

Early Prediction of Heart Disease: In-Depth Study

(1). Mrs.**G.Arul Geetha**, Research Scholar,
KPR College of Arts, Science and Research, Coimbatore..

(2) .**Dr.T. Deepa**, Associate. Professor & Head,
Department of Computer Science,
KPR College of Arts, Science and Research, Coimbatore.

Abstract

Nowadays it's very common that all age groups of people are suffering from heart disease. After COVID-19 it causes sudden cardiac arrest among the people. According to the World Health Organization (WHO), 17.9 million people depart their lives every year out of which 272 per 100000 population people are from India alone. This paper aims to realize an early prediction of heart disease based on various techniques such as logistic regression, decision trees, random forests, support vector machines, artificial neural networks, and convolutional neural networks. The methods achieve an accuracy of about 90% using the Random Forest. After using 10 epochs, the accuracy of the combination of SVM and ANN is about 90.56%. In this study, a comprehensive and comparative analysis of these techniques is carried out to understand the architectures, parametric characteristics, and datasets involved in heart disease prediction.

1. Introduction

In today's world, machine learning algorithms are extensively utilized across various sectors, including healthcare. These algorithms play a crucial role in early disease prediction, significantly improving patient outcomes and saving lives. Despite these advancements, heart disease remains a leading cause of death globally. Effective early detection through machine learning could potentially reduce the mortality risk associated with heart disease.

The heart is a vital organ, and when it fails to function properly, it can lead to severe complications and rapid death, not only affecting older adults[1]. Diagnosing heart disease is challenging due to its numerous risk factors. Common symptoms include physical weakness, chest pain, shortness of breath, and irregular heartbeats[2]. In the United States, heart disease is alarmingly prevalent, with one person dying every 34 seconds from this condition. Worldwide, approximately 26 million people are affected, and heart disease accounts for 32% of global deaths[4]. Between 2005 and 2015, heart disease cost India around \$237 billion, according to the World Health organization[5][6]. Both men and women are susceptible to heart disease, which often results from prolonged exposure to unhealthy lifestyles [6]. By applying machine learning algorithms to predict heart disease at an early stage, we can significantly improve patient outcomes and potentially save millions of lives. This research aims to evaluate the accuracy of six different machine learning algorithms to determine the most effective method for early heart disease prediction, ultimately contributing to better health outcomes and substantial economic savings.

2.Methods Used

2.1Support Vector Machine (SVM):

The primary analytical tool employed in this study is the Support Vector Machine (SVM), a supervised machine learning algorithm renowned for its effectiveness in classification tasks. SVM is used to analyze ECG data obtained with the goal of predicting the likelihood of heart disease in individuals.[7]

Data Input

The SVM algorithm processes multiple types of data inputs, including:

- ECG Data: This includes heart rate variability and morphological features of the ECG, which are crucial indicators of cardiac health.
- Demographic Data: Factors such as age, gender, and other relevant demographic details are also incorporated to enhance the accuracy of heart disease prediction.

Risk Prediction

SVM is designed to identify individuals at high risk of developing heart disease by analyzing the input data. Its predictive capability is crucial for enabling early intervention and better management of heart disease. SVM as a Specialized Supervised Machine Learning Classifier Support Vector Machines (SVM) are specialized supervised machine learning classifiers used for both linear and non-linear dataset categorization. The primary mechanism of SVM involves a non-linear mapping function that transforms the original data into a higher-dimensional space where the data becomes more separable. In this transformed space, SVM identifies an optimal linear hyperplane that serves as the decision boundary between different classes. This hyperplane is constructed using support vectors, which are the critical data points that help define the margin between classes. SVM is highly effective in creating a precise decision-making boundary, especially in complex datasets where traditional linear classifiers may fail. However, the computational demands of SVM are significant because it requires solving quadratic optimization problems involving complex mathematical functions, which can be time-consuming. The hybrid approach in machine learning, often involving the integration of SVM with other algorithms, combines the strengths of different methodologies to create a more robust and efficient model. By merging distinct elements, each with unique characteristics, hybrid models can leverage the advantages of multiple approaches, resulting in enhanced performance and adaptability in various data-driven applications.

Assessment

The study assesses the performance of the SVM in accurately predicting heart disease risk based on the collected ECG data. The results demonstrate that the SVM algorithm effectively identifies high-risk individuals, marking a significant advancement in the early detection and management of heart disease. This study integrates the ECG data collection with the application of the SVM algorithm. Together, these methods provide a powerful approach to accurately predicting heart disease risk, contributing to improved patient outcomes through early detection.

2.2 Artificial Neural Networks (ANNs)

Heart disease remains a significant global health challenge, and predicting heart disease using artificial neural networks (ANNs) has emerged as a prominent research focus aimed at enhancing the accuracy of early diagnosis. However, challenges such as data complexity, feature selection, and over fitting still need to be addressed in ANN models for effective prediction. The use of ANN in heart disease prediction holds great potential, offering valuable support for medical decision-making, reducing heart disease incidence, and aiding personalized treatment and health management.

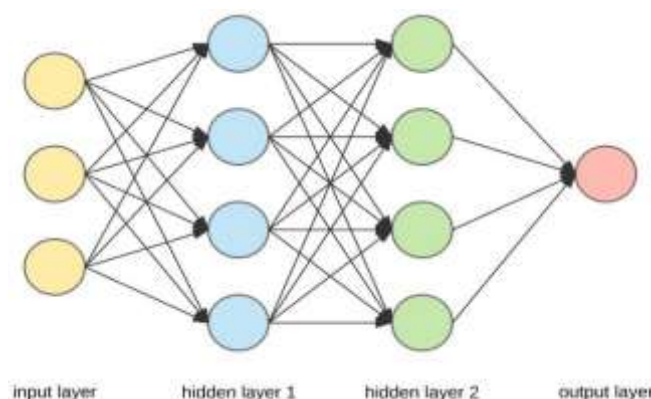


Fig 1: Model Architecture of ANN

This study introduces a heart disease diagnosis approach leveraging machine learning techniques, which balances the dataset using the SMOTE (Synthetic Minority Over-sampling Technique) and ENN (Edited Nearest Neighbors) algorithms. The research explores various machine learning models, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Gradient Boosting Decision Tree (GBDT), XGBoost (Extreme Gradient Boosting), Support Vector Machine (SVM), and ANN (Deep Neural Network) for predicting heart disease.

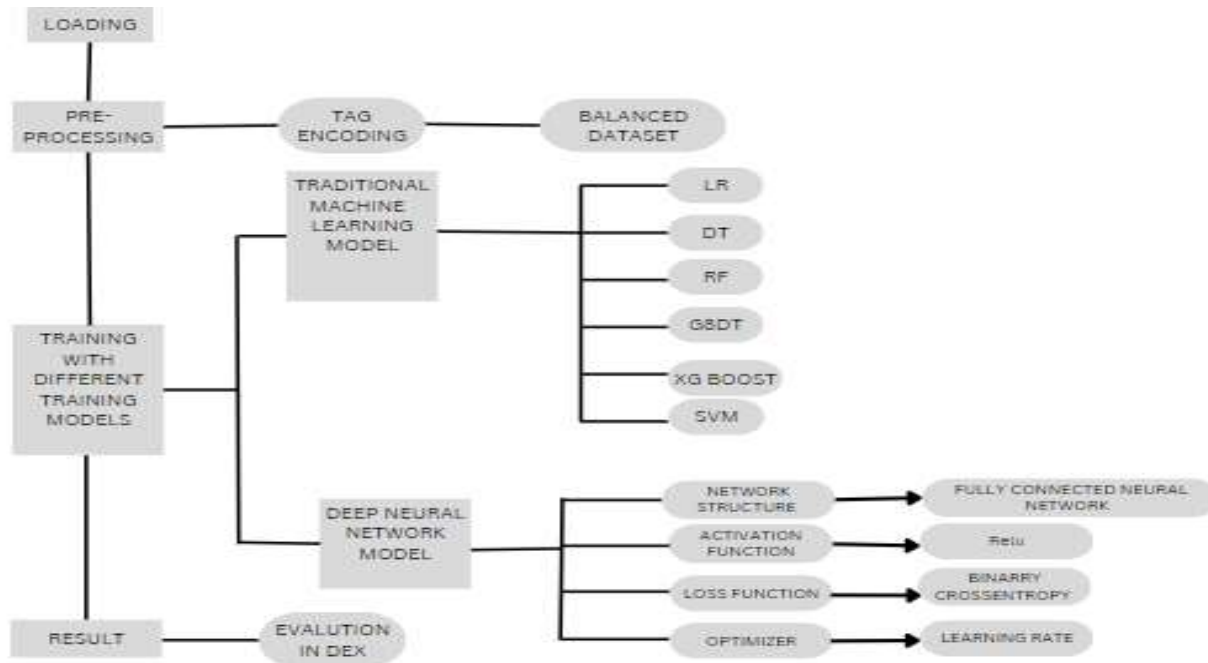


Fig :1 Experiment Workflow

MLP Classifier framework

MLP Classifier [9] is a classifier that applies back propagation Algorithms in multi-layer neural networks to train models that can be used for classification tasks. Multilayer perceptron (MLP), which is called a feed-forward artificial neural network (FFANN) as well, is frequently applied in different areas including audio processing, natural language processing, autonomous vehicles, computer vision, etc. The basic working principle of the MLP Classifier is to classify input data, gather input data through the connections of multiple neurons to the output layer, and ultimately determine the prediction results. In neurons, activation functions are used to determine whether the input signal should activate the neuron. Fig 2 shows the whole framework.

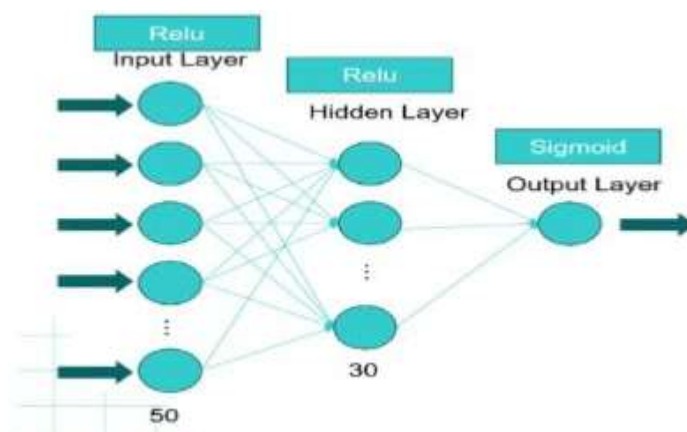


Fig:2 Frame Work

Dataset

Introduction Source of the dataset: The data platform provides the information for this study. In detail, this survey is composed of 319795 people who are at least 18 years old. In addition, the proportion of samples without heart disease to ones with that is approximately 10.7 to 1. The distribution is shown in Figure 3. Initially, this dataset came from a telephone survey conducted by the Centers for Disease Control and Prevention in the United States in 2020, which collected data on the health state of American residents. In the end, this study uses it for heart disease prediction to assist Professional doctors in diagnosing heart disease preliminarily.

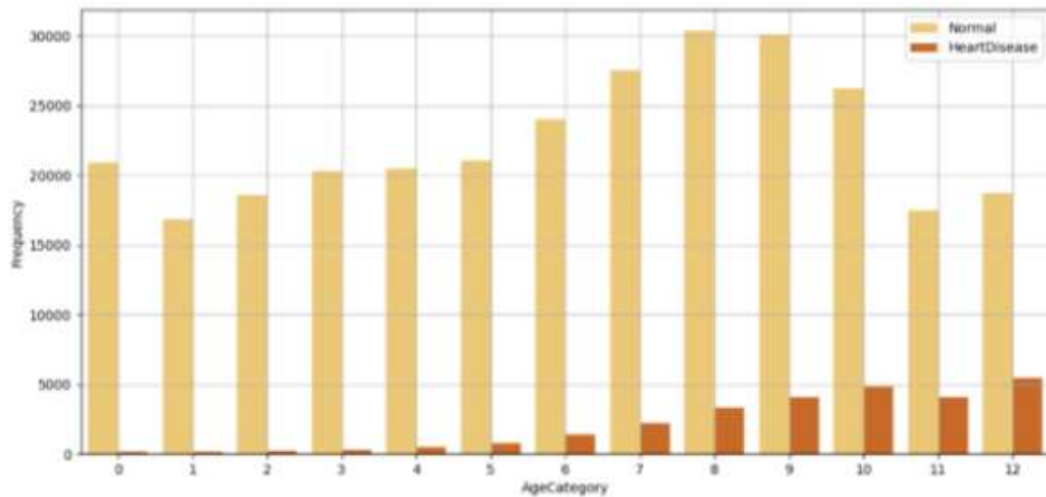


Fig 3: Distribution of Cases with Yes/No Heart Disease according to Age.

3. Results

Comparative Analysis of SVM and ANN for Heart Disease Prediction

The comparison between Support Vector Machines (SVM) and Artificial Neural Networks (ANN) in heart disease prediction highlights their respective strengths and optimal use cases. SVM is frequently recognized for its robustness and efficiency, particularly in high-dimensional spaces where it often surpasses traditional models like Logistic Regression and Decision Trees. Studies show that when combined with other techniques such as Rough Set Theory, SVM can achieve high accuracy rates, with one model reaching 93.8% accuracy. On the other hand, ANN models excel in handling complex and non-linear data patterns, often achieving high accuracy metrics such as an AUC of 0.808 and a recall of 0.81. However, ANN's performance is often challenged by issues like over fitting and data complexity, which can be mitigated through techniques like SMOTE and ENN. When combined, SVM and ANN can further enhance prediction accuracy, as demonstrated in hybrid models where a combination of both methods achieved around 90.56% accuracy after 10 epochs. Therefore, the choice between SVM and ANN depends largely on the specific characteristics of the dataset and the desired balance between accuracy and computational efficiency.

4. Conclusion

A comparison between Support Vector Machine (SVM) and Artificial Neural Networks (ANN) in heart disease prediction reveals that each method has distinct advantages and is suited to different scenarios. SVM is consistently recognized as a dependable and effective model, often surpassing traditional approaches like Logistic Regression and Decision Trees in numerous studies, with some identifying it as the most reliable for heart disease prediction. On the other hand, ANN models excel in accuracy-related metrics such as AUC and recall, with recorded values around 0.808 and 0.81, respectively. Despite these strengths, ANN faces challenges related to data complexity and overfitting. In summary, SVM is noted for its robustness and reliability, whereas ANN shows considerable promise in improving predictive accuracy, particularly in dealing with data complexity and imbalance. The selection between these models may hinge on the specific characteristics of the dataset and the clinical context in which they are applied.

References:

- [1] M. Heron, "Deaths: leading causes for 2008," National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, vol. 60, no. 6, pp. 1–94, 2012.
- [2] M. Durairaj and N. Ramasamy, "A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate," International Journal of Control Theory and Applications, vol. 9, no. 27, pp. 255–260, 2016.
- [3] P. A. Heidenreich, J. G. Trogon, O. A. Khavjou et al., "Fore casting the future of cardiovascular disease in the UnitedStates," Circulation, vol. 123, no. 8, pp. 933–944, 2011.
- [4] G. Savarese and L. H. Lund, "Global public health burden ofheart failure," Cardiac FailureReview, vol. 3, no. 1, pp. 7–11, 2017.
- [5] March 2022, https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.
- [6] C. Beyene and P. Kamat, "Survey on prediction and analysis the occurrence of heart disease using data mining techniques," International Journal of Pure and Applied Mathematics, vol. 118, no. 8, pp. 165–174, 2018.
- [7] T. Bala Krishna, Vimala Nunavath, P.S.N.M. Vinay, Nagudu Siddhardha, P Manohar "Early Heart Disease Prediction Using Support Vector Machine" 01 Jan 2023 (Springer International Publishing)
- [8] Pooja Dehankar, Susanta Das "Detection of Heart Disease Using ANN" 21 Jun 2024 1 - Advances in medical diagnosis, treatment.
- [9] Windeatt, T (2008). Ensemble MLP Classifier Design. In: Jain, L.C., Sato-Ilic, M, Virvou, M. Tsihrintzis, G.A., Balas, V.E., Abeynayake, C. (eds) Computational Intelligence Paradigms. Studies in Computational Intelligence, vol 137. Springer, Berlin, Heidelberg.
- [10] Rehan Ahmed', Maria Bibi 3, Sibtain Syed "Improving Heart Disease Prediction Accuracy Using a Hybrid Machine Learning Approach: A Comparative study of SVM and KNN Algorithm" IJCIM June 2023