

# Identification of Predictive Capability of Classifiers for Early Heart Disease Detection Using Machine Learning

Aquila Peeran<sup>1</sup>, U. Brinda Kumar<sup>2</sup>, N. Neha<sup>3\*</sup>, Nikita Ravi<sup>4</sup>

<sup>1,2,3,4</sup>Department Computer Science Engineering, Dayananda Sagar College of Engineering, Bengaluru, India

**Abstract:** Heart and cardiovascular diseases are the leading cause of death in today's world. In this era, even the younger generations are affected due to unbalanced and fast paced lifestyles. While most people can afford diagnostic tests, families with lower income cannot afford the cost of these expensive tests. This prediction system aims to aid such individuals in addition to issues such as lack of physicians in rural areas and places with low healthcare quality. By providing a prediction model for heart diseases at an early stage, this project helps reduce the cost of medical tests and the errors associated with it are also considerably reduced compared to manual testing. Since the model helps in predicting the diseases at an early stage, the dire consequences can be controlled and lifestyle changes can be made to reduce the further risks associated with a heart disease. The added feature of instant diagnosis can be very useful in case of an emergency. We check the capability of Deep Learning classifiers for cardiovascular disease identification and prediction in this paper along with a rigorous process of data mining to remove noisy data for a better decision making system with an extremely effective accuracy.

**Keywords:** Cardiovascular, Deep Learning, Prediction system, Data Mining.

## 1. Introduction

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Among all fatal diseases, heart attacks diseases are considered as the most prevalent. Due to the increase in workload, unhealthy diets and fast paced lifestyles, the younger generations have also fallen victim to heart complications.

The diagnosis of heart disease is usually based on signs, symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level, family history of heart disease, obesity, high blood pressure, and lack of physical exercise. However, with early diagnosis of heart disease, can help in making decisions on lifestyle changes in high-risk patients and help reduce the complications.

Health care industries collect large amounts of data that contain some hidden information, which is useful for making

effective decisions for providing appropriate results, using the data collected effectively we can make predictions and the results can be used to prevent and thus reduce cost for surgical treatment and other expensive tests.

## 2. Methodology

The dataset collected is the well verified UCI repository dataset. The proposed system is used for heart disease identification and prediction. The proposed system methodology has modules of Data preparation, Data mining, Splitting Data, Classifier Performance, Clinical Decision support system. The dataset undergoes a rigorous process of data mining in order to remove noisy data for a better decision making system with an extremely effective accuracy.

## 3. Literature Survey

AMIN UL HAQ et al. [27] proves the BPNN achieved 93% accuracy. Finally based on all experimental analysis and results that deep neural network algorithm (BPNN) is more effective for detection of heart disease. I Ketut Agung Enriko et al. [31] system has been trialled in three rural locations near Jakarta, which give results: prediction accuracy 76.47%, kNN processing time 1 seconds, and transfer time 8.97 seconds. The auto-diagnosis system using kNN algorithm is the novelty in this research. Sonakshi Harjai et al. [7] aimed an efficient approach towards prediction of heart disease using - Improved MLP classification. It adopted feature selection using correlation- based algorithm measures for improving the efficiency of heart disease classification. It was also found that the model can be improved by reducing the relative absolute error and the root mean square error of the classifier. Rishabh Magar et al. [2] found out that Logistic Regression algorithm has the most efficient out of the four with an accuracy of 82.89%. Decision tree and Naive Bayes had accuracy of 80.43% and 80.43% respectively, and SVM was having 81.57%. Thus these four algorithms were further implemented using a better user interface. Galla Siva Sai Bindhika et al. [16] found that by introducing new proposed Random forest classification, we find the problem of prediction rate without equipment and propose an approach to estimate the heart rate and condition.

\*Corresponding author: neha.nghode@gmail.com

Sample results of heart rate are to be taken at different stages of the same subjects, we find the information from the above input via ML Techniques. Archana Singh *et al.* [6] found that the accuracy of the algorithms in machine learning depends upon the dataset that used for training and testing purpose. Fahd Saleh Alotaibi [28] research discussed, proposed and implemented a machine learning model by combining five different algorithms. Rapid miner is the tool used in this research, which computed the high accuracy than Matlab and Weka tool. Dinesh Kumar G *et al.* [17] contributes the correlative application and analysis of distinct machine learning algorithms in the R software which gives an immediate mechanism for the user to use the machine learning algorithms in R software for forecasting the cardiovascular diseases. Rahul Katarya *et al.* [20] have summarized some algorithms which are useful while selecting the features, like hybrid grid search algorithm and random search algorithm, etc. So, in the future, it is better to use search algorithms for selecting the features and then applying machine learning techniques for prediction will give us better results in the prediction of heart disease.

#### 4. Conclusion

The proposed working model helps in reducing treatment costs by providing Initial diagnostics in time. The model can also serve the purpose of training tools for medical students and will be a soft diagnostic tool available for physicians and cardiologists. By using the data collected effectively we can make predictions and the results can be used to prevent and thus reduce cost for surgical treatment and other expensive tests.

#### Acknowledgement

Prof. Bhavana B. R. of Dayananda Sagar College of Engineering will be guiding us through the project.

#### References

- [1] Amin Ul Haq *et al.*, "Heart Disease Prediction System Using Model of Machine Learning and Sequential Backward Selection Algorithm for Features Selection", IEEE 5th I2CT 2019 Pun international conference, India, 2019.
- [2] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chest X-ray 8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in Proceedings of IEEE CVPR 2017, Honolulu, HI, USA, 2017.
- [3] Lamliili, E. N., Boutayeb, A., Derouich, M., Boutayeb, W., & Moussi, A. Fish Consumption Impact on Coronary Heart Disease Mortality in Morocco: A Mathematical Model with Optimal Control. *Engineering Letters*, vol. 24, no. 3, 2016.
- [4] Shigei, Noritaka, *et al.* "Land-Use Classification Using Convolutional Neural Network with Bagging and Reduced Categories." Proceedings of the International Multi Conference of Engineers and Computer Scientists IMECS 2019, pp. 13-15, 2019.
- [5] Tu, M. C., D. Shin, *et al.* "Effective Diagnosis of Heart Disease through Bagging Approach." *Biomedical Engineering and Informatics*, IEEE, 2009.
- [6] Enriko, I., Wibisono, G and Gunawan, D. Designing machine-to-machine (M2M) system in health-care modeling for cardiovascular disease patients: Initial study. In *Information and Communication Technology (ICoICT)*, 3rd International Conference pp. 528-532. IEEE 2015.
- [7] K. Srinivas, K. Raghavendra Kao, and A. Govardham, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques," -Pronk, John M. Clymer and the Community Preventive Services Task Force (May 2017).
- [8] Rivero, Jesus E., *et al.* "Thermal neutron classification in the hohlraum using artificial neural networks." *Engineering Letters*, Vol. 23, No. 2, pp. 87-91. 2015
- [9] Rukchart Prasertpong, Manoj Siripitukdet, "Rough Set Models induced by Serial Fuzzy Relations Approach in Semigroups." *Engineering Letters* vol. 27, no. 1, 2019.
- [10] Polaraju, K., Durga Prasad, D., & Tech Scholar, M. Prediction of Heart Disease using Multiple Linear Regression Model. *International Journal of Engineering Development and Research*, vol. 5, no. 4. 2017.
- [11] Patel, J., Upadhyay, P. and Patel, "Heart Disease Prediction Using Machine learning and Data Mining Technique" *Journals of Computer Science & Electronics*, 2016.