A Survey of Heart Disease Prediction using Deep Learning Methods

Ameet Shah, Dhanpratap Singh

Dept. Computer Science & Engineering, Lovely Professional University Punjab, India

Corresponding author: Ameet Shah, Email: ameetshah1981@gmail.com

Heart disease is a prevalent and life-threatening condition that affects a significant portion of the global population. Preventing negative consequences and better health outcomes may be greatly helped by early identification and precise prognosis of cardiac disease. Electrocardiogram (ECG) is the important apparatus which is used for diagnosticians of assessment of cardiogram in the clinical. This will Owing to its ability to learn complicated patterns from massive volumes of data, deep neural networks (DNNs) have recently emerged as strong tools for medical diagnostic and prediction tasks. In this part of research, people investigate the use of deep neural networks for cardiac illness forecasting which is easy way for analysis of cardio problem. We analyse various studies and approaches that have utilized Deep Neural Networks for heart disease prediction. This study gives us whole sheds light on the evolution of methods that use in deep neural networks to forecast cardiac issues. The findings from this survey can guide researchers and practitioners in designing and implementing effective DNN models for heart disease prediction, ultimately contributing to improved clinical decision-making and patient care.In this study, we will examine the various DNN techniques, including CNNs (Convolutional Neural Networks) and Stack Denoising Auto encoder, Deep Belief Network, Recurrent Neural Network (RNN), Long Short term Memory (LSTM) and Gated Recurrent Unit (GRU).

Keywords: Deep Learning, CNN, Stack Denoising Auto-encoders, cardiovascular diseases, ECG

2023. In Satyasai Jagannath Nanda & Rajendra Prasad Yadav (eds.), *Data Science and Intelligent Computing Techniques*, 643–656. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.56155/978-81-955020-2-8-58

1. Introduction

Heart disease is one of the leading causes of mortality worldwide, claiming the lives of millions of people every year.. In order to effectively intervene and prevent cardiac disease, early identification and precise prognosis are essential. There has been a significant uptick in research on how to use machine learning more specifically DNNs (deep neural networks to enhance heart disease forecasting. The structure along with function of the human brain served as inspiration for the development of a subset of artificial neural networks called deep neural networks. Neurons, which are the building blocks of these systems, are linked at numerous levels and are capable of processing and learning from massive amounts of data. DNNs' capacity to automatically extract complicated patterns and features from raw data, allowing them to capture deep correlations and make correct predictions, has garnered them a great deal of interest. The application of deep neural networks in heart disease prediction offers several advantages over traditional methods. Firstly, DNNs can handle large and diverse datasets, incorporating various types of data such as clinical records, medical images, genetic information, lifestyle factors, and sensor data. This comprehensive approach allows for a more holistic understanding of the risk factors associated with heart disease. Secondly, DNNs have the potential to capture both linear and nonlinear relationships between input variables and the target outcome, making them highly suitable for modelling the complex nature of heart disease. Traditional statistical models often assume linearity, which may not adequately capture the intricate interactions among multiple risk factors involved in heart disease development. Also, deep neural networks may learn useful features and representations automatically from the data, cutting down on the requirement for human-engineered features. The identification of significant risk variables may be difficult in the context of heart disease prediction owing to the availability of several possible predictors and their complicated interactions; therefore, this feature is very useful. The predictive power of deep neural networks for cardiovascular disease has been established in several research. To make trustworthy predictions, scientists have experimented with a wide range of DNN designs. These include feed-forward neural networks, CNNs, RNNs, and combinations of these. These architectures allow for the extraction of spatial and temporal patterns from medical images, time-series data, and sequential records, enhancing predictive performance. Additionally, researchers have investigated different pre-processing techniques to handle missing data, normalize features, and reduce noise, thereby improving the robustness and generalizability of DNN models. Feature selection methods, both manual and automated, have been employed to identify the most relevant predictors for heart disease, enhancing the interpretability and efficiency of the models. Commonly used evaluation criteria for DNN models in heart disease prediction for example precision, specificity, sensitivity, accuracy, and AUC-ROC. These measures help in clinical decision making and risk stratification by revealing how well the model distinguishes between people with and without cardiac disease. There are still obstacles and limits to using deep neural networks to forecast cardiac disease, despite the encouraging findings. The interpretability of DNN models remains a concern, as their black-box nature makes it challenging to understand the underlying features and mechanisms that contribute to the predictions. Furthermore, the availability of large and diverse datasets, which are essential for training accurate models, can be a limiting factor. The utilization of deep neural networks in heart disease prediction holds significant promise for improving the accuracy and efficiency of risk assessment. By leveraging the power of DNNs to learn complex patterns and relationships from diverse data sources, accurate predictions can be made, enabling timely interventions and personalized treatment strategies. Better patient outcomes and a less burden of heart disease will result from overcoming the obstacles of interpretability and data availability, which will lead to the wider use of DNN-based models in clinical practise. Heart Disease with additional correct and perfection. In this review paper, we are going to propose a Deep learning Process (Ali Isina, ICSCCW 2017, 24-25 August) along with deep learning methods. The neural network used in deep learning is a kind of machine learning, which saves information and captures data at several levels of a hierarchy. These days, deep learning and computer signal processing are the two main components of automatic classification technologies. The first category of automated ECG analysis methods includes, but is not limited to, frequency analysis, decision trees, KNN, SVMs, and ANN [1-5].

1.1 ECG Wave Form Analysis and Description

There are several beats in each cycle of the ECG signal, and each beat comprises multiple waves like the PQRST wave. There are regular amplitude and duration portions of signals such as PR and ST at each of the P, Q, R, S, T, and U points. ECG aspects include the peak, the gap between peaks, and the segments between peaks. Features extracted from an electrocardiogram wave representing the heart cycle are shown in Figure 1, with a corresponding explanation in Table I.

Table 1 Description of ECG Waveform

FEATURES	Description	Duration
RR	It's the time between the present R wave and the one that follows it.	0.6-1.50
P	It's the first brief wave in an electrocardiogram.	80-90 MS
PR	The onset of the P wave and the QRS complex are what are being monitored.	120-200 MS
ORS	The beginning of the P wave as well as the beginning of the QRS complex are what are being looked at and watched.	85-120 MS
PR	P wave and the ORS complex wave are both present.	60-120 MS
J-point	J-point refers to the location where the ORS complex wave terminates and the ST Segmentation wave starts. This is the point when the wave segmentation begins.	Not Applicable
ST	It is connected to the complexwave of T	80-125 MS
T	Mainly is in the upwarddirection of the waveform	160-180MS
ST	It is measured in the J point andends with a T wave	320 -330 MS
QT	The QRS complex onset to T wave termination is the metric used.	420 MS
U	The waves are weak and it is frequently non-existent.	Not Applicable

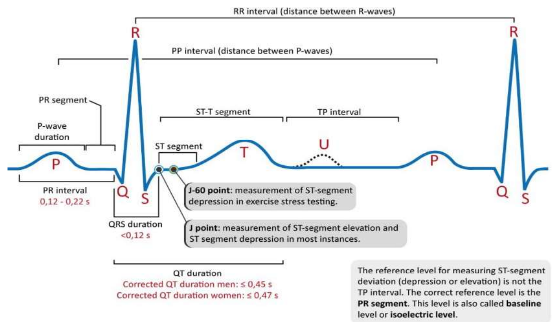

Figure 1 Wave Form of ECG Data

Electrical cardiac activity is represented by the waveform of an electrocardiogram (ECG). Electrodes attached to a patient's skin pick up and record the electrical impulses sent out by the heart at each beat.

The ECG waveform consists of several distinct components and intervals:

1. P-wave: The P-wave is symbolic of atrial depolarization, which refers to the electrical activity of the atria that occurs when the atria are contracting to pump blood into the ventricles.
2. QRS complex: The QRS complex is a representation of the electrical activity of the ventricles when they contract to pump blood out of the heart. This activity is caused by the contraction of the ventricles. The phenomena may be broken down into three distinct waves: the Q-wave, the R-wave, and the S-wave.
3. T-wave: The T-wave indicates that the ventricles have recovered and are ready to begin the next cardiac cycle.
4. PR interval: The amount of time that elapses between the start of the P-wave and the start of the QRS complex is what is measured by the PR interval. It is a measurement of the amount of time required for an electrical signal to get from the atria to the ventricles.

5. QT interval: The QT interval is the period that elapses between the start of the QRS complex and the end of the T-wave. Time spent in both the depolarized and repolarized states of the ventricles.

1.2 ECG Analysis

ECG analysis involves the interpretation of the waveform to assess heart's electrical activity and identify any abnormalities or indications of heart disease. Here are some key aspects of ECG analysis:

1. Heart Rate: The heart rate may be determined from the ECG waveform by determining how much time elapses between consecutive R waves. This provides valuable information about the heart's rhythm and overall cardiac health.
2. Rhythm Analysis: By examining the regularity of the P-waves and R-waves, healthcare professionals can determine if the heart is beating in a regular or irregular rhythm. Irregularities may indicate conditions such as a trial fibrillation or heart block.
3. ST Segment Analysis: That part of the waveform between the S-wave and the T-wave is called the ST segment. Myocardial ischemia (reduced blood supply to the heart) or infarction may be indicated by ST segment deviations from baseline (heart attack).
4. QRS Complex Analysis: Abnormalities in the QRS complex, such as widened or prolonged QRS duration, can suggest conduction abnormalities within the ventricles, like bundle branch blocks.
5. T-Wave Analysis: Changes in the T-wave morphology, amplitude, or duration may indicate electrolyte imbalances, cardiac drug effects, or myocardial ischemia.
6. Diagnosis and Treatment: ECG analysis aids in diagnosing various cardiac conditions, including arrhythmias, conduction disorders, ischemic heart disease, and electrolyte imbalances. The findings from the analysis guide healthcare professionals in determining appropriate treatment strategies and interventions.

2. Literature Review

For Establishing a framework for future researches, many researches has been trying to connected with problem of heart detection using the Machine learning and now with deep learning. Most of the user uses the Cleveland, Physionet and MIT-Arrhythmia Repositories of heart disease dataset. In this Condition the work in [6] for diagnosis of heart illness used a logistic regression classification technique and had accuracy of 77%.

Purushottama. C, KanakSaxenab, RichaSharma[7] explain an efficient framework that can tell us about the predicting factor which effect on heart disease. The aim of the study is to find out the factor effect heart disease. The method used is SVM, Bayesian Classifiers and C4.5.
Chaithra N and Madhu B [8] examines different sorts of approaches used to build up a predictive model concerning cardiovascular diseases. The algorithms used as Decision Tree, J48, and Naïve Bayes and Neural Network.

Kipp W. Johnson, BS, Jessica Torres Soto, MS, Benjamin S. Glicksberg [9] work on the medical practitioners on significant attributes of AI and ML. The algorithms used are of Neural Network and Deep Learning.

Cardiovascular disease prediction models were also proposed by Sumit Sharma et al. [10]. When it comes to improving the overall quality of cardiac disease classificationUsing Hyper Parameter Optimization techniques using Talos with accuracy of 90.78%,

SarangamKodati&Dr. R. Vivekanandam [11] work on the some of the factors which find it out heart disease and reasons for it.They Used the algorithm like SVM, Navive Bayes and KNN Algorithm.
Wankhede et.al [12] to reach an accuracy of 80% using the MLP (Multi-layer perceptron) classifier. Another author Allahverdi et.al [13] uses Neural Network and form a classification method for cardiovascular system which has accuracy of 82%.

Awang et al [14] used the method Naïve Bayes (NB) and Decision Tree (DT) found the accuracy of 82% and 80% respectively.

Key areas where the healthcare industry is having trouble were identified by AnandhavalliMuniasamy et al.[15] as computer-aided diagnostics, illness prediction, data integration, electronic record administration.

Based on the information provided, Chaithra N and Madhu B [16] analyse numerous DM methodologies utilised to construct a prediction about cardiovascular illnesses. These include Neural Network algorithms, Naive Bayes, J48 and Decision Tree. This strategy is developed to evaluate the predictive accuracy of NNs for heart disorders.

According to research conducted by ChalaBeyene and PoojaKamat[17], not all doctors are equally equipped to make life or death judgments. Some physicians have trouble making the right choices, which may have life-threatening consequences for their patients. Predicting the emergence of diseases is crucial for addressing such problems. SVM, KNN, naive bayes, decision tree, and ANN methods are used here. This approach has the drawback that not all possible implementation algorithms can have their performance predicted. The method's benefits arise from its use of many feature selection techniques and algorithms to improve and optimise the current decision-making system. The suggested method aids in the forecasting of heart illness to accomplish an early automated diagnostic and quickly retrieve data. Christalin Beulah Latha, Carolin Jeeva, S.[18]:

After applying ensemble methods like as bagging and boosting to the medical dataset, the accuracy of weak classifiers was improved by 7 percent.Manuel Fet al. [19] Different ML techniques were evaluated on the Framingham Heart Study dataset, includingSVM, NNs,Random Forest, Logistic regression, and Decision Trees. When applied to various models, the study's various algorithms produced contradictory findings.

Mohan Senthil, Thirumalai Chandrasekhar, and SrivastavaGautam[20] in this study, we use machine learning for cardiovascular illness forecasting. The overall efficiency was 88.7 percent accurate.

Laiqat Ali, et al. [21] Predictions of cardiac failures were made using two SVM models. The experimental findings reveal that the proposed model outperforms baseline SVM models by 3.3%.

The results were anything from 57.85% to 91.83 % accurate. H.Jayasree, Dr. S. S. K. T. Kumidini, R. T. Naren, and K. SaiSankeerth [22] said that HDPS uses machine learning to classify the likelihood of a person developing heart disease into one of five categories: extremely high, high, medium, low, or no.

AnkurBiswas, PrithaDatta, Krishna Roy, TanureeDey, and SanchayitaDhar [23] this method developed a hybrid machine learning algorithm that combined a Random Forest classifier with a basic KNN, and then compared the resulting predictions with those of theNaive Bayes as well as J48 tree classifier. Though several researchers have been working on ECG signal classification algorithms such as linear discriminant classifier as well as EM clustering algorithm [24], that can operate autonomously and proficiently perform feature classification or clustering, a patient-adaptable algorithm for ECG heartbeat classification is reported in [25].

This approach is patient-specific since it employs a linear discriminant classifier and an expectation-maximization clustering algorithm to categorize and cluster data derived from RR interval series. Many different ECG datasets were used to evaluate the system, demonstrating its superior performance in classifying heartbeats.

There are still challenges and space for development in the automated diagnosis system for electrocardiogram (ECG) arrhythmia. Classification accuracy of 99 percent was attained by another enhanced classifier published in [26] and employed in the automated diagnostic system for the classification of ECG arrhythmia.

This diagnostic method uses a mix of fuzzy clustering neural network algorithms to classify ten distinct forms of arrhythmia from the MIT-BIH database. Fuzzy C-mean clustering and NN for classifying cardiac arrhythmias are both defined in the literature [27].

In this work, we describe a model that filters ECGs, extracts feature from the RR interval wave using a wavelet transform, performs a fuzzy c-mean clustering-based pre-classification, and uses NNs to do the final classification. In this research, a 99.99 percent success rate in classifying data was observed. Another system for classifying cardiac rhythms, based on a synthetic network and fuzzy relationships, was utilised with an accuracy of 80%-85% in [28].

To achieve 85% accuracy, a heartbeat algorithm using ANN and fuzzy relation is used in the published literature [30]. This finding indicates that there is room for improvement in the feature extraction process; M. K. Das et al. offer a classification system using Machine Language Programs (MLPs) and Neural Networks (NNs) [31], which outperform previous methods.

In the literature [29] [32].An ANN-based approach to ECG picture classification is provided, giving us a new avenue to explore in the investigation of signals using images.Broad and foretelling classifications were found via a convolutional neural network with 13 deep layers, as described by Acharya et al. [33].

Jaymin Patel, Prof.TejalUpadhyay, Dr. Samir Patel [34] [35] carried out among various techniques of DT (Decision Tree) which will classify revealing the one effective performance concerning the diagnosis of the heart disease by adopting.WEKA. Algorithms that are being tested includes: Logistic model tree algorithm, J48, Decision Tree (DT) and Random Forest algorithm. Drawback of this method is accuracy and scalability of the system need to improve. The Advantage of this method is its fetch out hidden patterns with the help of DM Method and also anticipating heart disease in patients where the presence is rated.

Table 2 Various Method of Machine Learning Approach.

Sr. No	Author	Year	Methods	Accuracy
1	Kahramanli, K., Allahverdi, N.	2008	Hybrid Network With ANN(Artificial Neural Network and Fuzzy Neural Network	82 %
2	Palaniappan, S., Awang, R.,	2008	Navi Bayes(NB) and Decision Tree	82 % and 80 % Respective
3	Gudadhe, M., Wankhade, K., Dongre, S.,	2010	MLP-Multi Layer Petceptron	80%
4	Purushottama et al. [1]	2016	SVM, CMAR, Bayesian Classifiers and C4.5	89.7%
5	KaanUyar et al. [9]	2017	Genetic algorithm, RFNN - recurrent fuzzy neural networks	97.78
6	Chaithra N et al. [2]	2018	Decision Tree (DT),J48, NB (Naive Bayes) and NN (Neural Network)	95.66%
7	Kipp W. Johnson et al. [3]	2018	NN (Neural Networks) and Deep Learning.	94.28%
8	SarangamKodati and Dr. R. Vivekanandam	2018	SVM, NB (Naïve Bayes), Random Forest and KNN	91.37%
9	ChalaBeyene et al. [5]	2018	SVM, DT (Decision Tree), KNN (K-Nearest Neighborhood), NB (Naïve Bayes) and ANN (Artificial Neural Network)	94.56
10	S. Bagavathy et al. [10]	2018	SVM and K-NN	89.00%
11	Marinho et al.[15]	2019	NB, SVM, OPF	94.30%
12	Yildirim et al.[16]	2019	CAE and LSTM	99.00%
13	SumitSharma, and Mahesh Parmar. 2020	2020	Hyper Parameter Optimization Techniques using Talos	90.78%
14	Elnawawy, M., Sagahyroon, A., Shanableh, T.	2020	Naïve Bayes (NB) and SVM(Support Vector Machine)	85

15	AnandhavalliMuniasamy et al.	2020	computer-aided diagnostics,	81
16	Li et al.[17]	2020	Deep residual network	99.06%
17	Pandey et al.[18]	2020	LSTM	99.37%
18	Zheng, Z et al [19]	2020	CNN-LSTM	99.01%

This paper's following sections are structured as follows: Section III below presents the theory, structure, and data-processing processes of the CNN model; the experiments, methods, and comparisons to other algorithms are presented in Section IV, and the thesis is summed up and restated in Section V.

3. Methodology

While developing the methodology the problem statement and well structure dataset is significantly important for the effective computational model performs. Many of the researchers used many well available dataset on the different repository like Kaggle, MIT –Arrhythmia and Many More for much trained and well tested dataset. For the proper approach of the system, person'sboth health demands & decreased healthcare costs have a move toward personalised treatment. The process of deep learning and predictive analytics has begun (see Fig. 2). There are many tools like (Keras, TensorFlow), frameworks, and approaches which is available within the realm of deep learning allow it to tackle issues relating to creation, prediction, discovery, comprehension, perception, and categorization. The current technology curve, which is significantly increasing domains, has begun to play a critical role in the production of vast volumes of medical care information practises. CNN, DNN, and exploratory RNN models are suggested for cardiovascular illness prediction. This representation of this CNN, DNN & RNN is going to be uses by many researchers for the finding out the accuracy but before that there is process which should be follow for finding the prediction of heart with proper values to be inputted to system.

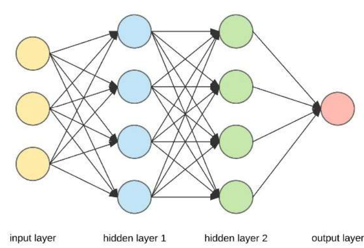

Figure 2 A representation of Deep Learning with two hidden layers.

This is Processing of datais going to be divided into different process forconduction like, Dataset and Pre-processing Techniques, Feature Extraction, Selection: Model Selection, Training, and Validation, Model Evaluation, Model Optimization, and Testing: To begin with, the wave signal threshold mechanism used to filter high-frequency noise, while additional wavelet transformations and the work primarily are used to rectify low-frequency noise baseline drift. The R peaks that were discovered using the wavelet transform approach were used to segment ECG signals with a decrease in dimension. Finally, to conclude the feature extraction and categorization of ECG signals, the treated heartbeat segments are immediately fed into the CNN model as input data.

This is the step which is used for execution of the data (Figure 3).

Step 1: Define the objective of the Problem Statement: By accurately predicting the likelihood of heart disease, healthcare professionals can identify high-risk individuals and provide timely interventions, preventive measures, or necessary medical treatments. This objective ultimately aims to improve patient outcomes, reduce mortality rates associated with heart disease, and optimize healthcare resource allocation by focusing on individuals who are more likely to develop the condition.

Step 2: Data Gathering: The data typically consists of both medical and lifestyle-related information about individuals. Here are some common types of data that are collected for heart disease prediction: 1) Demographic Information: 2) Medical History: 3) Symptoms and Complaints:4) Lifestyle Factors: 5) Physical Examination: 6) Diagnostic Tests 7) Follow-up Data:

Step 3: Data Preparation: Data preparation is a crucial step in heart disease prediction as it involves organizing, cleaning, and transforming the collected data into a format suitable for analysis and model training. Here are the key steps involved in data preparation for heart disease prediction:
1) Data Cleaning:2) Feature Selection: 3) Data Transformation: 4) Data Splitting 5) Handling Imbalanced Data: 6) Feature Engineering: 7) Data Normalization: 8) Data Validation:

Step 4: Exploratory Data Analysis: Exploratory Data Analysis (EDA) is an essential step in heart disease prediction to gain insights, understand the data distribution, identify patterns, and explore relationships between variables. Here are some key components of EDA in heart disease prediction:1) Descriptive Statistics: 2) Data Visualization: 3) Class Distribution:4) Correlation Analysis: 5) Feature Relationships 6) Data Distribution 7) Outlier Detection: 8) Feature Importance 9) Data Quality Check: 10) Subgroup Analysis:

Step 5: Building a Machine Learning Model: Building a machine learning model for heart disease prediction involves several steps. Here is a high-level overview of the process: 1) Data Preparation 2) Model Selection 3) Train-Test Split: 4) Model Training:5) Model Evaluation 6) Model Optimization 7) Model Interpretation 8) Deployment:9) Model Monitoring and Updating:

Step 6: Predictions: Predictions in the context of heart disease involve using a trained machine learning model to predict the likelihood or risk of an individual having heart disease based on their input data.

Step 7: Model Deployment: Model deployment in the context of heart disease involves making the trained machine learning model available for use in a real-world setting, such as a healthcare system or a software application.

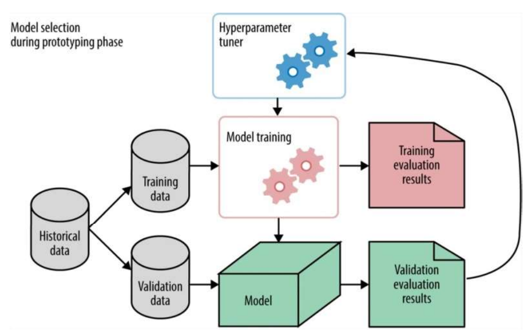

Figure 3. Step Wise Execution of the Data:

4. Classification Strategies

In this Classification algorithm, they exit many classification algorithms, but few among those which is well known algorithm are only investigated. The classification strategies used include Naive Bayes, Random Forest, KNN, Logistic Regression, and SVM. Table 2 will demonstrate, the accuracy using the UCI heart disease dataset which is been uses by some many research up to now.

Table 3 Algorithm with Accuracy implemented by other authors

Algorithms	Accuracy
Logistic Regression	85.25%
K-nearest neighbors	90.16%
Support Vector Machine	81.97%
Random Forest Classifier	85.15%
Naïve Bayes	85.25%

4.1 Support Vector Machine (SVM)

In SVM it is a supervised learning approach for regression as well as classification that includes prediction tools to improve accuracy. It implements a linear function's hypothesis space and has been utilized in a variety of applications, most notably pattern classification techniques. In the initial form, it has a classifier as a binary classifier; the output gives us either negative output or positive output. One of the key features of SVMs is their ability to maximize the classification margin, which refers to the distance between the decision boundary (hyper plane) and the nearest data points from different classes. This helps in achieving better generalization and robustness of the model. SVMs work to maximize the classification margin: Choosing the Hyper plane: In a binary classification problem, the SVM aims to find a hyper plane that best separates the two classes. Maximizing Margin: The goal is to find the hyper plane that maximizes the margin. This margin is defined as the distance between the two parallel hyper planes that pass through the support vectors of each class. Soft Margin: In practice, perfect separation of classes might not be possible, especially when dealing with noisy or overlapping data.By maximizing the classification margin, SVMs aim to find a robust decision boundary that generalizes well to unseen data.
Equation 1 Supoort Vector Machine

$$f(x) = w^T * x + b \quad (1)$$

Where:
1) f(x) is the decision function that classifies a new input sample x into one of two classes. 2) w is the weight vector perpendicular to the hyper plane that separates the classes. 3) b is the bias term that shifts the hyper plane away from the origin. 4) x is the input feature vector.

4.2 Ad Boosts Classifier

It is an abbreviated variant of the Adaptive Boosting algorithm, which improves performance by leveraging a machine learning, meta-algorithm. The key aim of the Adaptive Boosting Algorithm is to decrease the error rate of week classify using a new classify which takes an input of SVM output and collective with the Adaboost classifier and takes the new and final output of the classifier we are getting.In the AdaBoost (Adaptive Boosting) algorithm, weak classifiers are built and merged into an ensemble by iteratively training them on modified versions of the training data. AdaBoost assigns weights to each training sample, and these weights are updated at each iteration to focus on the samples that were misclassified by the previous weak classifiers. This output is very noise sensitivity the AdaBoost algorithm is shown as follows,
Equation 2 Adboost Classifier Equation

$$f(x)=sin\left(\sum_{t=1}^{T}-\alpha h t\,(x)\right) \quad (2)$$

Here α is the error rate and ht is the week classify.

4.3 Artificial Neural Network

ANNs are a simulation of biological NNs with potential uses in areas such as data categorization and pattern recognition. Depending on characteristics of the input pattern, ANN's decision-making is crucial as well as excellent for the classification of biological data. In ANN, the backpropagation approach is utilised to reduce errors as much as possible. The ECG database is sorted into normal and pathological signals using the ANN classifier. ANNs can be applied to heart disease prediction, along with information about their computation, classification power, and considerations. The goal is to train the ANN to predict whether a patient has heart disease based on these features. Training an ANN involves iteratively adjusting the weights of the network's connections to minimize the difference between predicted outputs and actual outcomes. ANNs have a high classification power, making them suitable for complex tasks like heart disease prediction. ANNs can capture intricate relationships between input features and the likelihood of heart disease, even when those relationships are non-linear and not easily expressed by traditional algorithms.
Equation 3 Artificial Neural Network

$$z = (w1 * x1) + (w2 * x2) + \ldots + (wn * xn) + b \tag{3}$$

Where:
1) z is the weighted sum of inputs and biases. 2) w1, w2,wn are the weights associated with each input x1, x2,xn. 3) x1, x2, ..., xn are the input values.4)b is the bias term.

4.4 Convolutional Neural Network (CNN)

The feature learner, or Convolutional Neural Network, has the capacity to automatically extract high-quality features from raw data. The classification of ECG beats occurs in three distinct phases. Beat detection in an electrocardiogram (a), sample extraction (b), and classification (c) Beat detection is a method used to identify the heartbeat from an electrocardiogram (ECG) recording [30]. Two components make up convolutional neural networks. The first portion, a feature extractor, is responsible for gleaning features from raw data, while the second, a fully connected multi-layer perceptron (MLP), is responsible for doing classification using the feature extractor's training features.

Let xio= [x1, x2.... xn] as the data input vector for beat samples, whereas the number of samples per beat is signifies by n. The convolution layer's output is as follows:

Equation 4 Convolutional Neural Network

$$Output(\mathrm{i,j,k}) = \sum\nolimits_{m,n,c} \mathrm{Input(i+m, j+n, c)} * \mathrm{Filter(m, n, c, k)} \tag{4}$$

Here, Output (i, j, k) signifies the value at position (i, j) of the kth feature map in the output. Input (i+m, j+n, c) signifies the value at position (i+m, j+n) of the cth input feature map, and Filter (m, n, c, k) represents the weights of the filter for the m^{th} row, n^{th} column, c^{th} input channel, and k^{th} output channel.

4.5 Stacked Denoising Auto-Encoder

Autoencoders (AEs) are neural networks that have been trained to exactly replicate the input x, and then output the encoded version of the data. As input, it is often down sampled to a linear feature, and as output, it is up sampled to restore the original dimensions. Stacked denoising AE (SDAE) is an architecture with the goal of re-creating the original, uncorrupted input from one that has been purposefully corrupted [31]. A basic generator is a function that takes an input and outputs a result V ∈RdVto a hidden representation h ∈Rdh, which can be stated as
Equation 5 Auto Encoder equation

$$h = f(W_e x + b_e) \tag{5}$$

4.6 Recurrent Neural Networks (RNNs)

Unlike other topologies, the internal state of these networks of feedback loops is used to process incoming data. As the gradients of vanilla RNNs decrease with time, alternatives like as LSTM (Long-Short Term Memory) [32] are suggested as a means of long-term data storage.
Equation 6 Recurrent Neural Network

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \tag{6}$$

4.7 K-Nearest Neighbour (KNN)

K-Nearest Neighbours (K-NN) is a simple and intuitive machine learning algorithm used for both classification and regression tasks. It's a non-parametric and instance-based algorithm, which means it doesn't make strong assumptions about the underlying data distribution and uses the actual training data for making predictions. The basic idea behind K-NN is to classify a new data point (or predict its value) based on the majority class (or average value) of its K nearest neighbours in the feature space. The "nearest" neighbours are determined by a distance metric, commonly the Euclidean distance, although other distance measures can also be used.

Here's how the K-NN algorithm works:

- **Training Phase:** Store the training data points and their corresponding labels (for classification) or values (for regression).
- **Prediction Phase:** Given a new input data point, calculate the distance between the new point and all the training data points. Select the K data points with the shortest distances to the new point.
- **Classification:** For classification tasks, assign the class label that appears most frequently among the K neighbours as the predicted class for the new point.
- **Regression:** For regression tasks, calculate the average (or weighted average) of the target values of the K neighbours and use that as the predicted value for the new point.

Key Parameters:K: The number of neighbours to consider. It's a hyper parameter that you need to specify before training the model. Choosing the right K value depends on the dataset and problem domain. Smaller K values can lead to noisy predictions, while larger K values can result in overly smoothed predictions. There are several ways to determine this distance, the most popular being the Euclidian, Manhattan (for continuous), and Hamming distances (for categorical).

Among the entire algorithm we find the KNN accuracy from the table 2 it's seen that it give more accuracy results as compare to the other algorithm.

5. Performance Metrics

Several finding and criteria will include like sensitivity, precision, F1-Score and accuracy all comparison from a matrix will be selected to calculated performance. The Matrix includes true negative (TN), false negative (FN), false positive (FP), and true positive (TP). The overall model performance formula is as shown below.
Equation 7 Accuracy of Performance

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} * 100 \tag{7}$$

Sensitivity: It is define in ratio of heart patients to the total number of heart patients
Equation 8 Sensitivity of Performance

$$\text{Sensivity} = \frac{\text{TP}}{\text{TP+FN}} * 100 \tag{8}$$

Precision: It is define as ratio of the true positive score to that of predicted positive score
Equation 9 Precision of Performance

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} * 100 \tag{9}$$

F1-Score: It is define as the ratio of the true positive score to that of the predictive score
Equation 10 F1-Score

$$\text{F1} = \frac{2*(\text{Precision}*\text{Recall})}{\text{precision+Recal}} \tag{10}$$

6. Challenges of ECG Data

There are several issues and challenges in ECG classification that researchers and practitioners encounter. Here are some of the key issues in ECG classification:

A. Noise and Artifacts: Signals from electrocardiograms are often distorted by a wide variety of artifacts and noise, the most common of which are baseline drift, power line interference, muscle artifacts, and electrode motion artefacts. These disturbances can significantly affect the accuracy of ECG classification algorithms by introducing false patterns or obscuring important features.

B. Imbalanced Data: ECG datasets commonly suffer from class imbalance, where the number of instances in different classes is not evenly distributed. The model's accuracy may be greater for the dominant class and worse for the minority class if the data is imbalanced.. This issue is particularly relevant in detecting rare arrhythmias or abnormal cardiac conditions.

C. Inter- and Intra-Patient Variability, ECG signals can vary significantly among different individuals (inter-patient variability) and even within the same individual over time (intra-patient variability). This variability poses challenges for developing accurate and generalizable ECG classification models that can handle variations in signal morphology and dynamics.

D. Over fitting, when a machine learning model over-fits its training data, it becomes highly specific and cannot accurately predict new data. In the context of ECG classification, overfitting can lead to poor performance on new ECG signals, limiting the applicability and reliability of the model in real-world scenarios.

E. Limited Dataset Size Collecting large-scale annotated ECG datasets with diverse cardiac conditions is a challenging task. Limited dataset size can hinder the development of robust and generalizable ECG classification models, as the models may not capture the full range of variability and complexity present in real-world ECG data.

F. Computational Complexity: ECG classification often involves processing long sequences of time-series data, which can be computationally intensive. Developing efficient algorithms that can handle real-time ECG analysis and operate on resource-constrained devices, such as wearable devices or implantable cardiac monitors, is a significant challenge.

7. Conclusion and Future Work

In this survey paper, we are presenting the detail information about the existing heart disease prediction systems. As a part of the survey paper, we have consider dataset from MIT-arrhythmia and Cleveland and Physionet dataset for understanding the machine learning and deep learning algorithms, we have also study previous heart disease prediction models which is been carried out by the different scholar and find out their result in term of accuracy, This research consider performance metrics etc. Former researcher scholar used disease prediction models by exploring the advantages of the methods Here in this survey we have examine the different method like Logistic Regression, KNN, SVM, Radaom Forest, Navie Bayes methods along with this methods we also like to exploring the methods which is used for classification methods like Ad boosts, Artificial Neural Network, Convolution Neural Network, Auto Encoder, Recurrent Neural Network. All this methods is used by the former authors for finding the accuracy and performance also. The current heart disease prediction system are facing the challenges are mention as ECG Challenges. The Main Objective of this survey study is to represent the current scenario of heart disease prediction system and their methods. In the future it required such kind of the machine or system which predicated with high accuracy and performance about the heart disease based on the dataset inputted to them on the algorithm applies. For this purpose we required the trained dataset along the pre process and classification process and feature extraction methods. We have study here different methods like KNN and Radaom Forest and Navie Bayes technology should be more robust and effective. While we predicating the heart disease prediction to the patients we required proper attributes which have impacts on the results, so for that attributes which have less impact on the attribute which can save the time and process of the prediction and give high accuracy.

References

[1] Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans. Biomed. Eng. 2016, 63, 664–675.

[2] Minami, K.; Nakajima, H.; Toyoshima, T. Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. IEEE Trans. Biomed. Eng. 1999, 46, 179–185.

[3] Inan, O.T.; Giovangrandi, L.; Kovacs, G.T. Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. IEEE Trans. Biomed. Eng. 2006, 53, 2507–2515.

[4] Acharya, U.R.; Sudarshan, V.K.; Koh, J.E.; Martis, R.J.; Tan, J.H.; Oh, S.L.; Muhammad, A.; Hagiwara, Y.; Mookiah,M.R.K.; Chua, K.P.; et al. Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals. Biomed. Signal Process. Control 2017, 31, 31–43.

[5] Melgani, F.; Bazi, Y. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. IEEE Trans. Inf.Technol. Biomed. 2008, 12, 667–677

[6] Luca Brunese, Fabio Mart inelli, Francesco Mercaldo, and AntonellaSantone. 2020. Deep learning for heart disease detect ion through cardiac sounds. Procedia Computer Science 176, 2202-2211. https://doi.org/10.1016/j.procs.2020.09.257

[7] Purushottama. C, KanakSaxenab, Richa Sharma (2016), “ Efficient Heart Disease Predict ion System”, Elsevier, Procedia Computer Science, No. 85, pp. 962 – 969.

[8] Chaithra N and Madhu B (2018), “Classification Models on Cardiovascular Disease Predict ion using Dat a Mining Techniques”, Journal of Cardiovascular Diseases & Diagnosis, Vol. 6, pp.1-4.

[9] Kipp W. Johnson, BS, Jessica Torres Soto, MS, Benjamin S. Glicksberg (2018), “ Artificial Intelligence in Cardiology”, Elsevier, Journal Of The American College Of Cardiology, Vol. 71, No. 23, pp. 2668 - 2679.

[10] SumitSharma, and Mahesh Parmar. 2020. Heart Diseases Prediction using Deep Learning Neural Network Model. International Journal of Innovative Technology and Exploring Engineering, 9(3). https://doi.org/10.35940/ijitee.C9009.019320

[11] SarangamKodat i&Dr. R. Vivekanandam (2018), “ Analysis of Heart Disease using in Data Mining Tools Orange and WEKA”, Global Journal of Computer Science and Technology, Vol. 18.

[12] Gudadhe, M., Wankhade, K., Dongre, S., “Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network,” International Conferenceon Computer and Communication Technology, 741–745, 2010, https://doi.org/10.1109/ICCCT.2010.5640377

[13] Kahramanli, K., Allahverdi, N., “Design of a Hybrid System for the Diabetes and Heart Diseases,” Expert Systems with Applications, 35, 82–89, 2008, https://doi.org/10.1016/j.eswa.2007.06.004

[14] Palaniappan, S., Awang, R., “Intelligent Heart Disease Prediction System Using Data Mining Techniques,” International Journal of Computer Science and Network Security, 8, 8, 290–294, 2008 https://doi.org/10.1109/AICCSA.2008.4493524

[15] AnandhavalliMuniasamy, SehrishTabassam, Mohammad A. Hussain, Habeeba Sultana,VasanthiMuniasamy, and RoheetBhatnagar. 2020. Deep Learning for Predictive Analytics in Healthcare. Advances in Intelligent Systems and Computing, 921, 32–42. https://doi.org/10.1007/978-3-030-14118-9_4

[16] Chaithra N and Madhu B (2018), “Classification Models on Cardiovascular Disease Predict ion using Data Mining Techniques”, Journal of Cardiovascular Diseases & Diagnosis, Vol. 6, pp.1-4.

[17] ChalaBeyene, PoojaKamat (2018), “Survey on P rediction 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

[18] C. Beulah ChristalinLatha, S. CarolinJeeva.: "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques", Informatics in Medicine Unlocked, vol.16, 2019

[19] Juan Jose Beunza, Enrique Puertas, Ester Garcia Ovejero, GemaVillalba, Emilia Condes, GerganaKoleva, CristianHurtado, ManuelF. Landecho.: "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)", Journal Of medical Informatics, vol. 97, 2019.

[20] Senthilkumar Mohan, ChandrasegarThirumalai, GautamSrivastava.: "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE, vol. 7, 2019.

[21] Laiqat Ali, AwaisNiamat, Javed Ali Khan, NoorbakhsmamiriGolilarz, XiongXingzhong, Adeeb Noor, RedhwanNour, SyedAhmad Chan Bukhari.: "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of HeartFailure", IEEE, vol. 7, 2019.

[22] H.Jayasree, D. S. S. K. R. T. Naren, K. SaiSankeerth, T. Kumidini.: "HEART DISEASE PREDICTION SYSTEM", www.researchgate.net/publication/334612815, July 2019.

[23] SanchayitaDhar, Krishna Roy, TanureeDey, Pritha, Datta, AnkurBiswas.: "A Hybrid Machine Learning Approach for Prediction of Heart Diseases", IEEE, 2018

[24] Saramaki, T., Neuvo, Y., and Mitra, S. K., Design of computationally efficient interpolated FIR filters. IEEE Transactions on Circuits and Systems 5(1):70–88, 1988.

[25] Y. Bengio et al., “Scaling learning algorithms towards AI,” Large-Scale Kernel Mach., vol. 34, no. 5, pp. 1–41, 2007

[26] Luo, K.; Li, J.; Wang, Z.; Cuschieri, A. Patient-specific deep architectural model for ECG classification. J. Healthc. Eng.2017, 2017, 4108720.

[27] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli,“Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals,” Computers in Biology and Medicine, vol. 100, pp. 270–278, 2018.

[28] R. Acharya, A. Kumar, P. S. Bhat, et al, “Classification of cardiac abnormalities using heart rate signals”, Medical and Biological Eng. And Computing, vol. 42, no. 3, pp.288-293, 2004.

[29] B. Mohamed, A. Issam, A. Mohamed, et al, “ECG image classification in real-time based on the harr-like feature and artificial neural networks”, In procedia Computer Science, vol.73, pp.32-39, 2015

[30] R. Acharya, A. Kumar, P. S. Bhat, et al, “Classification of cardiacabnormalities using heart rate signals”, Medical and Biological Eng. AndComputing, vol. 42, no. 3, pp.288-293, 2004.

[31] Saramaki, T., Neuvo, Y., and Mitra, S. K., Design of computationally efficient interpolated FIR filters. IEEE Transactions on Circuits and Systems 5(1):70–88, 1988

[32] B. Mohamed, A. Issam, A. Mohamed, et al, “ECG image classification in real-time based on the harr-like feature and artificial neural networks”, In procedia Computer Science, vol.73, pp.32-39, 2015.

[33] S. C. Lee, “Using a translation-invariant neural network to diagnose a heart arrhythmia,” in IEEE Proc. Conf. Neural Information Processing Systems, Nov. 1989.

[34] R. Kavitha, T. Christopher, “A Study on ECG Signal Classification Techniques”, Intl. Journal of Computer Applications, vol. 86, no. 14, pp. 9-14, 2014.

[35] Jaymin P atel, Prof.TejalUpadhyay, Dr. Samir P at el (2015), “ Heart Disease Predict ion Using Machine learning and Data Mining Technique”, IJCSC, Vol.7, pp.129-137.