset from UCI repository. The organized dataset describes variables correlations with class-level target variables. This experiment has analyzed the variables by different machine learning algorithms. The authors have considered prediction-based previous work and finds some machine learning algorithms did not properly work or do not cover 100% classification accuracy with overfitting, underfitting, noisy data, residual errors on base level decision tree. This research has used Pearson correlation and chi-square features selection-based algorithms for heart disease attributes correlation strength. The main objective of this research to achieved highest classification accuracy with fewer errors. So, the authors have used parallel and sequential ensemble methods to reduce above drawback in prediction. The parallel and serial ensemble methods were organized by J48 algorithm, reduced error pruning, and decision stump algorithm decision tree-based algorithms. This paper has used random forest ensemble method for parallel randomly selection in prediction and various sequential ensemble methods such as AdaBoost, Gradient Boosting, and XGBoost Meta classifiers. In this paper, the experiment divides into two parts: The first part deals with J48, reduced error pruning and decision stump and generated a random forest ensemble method. This parallel ensemble method calculated high classification accuracy 100% with low error. The second part of the experiment deals with J48, reduced error pruning, and decision stump with three sequential ensemble methods, namely AdaBoostM1, XG Boost, and Gradient Boosting. The XG Boost ensemble method calculated better results or high classification accuracy and low error compare to AdaBoostM1 and Gradient Boosting ensemble methods. The XG Boost ensemble method calculated 98.05% classification accuracy, but random forest ensemble method calculated high classification accuracy 100% with low error.

KEYWORDS

AdaBoost, Chi-Square, Decision Stump Algorithm, Gradient Boosting, J48 Algorithm, Pearson Correlation, Random Forest Classifiers, Reduced Error Pruning, XGBoost

1. INTRODUCTION

Virani SS et al., (2020), introduced about heart organ importance. The heart is an important organ of human body in which various vessels support in blood flows to different parts of the body. Our heart is a complex structure consisting of muscles. It suffers from various types of diseases in the heart by more or less blood secretion. The cardiovascular is a heart diseases and the main cause of cardiovascular diseases is poor lifestyle, stress, coma, not exercising and irregular food intake. These life style leads to disability on heart and invites serious heart related diseases. So, healthy heart is very important for human life because it is necessary to keep the body healthy. If a person has been faced the problem of heart attack, then he should change his lifestyle to keep the heart healthy. Sometimes heart diseases show some of their symptoms then we should not avoid it easily.

1.1. Types of Heart Disease

1.1.1. Coronary Artery Disease

This disease is very common, which is caused due to the accumulation of arteries in the arteries and stops the flow of blood in the heart and brings the heart to a state of zero and increases the risk of stroke.

1.1.2. Hyper Tensile Heart Disease

This disease is caused by high blood flow in the arteries, as a result of which the blood vessels become heavy and the incidence related to heart disease problem.

1.1.3. Rheumatic Heart Disease

This disease arises due to damaged heart valve and the patient faced coma as heart disease which is associated with rheumatic fever. Due to this disease, the tissues connecting the brain, the joints of the brain and the skin get affected and these organs become inflamed.

1.1.4. Congenital Heart Disease

This disease is visible due to the lack of proper heart structure at birth, because it is not able to be in normal state due to bad heart condition.

1.2. The Main Causes of Heart Disease

- Increased cholesterol
- Smoking more
- Excess alcohol consumption
- Maintaining a state of constant tension
- Heredity
- Excess body fat
- High blood pressure
- Stretching with pain due to artery block in the heart and increasing the possibility of heart attack
- Excess of nausea
- Chest irritation and imbalance in digestive functions
- Excess of pain in the hands and down the left shoulder
- Excessive phlegm and white or pink color
- Excess of sweat from the body while lack of physical activity

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/analysis-of-heart-disease-using-parallel-and-sequential-ensemble-methods-with-feature-selection-techniques/268417

Related Content