Abstract No: MO-16   **Multidisciplinary Research**

Cardio forecast: A multi criteria weighted voting ensemble model for the prediction of cardiovascular disease level

F. Zahara Kitchilan[1*] and W. G. D. M. Samankula[2]

[1]Northshore College of Business and Technology, Church Road, Colombo 15
[2]Department of Statistics and Computer Science, Faculty of Science,
University of Kelaniya, Sri Lanka
*Email: zahkit@gmail.com

Computational intelligence and data mining in medical domain involve the discovery of unknown patterns and useful knowledge from massive sets of healthcare data. Different machine learning algorithms and statistical analysis tools perform diagnostic procedures. This research presents an intelligent medical knowledge driven approach for the prognosis of heart disease using a weighted voting ensemble classifier. The proposed ensemble classifier overcomes the limitations of single and hybrid data mining techniques by combining outcomes of three heterogeneous classifiers, namely Gaussian Naïve Bayes, Extra Tress and Ada Boost. The clinical data was obtained from processed Cleveland heart disease data set at UCI (University of California, Irvine) Machine Learning Repository. Different pre-processing techniques are used to clean data. There are thirteen (13) dependent attributes and one (1) independent attribute for the prediction in heart disease dataset.

This study executes as quantitative research and many related works on data mining and heart disease reviewed and analyzed methodically to get a deep insight of the field. This study is based on statistics and numeric, therefore the research is quantifiable and various relations found between clinical data features. The main scope of study is to enhance the overall accuracy of cardiovascular disease prediction.

The original dataset holds predictable attributes ranging from 0 to 4, a healthy heart beginning from 0 to a severely unhealthy heart at 4. In the analysis, the multiclass label changed to binary class. The predicted class either 0 or 1, which means the heart is either 0 'negative heart disease' or 1 'positive heart disease'. K-fold cross validation technique is used to evaluate the performance of a proposed ensemble model. The weighted voting ensemble classifier carries out high categorization and diagnosis accuracy by enhancing several performance metrics to a reasonable margin when comparing to other individual and ensemble classifiers. It is obvious from the outcomes that the proposed voting classifier deals with all types of attributes and outperforms the highest accuracy of 92.11%, F1 score 91.89%, precision 92.24% and recall 92.47%. For the proper conclusion for this research to be applied clinically, it requires to further elaborate with the aid of professionals in both cardiac and machine learning domains.

Keywords: Ada Boost, Binary class classification, Extra Tress, *Gaussian Naïve Bayes*, *K*-fold cross validation