STUDY OF HEART DISEASE PREDICTION USING CNN ALGORITHM

By
DURGESH KUMARI
(1724031501)

Under the Supervision of
Mr. GYANENDRA SINGH
(Associate Prof.)
SUNDER DEEP ENGINEERING COLLEGE (240)

ABSTRACT

The medical field is expanding at a faster rate, with new diseases emerging on a daily basis, necessitating the development of a proper course of treatment. The heart is a muscular organ the length of a clenched human that is in charge of blood circulation. Though the title "heart/cardiac disease" introduce to disorder that affect the heart overall and several diseases that fall under this umbrella, such as Coronary Artery Diseases (CAD), cardiomyopathy, Cardio Vascular Disease (CVD), and others that are caused by the circulation of blood throughout the body. The heart disease data prediction has been done by analyzing medical data with clinical expertise to assist clinicians in the detection of heart disease. By improving these predicting structures, the provision of health diagnostic decisions for heart disease could be improved.

Heart-related illness and CVDs, which have arises as the main life-threatening disorder not only in India as well as globally heart disease is the main justification for a large number of deaths over the last few decades. So, in order to identify such disorders in time for actual remedy, a dependable, precise, or feasible procedure is necessary. ML and deep learning methodologies & approaches have been implemented to huge quantity of information in the field of medical for data processing. Researchers use a variety of deep learning as well as machine learning techniques to analyze large data sets and aid in the accurate prediction of heart diseases. In this report, suggested a Random Forest with Bi-LSTM as a hybrid approach has been found to have more accurate as compared to other algorithms.

CHAPTER-1

INTRODUCTION

1.1 General

The medical field is expanding at a rapid pace as new diseases emerge on a regular basis, necessitating the development of appropriate treatment options. For appropriate diagnosis and treatment, an accurate and efficient mode of operation is required. If the scheme is automated, it could be extremely useful. However, the problem is that, medical practitioners are not effective enough in each sub-specialty, resulting in a scarcity of resource people. As a result, implementing an effective automatic medical diagnostic scheme could be highly beneficial to all stratification involved in this process.

The heart is a muscular organ that is in charge of blood circulation. The human heart functions as a compressor, regulating blood flow in the circulatory system. The heart pumps de-oxygenated blood from other areas of the body into veins, as well as oxygenated blood is pumped back from the lungs to different areas via arteries[1]. Lungs help in the process of oxygenation. The Sinus node is an electric impulse system found in the heart that organizes the frequency of pumping. It acts as a natural pacemaker but is located at the top of the right atrium. The control the contraction and relaxation of the atrium and ventricles are actuated via messages sent over the heart muscle tissues defined [2] .

As per to WHO statistics, heart diseases claim the lives of 12,000,000 people worldwide every year. CVD report for nearly $\frac{1}{2}$ of deaths in the US or another advanced nations, and the same could be said for creating countries such as India. As a result, cardiac disorder are the leading source of death in adults worldwide.

As per [3], a recent cardiac health-related survey, nearly 1.2 billion population die every year as a outcome of heart diseases. There is no single solution to the increasing load of HD. There is no single solution to the rising burden of heart disease, given the massive transformation in ethnic, as well as economic environments. Heart failure prognosis has historically been an extremely thought-provoking task in the eve of high-cost ratios. The price of a wide range of modern imaging and clinical methodologies for the diagnosis of heart disease is prohibitively high. Leading causes of cardiac disease involve chest discomfort, dyspnoea, fatigue, edema, palpations, as well as syncope, as well as cough, hemoptysis, and cyanosis. Figure 1.1 depicts the 2017 heart disease death rate framework.

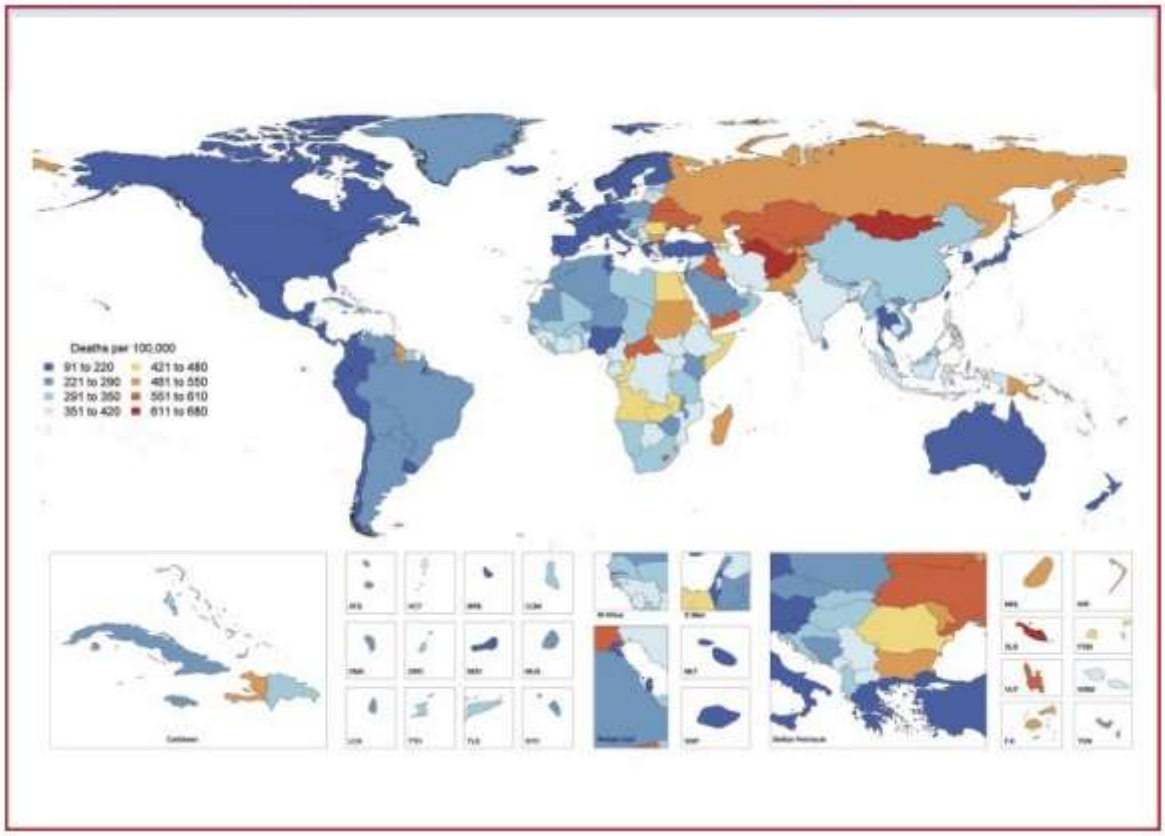

Fig. 1.1 Global Map Standardized Death Rate of CVD in 2017[3]

Though heart/cardiac disease is the common term for diseases related to the heart, there are several diseases that fall under this umbrella, such as Coronary Artery Diseases (CAD), cardiomyopathy, Cardio Vascular Disease (CVD), and so on, based on blood circulation throughout the body[4].

- **Coronary Artery Disease:** Coronary artery disease is a form of discomfort caused by a decrease in blood circulation. The reduction of artery supply will harm the vein and cause discomfort to the heart's normal systolic and diastolic function. CVD is a leading cause of serious illness, disability, as well as death. High blood pressure, CAD, and rheumatic fever/rheumatic heart disease are all risk characteristics for CVD.

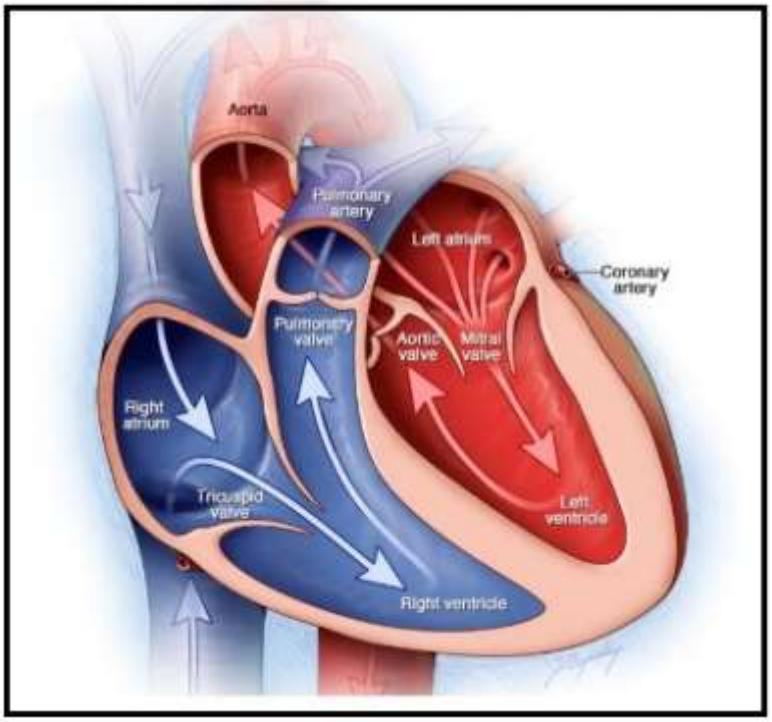

Figure 1.2 : Anatomic Structure of Heart[4]

Reduced blood supply to the heart is caused by the narrowing of coronary arteries, which leads to Coronary Heart Disease (CHD), which in turn leads to MI, also called as a HA as well as CP. This could also be caused by obstructing the artery with plaques or fat deposits, causing formation of blood clots. As a consequence, the blood supply to the heart muscle is insufficient, resulting in severe chest pain. The different kinds of heart disease are depicted in Figure 1.3 based on clinical scenarios.

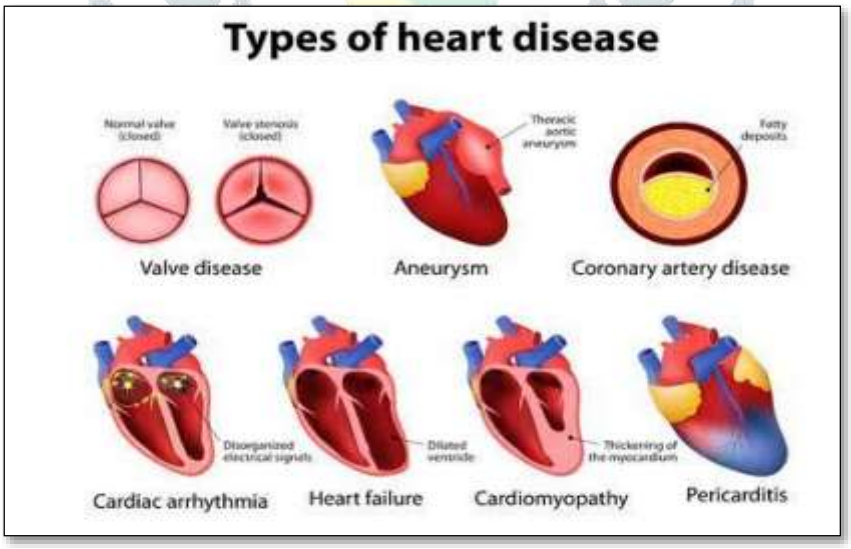

Figure 1.3 Types of Cardiac Disease[5]

- **Acute myocardial infarction**

It is the medical name for a cardiac arrest. A cardiac arrest is a situation in which fatty substances in the blood reduces the profitability of flow, causing tissue damage on the arteries. The blocked arteries may be unable to supply the body with oxygenated blood which will causing inducing other organs to malfunction. Figure 1.64 explains a type of heart arrest lead by intense pressure.

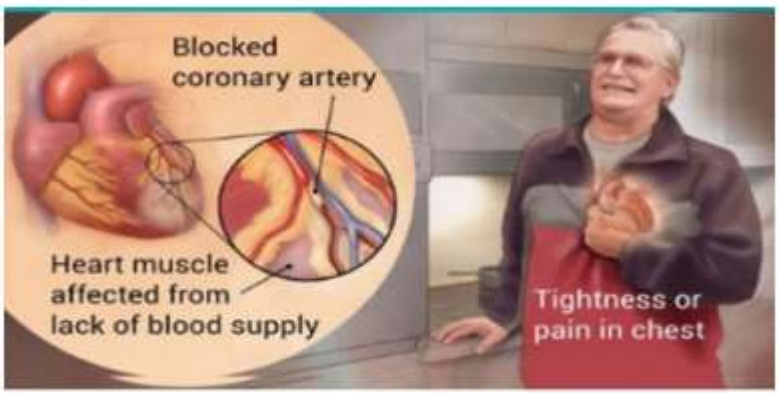

Fig. 1.4 Acute Myocardial Infarction[6]

- **Chest Pain (Angina)**

Angina is the medical term for chest pain. It is common for patients to require emergency medical attention. Patients must be addressed promptly with ventilators If humans experience this form of discomfort. The lack of blood flow causes pressure on the blood walls as well as affects the blood vessels[7]. This causes pressure on the blood vessals, resulting in chest pain. In peritoriam, the disease causes stable angina. Inconsistent blood flow among the peritoris walls. The primary causes of unstable angina are lifestyle changes and behavioral habits. Figure 1.5 depicts the basic case of unstable angina triggered by a coronary artery.

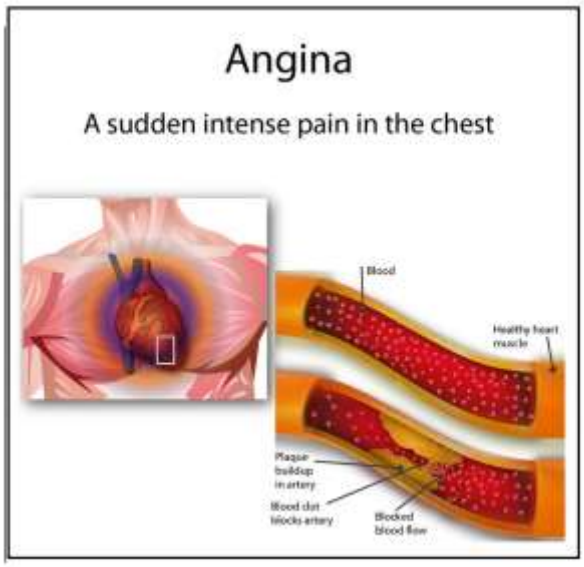

Fig. 1.5 Angina[7]

Adults in established countries like the US , UK, Canada, as well as Australia are more likely to develop heart disease. Cardiac diseases are refer as one or more heart problems. Definite kinds of heart disease have been proposed [8]:

- Randomly heartbeat (arrhythmias)
- congenital heart defect
- weak heart muscles (cardiomyopathy)

- heart valve issues
- heart infections
- CVD

Cardiac failure occurs when the amount of blood to the heart is inadequate to satisfy the body's needs. It does not happen overnight, but rather worsens gradually over time. The following are the reasons of heart failure:

- Cardiomyopathies
- Coronary Artery Disease
- Diabetes
- Heart valve defects
- Heart Defects present at Birth
- Hypertension
- Lung conditions such as Emphysema
- Prior Heart Attacks

1.1.1 Risks Factors of Heart Disease

The following subsections describe the factors that contribute to cardiac disease. Various factors are exacerbating the risk of acquiring cardiac disease[9]. The manageable endanger elements of cardiac discomfort are depicted in Figure 1.9. Age, gender, previous case reports of patients, cholesterol levels, smoking, and diabetes are among the elements.

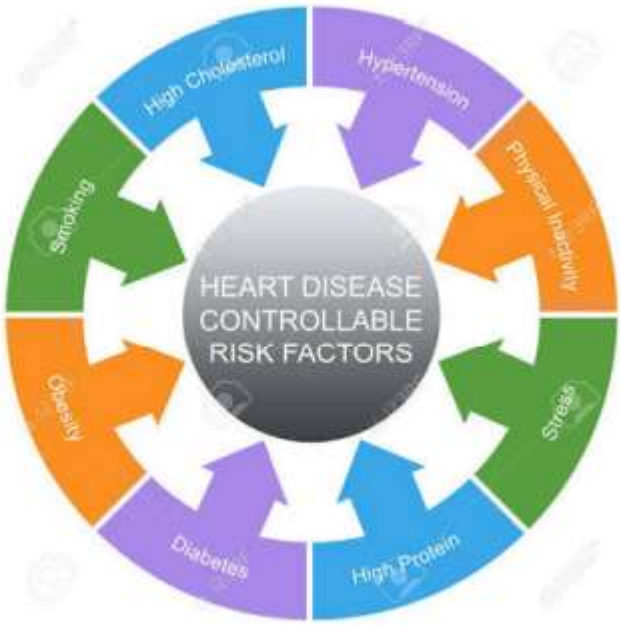

Figure 1.6: Risk Factors of Heart Disease[9]

- **Smoking:** Smoking is a significant risk factor for cardiac arrest because it creates patient blood as well as enzymes to coagulate with one another, increasing the risk of heart failure. Smoking causes artery thinning leads to the effect of atheroma on the artery walls, allowing damage to the heart muscle and a reduction in blood demand to the heart[10]. Figure 1.7 depicts the significance of a risk factor premised on a behavioral aspect of smoking.

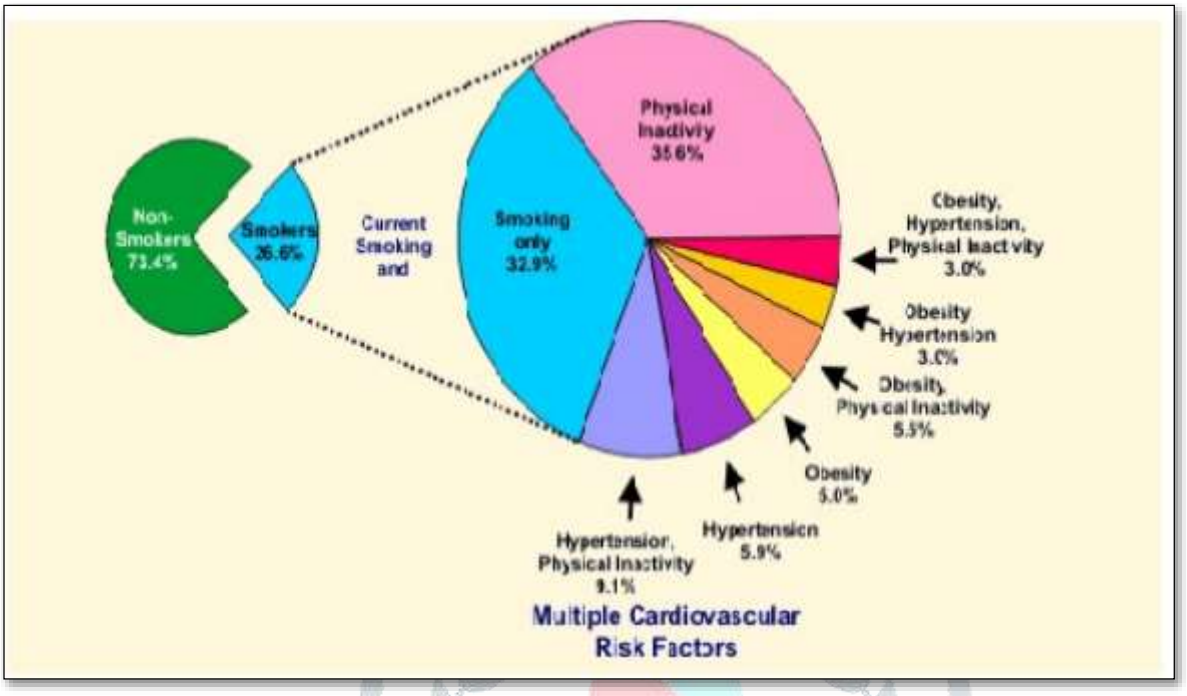

Fig. 1.7 Risk factor based of behavioral aspect-smoking[10]

- **Cholesterol:** An rise in blood cholesterol levels, particularly Low Density Lipoproteins, is another essential risk factor for heart disease (LDL). This leads to an increase in the accumulation of fat deposits in the blood vessels, a condition known as atherosclerosis. Higher cholesterol levels in the blood rise the risk of heart disease Atherosclerosis is caused by fatty material with high cholesterol levels. Lipid characteristics are classified into two types. There are two types of lipid profiles: low level lipid profiles and high level lipid profiles.

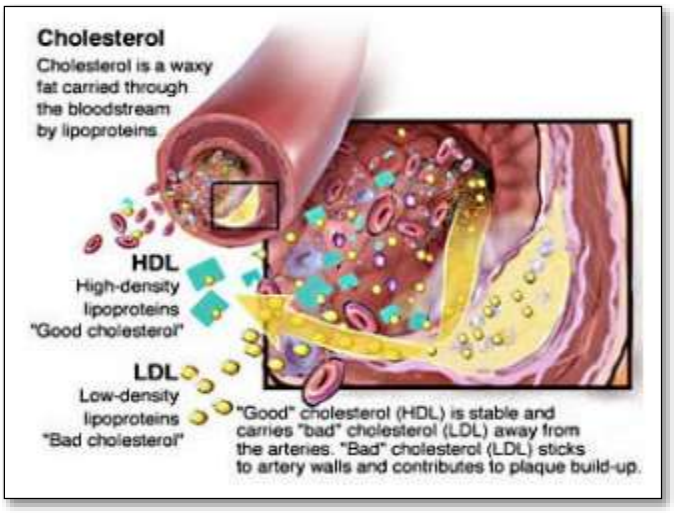

Fig. 1.8 Cholesterol Levels in Blood

- **Obesity:** This may result in obesity and also increase cholestrol stage , which is one of the common complications.

- **HBP :** Figure 1.9 depicts the chance of heart disease predicated on surveillance behavioral factors. In [11] describes how irregular blood flow increases the risk of cardiac disease.

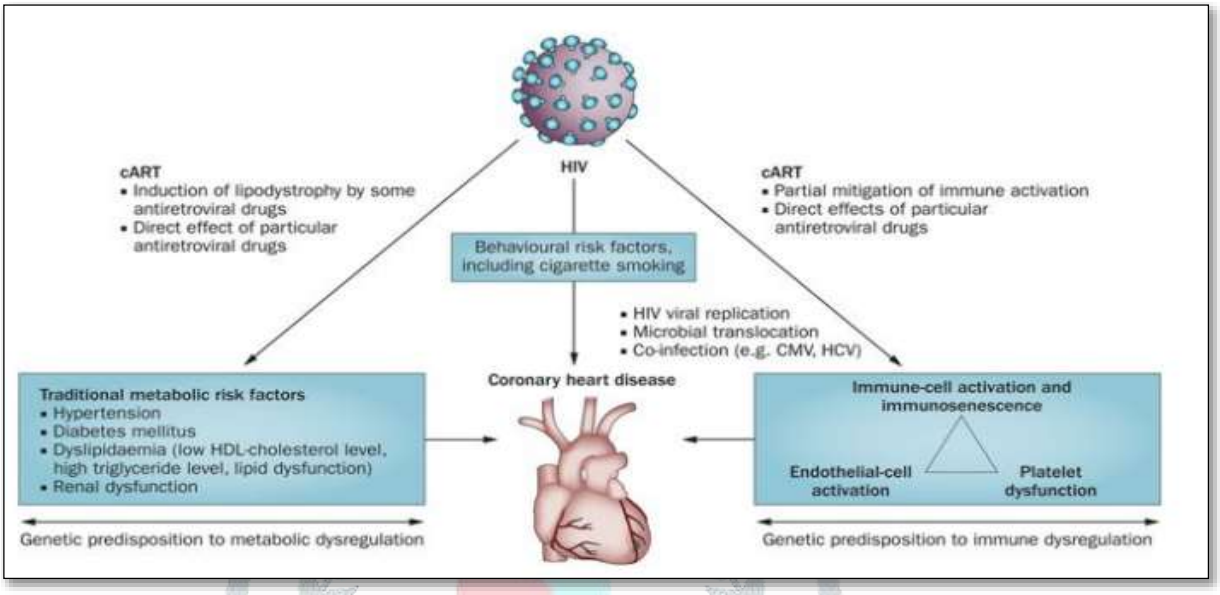

Fig. 1.9 Behavioral risk factors of surveillance[11]

Early indications of Heart Disease:

- Dizziness or fainting spells are early signs of heart disease.
- Discomfort following a meal, especially if it is prolonged.
- Breathing difficulties, even with light exertion.
- Fatigue of unknown origin.
- As a prevalent symbol of coronary insufficient Pain behind chest bone that radiates to the shoulders or a feeling of numbness and intense pain in the middle of the chest has been noted [12].
- Heart palpitation.

1.2 Heart Disease Prediction

The Heart Disease Forecasting has been created by analyzing medical data with clinical expertise. The brilliance of medical diagnostic choices for heart diseases could be improved by improving these forecasting structures. Many hospitals keep proper medical records and also have an effective hospital information system. The hospital information system must be correctly utilized or obtained so that it can assist physicians in making proper diagnoses and clinical decisions [13].

Delays in diagnosis could even lead to delays in diagnosis, which can have significant effects in these deadly diseases. In other cases, too several tests are performed, which may result in negative outcomes, resulting in a waste of both time as well as money. These are induced by the lack of experience or expertise of the doctor.

Health systems offer a variety of services to improve quality and efficiency, require impacted people and their families, improve care as well as coordination or public health, and protect patients', safety and confidentiality of health information of patients.

Heart failure is a major as well as deadly health issue today, affecting primarily older patients as a result of changes in lifestyle and the use of nonsteroidal anti-inflammatory drugs as well as finally leading to death. Cardiovascular Diseases are the major type of heart disease. This community of illnesses could be anticipated based on symptoms [14]. A several ML and DL learning models are used to forecast heart diseases. Figure1.10 depicts the relationship among methods that are thought to be efficient as well as are currently state-of-the-art for sensing, forecasting, segmenting, categorizing, as well as acknowledging objects in images and videos.

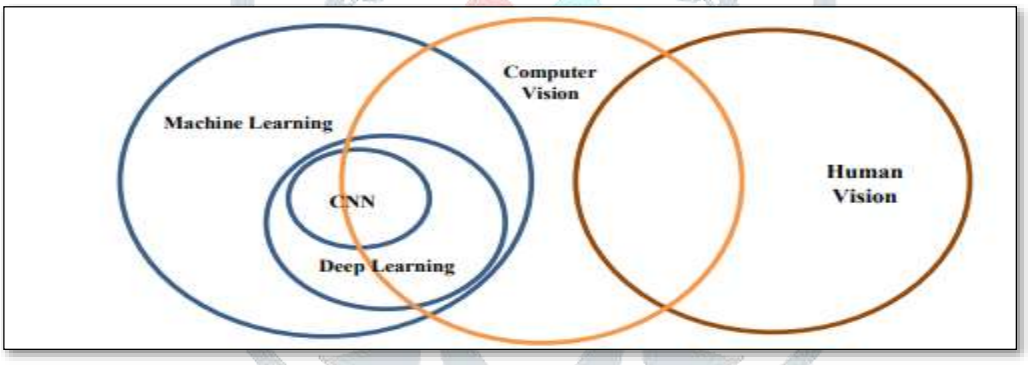

Figure 1.10: The relation between machine learning, deep learning, CNN, human vision and computer vision[15]

1.3 Machine Learning

Establishing medical data apps is extremely diverse, and at a greater level, one can evaluate patient sensitive data like temperature, glucose, blood pressure, and sugar level for previous disease identification and prevention. A single attribute, like blood pressure or blood sugar level, could be analyzed at the low level to offer appropriate medications. Different approaches are used in all elements to derive significant knowledge from existing real-world unprocessed data. In this study, developers describe a hybrid approach traditional statistical methods with model-based machine learning methods to effectively categorize heart disease datasets with smaller feature sets and higher accuracy indicators.

Machine learning methods are critical in data analysis as well as prediction[16]. Traditional methods fail to deal with medical data extraction because it is enormous as well as complicated. The primary motive of the ML case research is to form a recommendation system for heart disease diagnosis or present novel & high opportunity methods for early treatment of heart disease diagnosis. Developers focus on heart disease databases in specific as well as look at apps in chronic disease management. In everyday life, machine learning is used in critical decision-making structures like medical diagnosis and clinical decision-making.

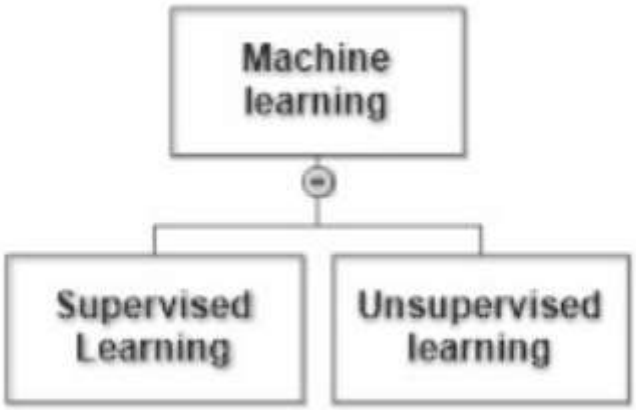

Figure 1.11 Classification on Machine Learning Techniques[16]

It is the best choice for solving issues involving a huge amount of data and numerous features. These methods construct classification models from the input dataset as well as perform classification. It employs supervised, semisupervised, as well as unsupervised learning methods. In the case of supervised designs, classification designs are constructed with known input and output data. Unsupervised learning, on the other hand, contracts with hidden patterns in data. The classification techniques are depicted in Figure 1. 11. There are many different kinds of machine learning methodologies, but they are widely divided into four categories based on their objective.

- **Supervised learning:** In this learning, a feature is implied from labeled training data that defines an incoming information to an output based on the function learned from a set of labeled training instances [18].
- **Unsupervised learning:** Due to the unlabeled dataset, prior training is not provided in this aspect of teaching. The machine is limited to grouping unsorted information by itself based on similarities, differences, as well as hidden patterns in the data [18].
- **Semi-supervised learning:** It drops somewhere among supervised as well as unsupervised learning because the input information in semi-supervised learning is partly labeled [19-20].
- **Reinforcement learning:** A machine is trained using the trial-and-error technique in this learning. The algorithm learns from its previous experiences until it has explored all possible states and has determined an ideal behavior to achieve optimal performance. It's commonly utilized in robotics, gaming, as well as navigation [21].

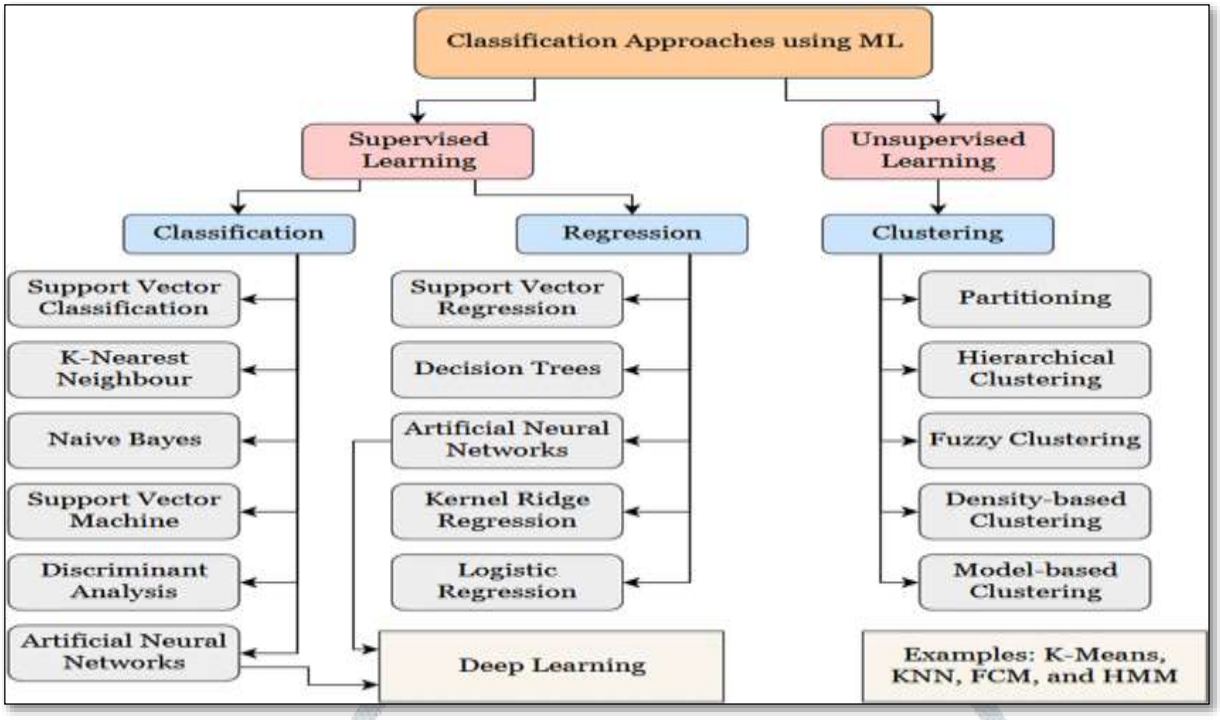

Figure 1.12 Classification of Machine learning techniques[17]

Figure 1.12 defines the numerous types of approaches used in different fields. SVM, LR, decision trees, RF , K-nearest neighbour (K-NN), nave bays (NB), neural networks (NN), and linear discriminate analysis (LDA) are some of the most widely used applications in solving supervised learning issues. SVM, LR, DT , RF , K-NN , nave bays (NB), neural networks (NN), and linear discriminate analysis (LDA) are some of the most widely used applications in solving supervised learning issues.

- **SVM :** The SVM approach could be utilized for both regression as well as classification issues , but it is most commonly used to accomplish classification goals. By deciding a hyper-plane, the data is divided into classes [22]. In way to select the best hyper-plane, the method aims at maximizing the distance among the data points of the classes. Margin maximization is a term used to describe the concept of distance maximization [23]. There are 2 kinds of SVM algorithms: linear SVM as well as non-linear SVM. A hyper-plane is used to classify training data in linear SVM. A non-linear SVM is achieved by implementing a kernel trick to highest-margin hyper-planes, in which each pixel value is substitute via a NL kernel variable.

- **Decision Tree (CART)**: It is a visual representation that uses branching technique to show all of the possible results of a decision based on some circumstances. The root node is the starting point in a decision tree, in which each internal node denotes a test a function, every tree branch reflects a decision rule, & every leaf node indicates an outcome (target variable) [24]. This method gives a graphical representation of a decision problem or aids in communication improvement and decision making under unstable situations. It also aids in decision making by enabling a data scientist to navigate forward & backward calculation routes. In addition, decision trees are resistant to missing values or errors. The

method is best suited for troubles with attribute-value pairs as instances. The decision tree is categorized into two categories:

The decision tree is categorized into two categories:

i) classification Tree, and

ii) Regression Tree.

iii) The algorithm has applications in a variety of fields, including finance, banking, remote sensing, and medical centers [25].

- **Random Forest (RF)** This method creates a slew of decision trees from a random subset of data and is typically equipped using the bagging method [26]. The classifier is constructed numerous times on random samples before making the final estimation by integrating the outcome of all decision trees to attain the optimum prediction [28 This method is resilient to outliers as well as maintains accuracy even missing data is present. It is extremely capable of handling binary, numerical, and categorical features without transformation or alteration. It also estimates the characteristics that are most important for classification. This method has recently been used in banks, the automobile industry, the healthcare industry, speech recognition, as well as the classification of images and texts.

- **Naïve Bays (NB):** The Bayes theorem of probability [28] is the foundation for the Nave Bays classifier, which has a low variance as well as a high bias. This classification method's machine learning techniques are especially helpful for document classification & disease prediction. This simple classifier becomes a great option when a moderate or big training data frame with provisionally independent features is accessible. This application areas of classification methods include sentiment analysis, document categorization, email spam filtering, as well as news article classification, among others.

- **K-NN:** k-NN is a non-parametric technique, which means it makes no assumptions about data distribution [29]. The methodology could be utilized to resolve both classification as well as regression issues. It is a straightforward method in which all obtainable cases are regarded, as well as forecasting on a specific case is conducted using similarity measures. The k-NN method measures similarity using Euclidian, Manhattan, Minkowski, and Hamming distance functions. It should be mentioned that the Hamming distance measure is only appropriate to categorical target variables, whereas the other three are accurate for continuous target variables.

- **Regression Models:** Regression models are the most important component of the predictive analysis process. In overall, regression analysis use statistical methods to calculate the criterion of the distinction among forecasted and expected values. The regression design makes use a variety of data types. It employs mathematical formulas to facilitate the design that depicts the relationship among the various parameters in the given database.

- **Clustering :** In simple terms, it makes it easier for the user to determine the principles of the database. A good clustering technique will always have a high intra-class clustering and a low interclass clustering. The accuracy with which hidden patterns linked with information are clustered. Stream analysis, economic science, and several other fields make extensive use of clustering methods.

1.4 Deep Learning Approach

DL is a new subfield of ML that was developed to resolve the shortcomings of conventional machine learning methods in the automation period. Because the quality of most classification systems based on traditional methods is dependent on the feature extraction stage. It is frequently hard and time-consuming to collect valuable features from information. Furthermore, extensive prior domain skill is needed in the layout of a feature extractor in order to collect efficient knowledge from big data. Deep networks, as opposed to hand-crafted feature selection, obtain sophisticated hierarchies directly from raw data to produce a hierarchical data representation[30].

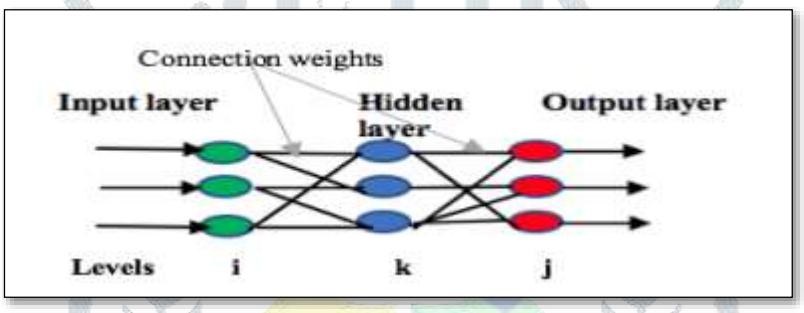

Figure 1.13 : Neural network[31]

The layers among the input and output are known as hidden layers, and they are made up of a variety of linear and non-linear transformations. Deep Learning is an effective learning technique that provides neural networks to complete tasks. Neural Networks are analogs that simulate the behavior of biological neurons in the human brain[32].

Neural networks are fully integrated graphs where every node is associated with an input value and each edge with a weight, both of which are initially random values with a bias that is always set. A neural network works by measuring sum, where w represents the weights and x represents the inputs.

$$\text{Weighted Sum} = \sum w_i x_i + b \qquad \text{(Equation 1.1)}$$

To maximize the output, the weighted sum is used as a special function. These special functions are known as Activation Functions, and they could be utilized to create the output of NL, allowing classification.

ReLu is rectified linear unit which is used to obtain only positive values and zeros.

$$\text{ReLu} = \max(0, x) \qquad \text{(Equation 1. 2)}$$

The sigmoid function is a special S-shaped curve which limits the values between 0 and 1.

Sigmoid = 1/(1+e-x) (Equation 1.3)

The training dataset in a DL technique is a collection of instances that are consumed to a specific classifier to build the machine. Because the alternatives to the input instances are already known, the deep learning classifier obtains the knowledge during the training system in order to organize the data into a specific class as well as provide the desired output as well as training accuracy. The testing dataset is a collection of examples used to put the neural network to the test and see how well it learned to categorize from the testing set. Because the alternatives to these inputs are already recognized, the neural network is evaluated on these instances to see if it is making the desired assumptions while being evaluated on new data that is distinct from the training data.

1.4.1 Significance of DL

It has been used in a range of systems with remarkable results over the last five years. The deep learning algorithm attempts to learn without the use of any supervision. Deep Learning employs both supervised and unsupervised classification. Hidden layers learn features, which are then trained to develop a model. Artificial Neural Networks are reliable as well as generate efficient output by establishing numerous hidden layers from the input given. Furthermore, repeated computations at the hidden layers could be summarized at subordinate layers, resulting in variable reduction[33]. Even so, these are only the most fundamental priors that distinguish deep learning from other methodologies, but there are a plenty of others that remain unexplored.

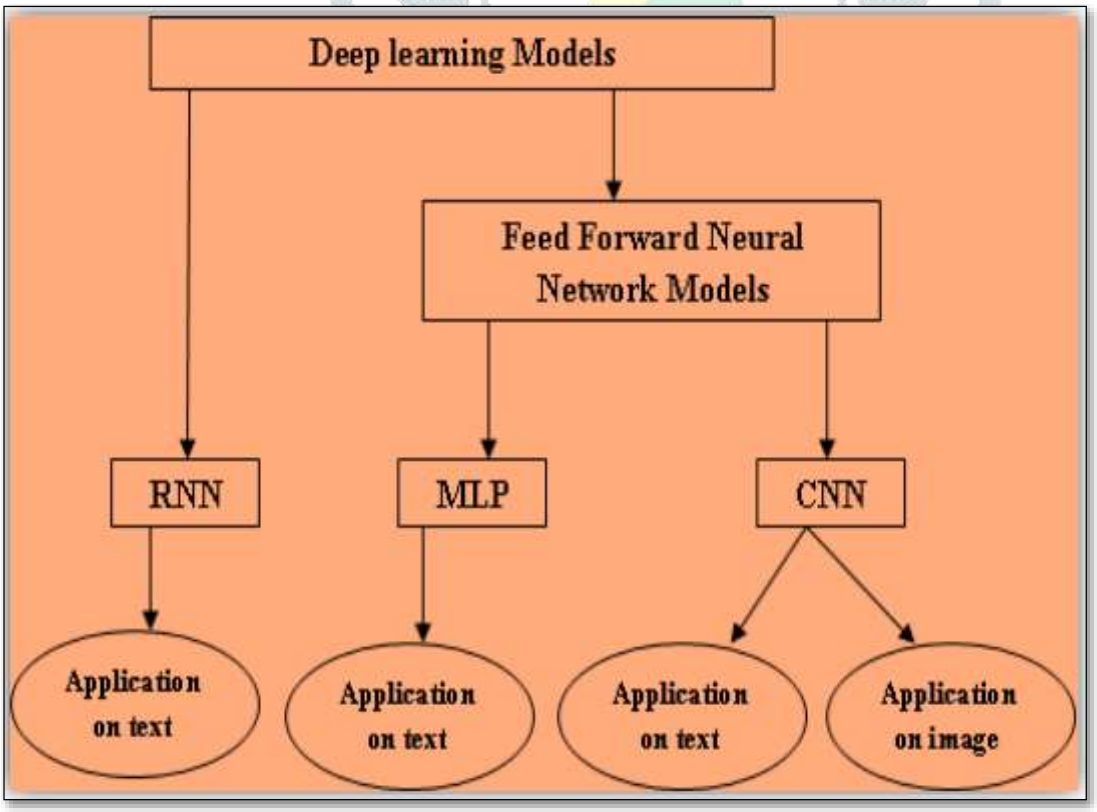

Figure 1.14: Deep learning Models[33]

Deep Learning, a subgroup of Machine Learning, has grown in importance in the development of automatic navigation systems, text processing, speech recognition, as well as a variety of other apps. In aspects of classification and clustering, DL approaches perform better than ML algorithms on large data sets. Figure 1.14 depicts the various deep neural networks utilized to diagnosis a heart disease.

1.4.2 Feed Forward NN

It designs as well as recursive neural network (RNN) designs were used to classify DL techniques for sentiment classification. Feed-forward neural network models were divided into two types: multilayer perceptron models and convolutional neural network models. LSTM, a type of RNN is used in suggested model with some sort of Enhancement.

1.4.3 CNN

CNNs are deep artificial NN that are primarily used to extract features, cluster focused on the similarities, as well as recognize objects. Similar methods are utilized to recognize faces, street signs, tumors, platypuses, as well as other elements of graphic data[32].

1.4.4 RNN

RNNs are a kind of neural network in which the current inputs of hidden layers are determined by the previous outputs of hidden layers. This enables them to interact with a time sequence with temporal relations, like speech recognition as well as text classification, among other things. RNNs are found to be more effective in sentiment analysis unlike CNNs in the literature[34]. Consider the sentence or document Di as an example. The document at hand is a sequence of things. Features are obtained from the words and represented in a high-dimensional vector space.

1.4.5 LSTM

NN typically operate as a black box, making decisions based on inputs. It includes data about learning opportunities in static memory in the form of weights[35]. The LSTM system was created to present explicit representation for memory in RNNs. The memory unit in the system is referred to as a 'cell,' but these designs are RNN adaptations that are suitable for sequential data. Figure 1.15 depicts how the algorithm works.

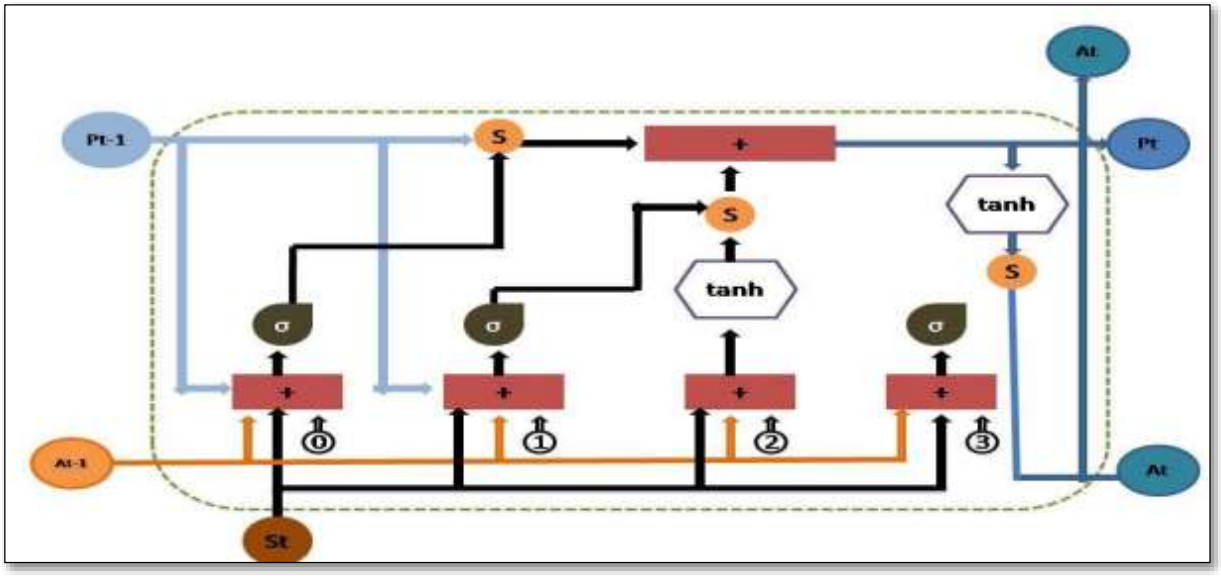

Figure 1.15: Working of LSTM[35]

As per Figure 1.15, the LSTM network accepts three inputs: 'St,' 'At-1,' and 'Pt-1.' 'St' is the current time step's input vector. The preceding LSTM unit's output or hidden state is represented by 'At-1.' And 'Pt-1' is the previous unit's memory component or cell state. It has two outputs, 'At' and 'Pt,' where 'At' is the current unit's output and 'Pt' is the current unit's memory component. Each decision is made after taking into account the current input, previous output, as well as memory data. The memory is modified when the current output is acquired. The 'S' represents the 'Forget' component of multiplication. When the forget component is set to '0,' it neglects ninety percent of old memory.

The unit allows a portion of old memory for all other values like 1, 2, and 3. The plus operator is used for piecewise summation to combine old as well as new memories. The 'S' sign calculates the quantity of old memory. Pt-1 is altered to Pt as a function of multiple operations. The activation functions depicted in figure 5.3 are the sigmoid and tanh activation functions with forget valve output. Because it contains old memory while processing new inputs, the second activation valve is referred to as a new memory component. The amount of memory to be allocated to the next unit is determined by the old memory, previous output, as well as current input, and bias vector. The LSTM is collection of 3 gates: the Input gate, the Forget gate, Output gate. These are indicated by the sigmoid activation function in the range of '0' to '1', where '0' prevents all data from entering and '1' does the inverse. The function is expected to return a positive result that is accurate. The purpose of desertification is to investigate the effectiveness of a Machine and DL approaches for estimation of heart disease utilize a hybrid approach(Bi-LSTM and Random forest).

1.4.6 Bi-directional LSTM

The Bi-LSTM is a deep learning framework that collects information both forward as well as backward to decrease classification error. The Bi-LSTM is made up of a forward as well as backward surface that analyzes information and tokens. LSTM is a form of RNN that can store the most significant data. BiLSTM is made up of two LSTM layers that analyze data serially from the previous as well as future token

contexts[36]. One layer processes data information from left to right, whereas other layer processes data from right to left. A secret forward level comprises a hidden unit function h at every time step t, which is stored predicated on the previous phase $\overrightarrow{h}\ t-1$with I/O data at the current time step $\overrightarrow{h}$. The backward layer performs the similar hidden unit variable predicated on future content $\overleftarrow{h}\ t+1$ with existing information. The forward as well as backward data presentation integrated to long vector.

The dropout method is utilized after the I/O layer in BiLSTM-RF to minimize overfitting of the training information. This method aids in the prevention of complicated co-adaptation on training samples.

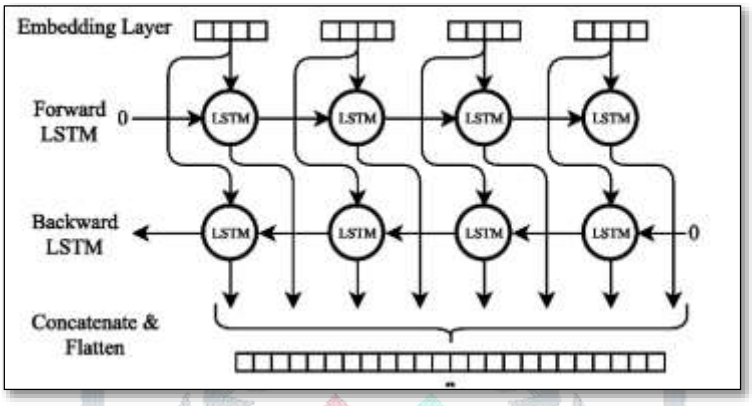

Figure 1.16:Bi-LSTM[37]

1.5 Bi-direction long short term memory –Random Forest

Bi-LSTM and RF are used as a Hybrid Method in the presented design to forecast heart disease. The Bi-LSTM technique, which analyzes data in both forward and backward directions to decrease classification error, as well as random forest lead to the conclusion that more trees could protect the right decision. Typically, the heart disease prediction framework examines a variety of patient factors to anticipate the risk of heart disease. The Bi-LSTM algorithm shows heart disease by analyzing different features in the available dataset. Deep learning methods are utilized in a large scale of research regions to improve prediction as well as classification accuracy. In this study, a new deep learning technique (BiLSTM–RF) for heart disease prediction is suggested. The BiLSTM–RF model was evaluated to a few standard techniques in this study.

As a result, the following are the various benefits of Random Forest and Bi-LSTM:
Bi-LSTM has the benefit of fixing the situation of fixed sequence to sequence prediction[37]. RNN has the restriction of having the same size for both input and output.

A few advantages of Random Forest are[38-39]:

- Random Forest is a bagging methodology that employs the Ensemble Learning approach . It grows as many trees as it can on the sample of features as well as then integrates the results of all

the trees As a result, it decreases the overfitting issue in decision trees, as well as the variance, and thus enhance the efficiency.

- Random Forest could be applied to both classification and regression issues.
- Random Forest is effective with both nominal and ordinal parameters.
- Random Forest could even manage missing values instantly.
- No feature scaling needed: Because Random Forest utilizes a rule-based method rather than distance calculation, no feature scaling (standardization and normalization) is needed.
- Handles non-linear variables effectively: non-linear variables have no influence on the productivity of a Random Forest Unlike curve-based methodologies. As a result, if there is a maximum level of non-linearity among the independent parameters , RF could exceed more curve-based methodologies.
- RF could indeed tackle misplaced values instantly.
- RF is generally resistant to outliers as well as could deal with them instantly.
- The RF approach is highly stable. Even if a fresh information value is added to the database, the complete method is unaffected because the fresh information may affect one tree but is unlikely to affect all trees.
- Disturbance has less impacted on Random Forest.

CHAPTER-2
LITERATURE SURVEY

Huaiyu Wen et al.,(2020) Classifiers based on Machine Learning (ML) have been used to identify HD. To enhance the forecasting efficacy of the machine training designs, the Chi square FS approach was utilized for linked feature selection. For method hyper-parameter tuning as well as best model identification, cross validation or technique were used. Prediction accuracy has also been evaluated using performance metrics like classification precision, specificity, sensitivity, Matthews' correlation coefficient, as well as running time. The suggested approach was performed using the Cleveland HD database. The experimental outcomes showed that the suggested method outperformed state-of-the-art methodologies in means of standard[40].

Rahul et al.,(2021) Utilizing database schemas to categorize the database for disease prediction. The classifying research is able to provide faster as well as more diverse alternatives. Deep learning or Gradient Boosted Trees are two algorithmic trends used to obtain the predicted values of 32.20 and 27.73. Deep Learning outperforms Gradient Boosted Trees, which also emerge in the investigation[41].

Sharma et al.,(2020) has used the Cleveland heart disease database to conduct deep neural network analysis. A few DNN methods have been created through research, every with completely different characteristics as compared to the others. The total goal is to develop a design that can estimate whether a patient will be identified with cardiovascular heart disease or otherwise. To achieve the best results from the study, distinct optimizing algorithms are used to maximize the loss function, various weight initialization methodologies to activate the model's variables, and a distinct set of layers of neurons. The test set includes 147 instances, as well as the DNN achieves an accuracy of 82.7 percent, a misclassification possibility of 17.3 percent, a sensitivity of 81.03 percent, as well as a selectivity of 82.8 percent. AdaGrad is known to be the optimizer in this case, generating 85 percent accuracy, as well as random uniform weight initialization produces fast outcomes in case of epochs. It has also altered the no. of hidden layers in the neural network as well as observed that as the set of hidden layers rises, so does the set of epochs required to train the design[42].

Xiao et al.,(2020) Predicated on an enhanced 3D U-net convolutional neural network DL model for heart coronary artery segmentation for disorder risk forecasting, it is applicable to several information sources with 2 backgrounds without as well as with the centerline. Utilizing a fresh local feature to retrieve ventricular data as well as DNN to automatically remove to revert the biventricular contour coordinates. Integrating features as well as deep belief networks, as well as training regression systems, could not only retrieve high-level data but also precisely split the left as well as right ventricles at a low operating complexity. The segmentation accomplish competitive on the dice coefficient was placed in 2 databases. The findings demonstrate that the centerline preprocessing design training impact outperforms the actual information. The experimental outcomes reveal that the dice coefficient of 0.8291 has the best outcomes. The study findings are still fairly fuzzy as well as smooth, but are not sensitive to captured image. As a result, it is unsuitable for even more complex medical photos[43].

Kefaya et al.,(2020) identifies this critical gap by examining a DNN model that accurate diagnosis arrhythmia relying on a personal raw electrocardiogram (ECG) heartbeat, as well as collating existing methods predicated on Heart Rate Divergence. To obtain 100 percent arrhythmia prediction accuracy, trained and tested the method utilizing ECG database containing 380000 heartbeats. Notably, the model recognizes heartbeat scenes as well as ECG morphological features that are class-discriminative and therefore important for arrhythmia prediction. It Conclude, contribution significantly develops current arrhythmia prediction methodology as well as satisfy the requirements of clinical practitioners by obtaining an effective as well as fully transparent system to assess arrhythmia prediction decisions[44].

Adeen et al.,(2021) The goal of analysis is to compare the productivity of numerous classification approaches in way to identify the most effective approach for estimating whether or not a patient will develop HD. Also compares the Nave Bayes, Help Vector Machine, Random Forest, as well as supervised

learning models to find the most effective approach. Database has been categorized into training & testing data as well as designs have been equipped or using Python the precision has been mentioned. A comparison of the performance of the methodologies is shown below and the table provides their accuracy scores, recall, accuracy, F1 score. Random Forest has been reported to have greater accuracy (95.08 percent), recall (0.91 percent), and F1-score (0.91 percent) than other methodologies[45].

Ankita et al.,(2015) For the prediction of heart disease, author propose an efficient GA hybrid model combined with the BP method. The primary goal is to create a design that could determine as well as obtain uncalled knowledge (patterns & relations) about HD from a previous HD dataset file. It could solve complex questions for identifying HD, assisting medical researchers in making smart clinical decisions that traditional decision support could not. It could help to minimize treatment costs by providing efficient treatments[46].

Touhidul et al.,(2020) To decrease attributes, PCA was utilized. In addition to a Hybrid genetic algorithm (HGA) with k-means for final clustering. The k-means technique is commonly utilized for data clustering. Since this technique is heuristic, it is prone to becoming stuck in local optima. To solve this situation, author used the Hybrid Genetic Algorithm (HGA) for data clustering. The proposed method has a 94.06 percent accuracy in predicting early heart disease[47].

Sayali et al.,(2018) Because of the enormous volume of information expansion in the biomedical or healthcare fields, precise clinical data analysis has become advantageous for earlier disease detection as well as patient care. Even so, when medical data is missing, the precision suffers. To address the issue of lacking medical data, use data cleaning as well as imputation to convert missing values in the accurate data. On the basis of the dataset, authors are continuing to work on predicting heart disease using the Nave Bayes and KNN algorithms. And propose extending this work by predicting disease risk utilizing structured data. Author employ a unimodel disease risk predictive model focused on convolutional neural networks. The CNN-UDRP approach has a forecasting accuracy of more than 65%. Furthermore, this program gives answers to disease-related questions that people face in their daily lives[48].

Kavitha et al.,(2021) To anticipate heart disease, a novel machine learning framework is developed. The research project makes use of the Cleveland heart disease data source, as well as data mining approaches like regression as well as classification. Random Forest & Decision Tree ML approaches are utilized . The machine learning model's novel method is introduced . Three ML methods are utilized in the implementation: Random Forest, DT, as well as HM (Hybrid of RF & DT). The test outcomes reveal that the HDP model with the hybrid system has an accuracy level of 88.7 percent. The functionality is intended to collect the user's input parameters in order to forecast HD, for which author utilized a DT & RF hybrid system[49].

Saxena et al.,(2015) developed a system which could effectively explore the regulations to anticipate the risk status of customers depending on the provided health parameter. The regulations could be prioritized depending on the requirement of the user. The program's system is assessed in means of classification accuracy, as well as the findings demonstrate that the scheme has huge potential in forecasting the level of risk of heart disease more precisely[50].

Repaka et al.,(2019) The suggested technique consists of various phases: dataset collection, user registration as well as login (application-based), classification using NB, forecasting, & safe data transfer using AES. Following that, an outcome is generated . The report discussed as well as provides a variety of expertise abstraction methods that are used for HDP utilizing DM techniques. The results show that the standard clinical program is efficient in forecasting risk factors for heart disease[51].

Jayshril et al.,(2014) display a multilayer perceptron neural network-based prediction framework for HD. This program's NN acknowledges 13 clinical characteristics and is educated to use a back-propagation method to estimate the occurrence of HD in the patient with a 98 percent precision when compared to other processes. The result achieved with this scheme demonstrates that it is superior and more effective as compared to other technologies[52].

Srinidhi et al.,(2021) to conduct a thorough examination of LSTM-based Deep Learning DL designs with several productivity measures on the MIT-BIH arrhythmia database for heartbeat classification. For classification purposes, various variants of the LSTM- DL design are suggested. Across the variants, the bi-directional LSTM DL model performs well in the classification of Normal beats (97%), PVC beats (98%), APC beats (98%), as well as PB beats (99 %). In the categorization of heartbeats, the BiLSTM DL design has 95 percent sensitivity as well as 98 percent specificity when compared to existing works. The results demonstrate that the LSTM Ml algorithms are suitable for heartbeat classification[53].

Senthilkumar et al.,(2019) developed a mechanism for identifying important features using machine learning methods, leading to increased cardiovascular disease prediction accuracy. The forecast method is formulated with various feature combinations and many well-known classification algorithms. Author achieve an enhanced performance amount with an accuracy level of 88.7 percent by combining the combination RF with a linear design in the forecasting structure for heart disease (HRFLM)[54].

Fatma et al.,(2020) suggest a new hybrid model to CVD prediction utilizing distinct ML approaches like LR, AdaBoostM1, MOEFC , FURIA, GFS-LB, and FH-GBML. The accuracy as well as outcomes of every classifier were contrasted for this aim, with the finest classifier selected for a high precise CVD prediction. Author use two pieces of free software to achieve this goal (Weka and Keel). GFS-LogitBoost achieves the highest accuracy of 94.17, which is higher than the other two methodologies[55].

Abhay et al.,(2019) discovered that different methods had varying degrees of accuracy, ranging from ML algorithms to deep neural networks, but were unable to acquire good outcomes for silent heart attack forecasting. suggests a heart attack predictive design focused on Deep Learning methods, especially RNN, to estimate the degree of the patient's heart-related diseases. As a result of our analysis, author decided to use RNN and GRU to improve the system's accuracy and efficiency in predicting silent heart attacks and informing the user as soon as possible. This scheme has improved the frequency of heart attack forecasting to 92 percent and has proven to be an excellent source for forecasting silent heart attacks[56].

Shelda et al.,(2019) to evaluate the application of a DL method to simulate the rate of HD through a common benchmark database (the University of California, Irvine (UCI) database). The quality of a DL approach was contrasted to the performance of 4 common machine learning techniques (two linear as well as two nonlinear) in forecasting the rate of HD utilizing information from 567 participants from 2 groups from the UCI dataset. When contrasted to existing designs, the deep learning design achieved the highest accuracy of 94 percent as well as an AUC score of 0.964. Deep learning & nonlinear ML methods performed significantly strong than linear ML techniques as database size increased[57].

Kathleen et al.,(2018) An improved DNN learning system was created to help patients as well as healthcare experts, as well as to improve the correctness of HD diagnosis and prognosis in patients. The advanced DNN learning design relies on a deeper multilayer perceptron architectural design with deep learning regularization as well as dropout. The advanced DNN learning category comprises a classification design focused on training data & a prediction method for identifying new patient cases utilizing a database of 303 clinical instances from Cleveland Clinic Foundation patients treated with coronary heart disease. The DNN classification as well as prediction model obtained the required outcomes in testing: diagnostic accuracy of 83.67 percent, sensitivity of 93.51 percent, specificity of 72.86 percent, precision of 79.12 percent, F-Score of 0.8571, area under the ROC curve of 0.8922, K-S test of 66.62 percent, DOR of 38.65, and 95 percent confidence interval for the DOR test of [38.65, 110.28]. As a result, the established deep learning classifier as well as data sets can present highly consistent and valid diagnoses for coronary heart disease while lowering the amount of erroneous diagnoses that may harm patients[58].

Joon et al.,(2018) focused on developing as well as verify a deep learning-based echocardiography-based mortality prediction design for HD utilizing DL. There were 25 776 patients in the study, with 1026 deaths. The DL model's regions under the AUROC were 0.912, 0.898, 0.958, as well as 0.913 for internal validation, external validation, CHD & HF, in both, outperforming other comparative designs[59].

Shadab et al.,(2021) suggested a novel DL structure focused on a 1D CNN for categorization of healthy and as well as non-healthy people to achieve the shortcomings of conventional methods. Numerous medical variables are used to assess a patient's risk profile, which aids in early diagnosis. In the

suggested network, numerous strategies are used to avoid over-fitting. On the dataset, the suggested network achieves over 97 percent training accuracy as well as 96 percent test accuracy. The model's accuracy is contrasted in depth with other classification methods using several performance parameters, demonstrating the efficiency of the suggested architecture[60].

Nazir et al.,(2018) created a ML-based diagnosis framework for prediction HD through using a HD database. Author's utilized 7 common ML techniques, 3 feature selection approaches the CV technique, as well as seven performance evaluation metrics for classifiers like classification accuracy, specificity, sensitivity, Matthews' correlation coefficient, as well as execution time. The suggested device could simply distinguish between people with HD and those who are healthy. Moreover , ROC as well as area under the curves were calculated for every classifier. All of the classifiers, feature selection methods, preprocessing methods, validation methods, as well as classifier productivity evaluation metrics utilized by author. The suggested platform's performance has been verified on both full features as well as a subset of characteristics. The lowering of characteristics has an effect on classifier production in means of accuracy as well as execution time. The suggested ML-based decision support program will help doctors in efficiently diagnosing heart patients[61].

Saboji et al.,(2017) designed and analyzed a scalable approach for forecasting HD attributes. Author implemented the RF technique on the Spark system for forecasting heart disease as well as demonstrated that proposed system can accomplish 98 percent accuracy with as few as 600 dataset records[62].

P. Sujatha et al.,(2020) Decision Tree, Naive Bayes, Random Forest, SVM, KNN , as well as logistic Regression methodologies are used to indicate the degree of heart disease. The design was assessed utilizing metrics such as Accuracy, Precision, AUC, and F1-score. According to the observational data, the Random Forest is more precise for predicting heart disease than other supervised machine learning approaches, with an accuracy of 83.52 percent. Random forest classifiers have an F1-score of 84.21 percent, an AUC of 88.24 percent, and a precision score of 88.89 percent, in both[63].

Aslam et al.,(2021) presented a framework for determining similarity based on Hierarchical RF creation as well as NR design Model (HRFFNRM). Using this framework, which predicts cardiovascular diseases with a 90.3 percent accuracy[64].

Divya et al.,(2019) suggest an extensive preprocessing model for detecting Coronary Heart Diseases (CHD). Removing null values, re-sampling, standardization, normalization, classification, as well as estimation are all part of the method. The motive of study is to forecast the risk of coronary artery disease utilizing ML techniques such as RF, DT, as well as KNN. In addition, a comparison research of these methods rely on prediction accuracy is carried out. K-fold Cross Validation is also utilized to create

randomness in the information. These methods were compared on the "Framingham Heart Study" database , which contains 4240 records. RF, DT, as well as KNN accomplished accuracy rates of 96.8 percent, 92.7 percent, and 92.89 percent, respectively, in proposed experimental analysis. As a consequence of incorporating proposed preprocessing steps, RF classification produces more accurate outcomes as compared to other ML techniques[65].

Chunyan et al.,(2020) suggested to detect heart disease, a recursive improved RF with an improved linear design(RFRF-ILM) was used. aims to explore the main elements of cardiovascular disease prediction using machine learning methods. The predictive algorithm involves different feature subsets and formed classification techniques. Through the heart disease prediction model, it achieves a higher level of performance with greater precision. Actions must be implemented to protect this disease, as well as Diabetes requires an extra factor that should be considered in the incidence of coronary artery disease with 96.6 percent accuracy, 96.7 percent stability ratio, and 96.7 percent F-measure ratio[66].

Pranav et al.,(2020) suggest a ML method for estimating the likelihood of having heart disease utilizing numerous approaches. The structure is implemented by five methods: RF, NB, SVM, HDT, as well as LMT. The Cleveland database is utilized to train as well as test the design. The dataset is pre - processed, and then feature selection is used to pick the most prominent characteristics. The generated database is then used to train the structure. When the outcomes are merged, it is clear that Random Forest provides the best accuracy[67].

Yuepeng et al.,(2020) To monitor out the main features that may cause heart disease, built a heart disease prediction method focused on random forest and LSTM. Then, use the LSTM, KNN, and DNN methodologies to see if the prediction accuracy improves after testing, as well as eventually, select the most accurate method to build the heart disease prediction method[68].

Shaik Farzana et al.,(2020) ML algorithms like Gaussian NB, RF, KNN, SVM, as well as Xg-Boost are used to create an effective HDP structure. The work proposes 13 characteristics like age, gender, blood pressure, cholesterol, obesity, cp, and so on. It is a user-friendly device with several phases. In the first phase, author upload the database file as well as choose the method to run on the chosen database. The accuracy of every selected approach is forecasted along with a graph, as well as the modal is produced for the one with the highest number through training the database to it. In the next step , input for each heart variable is provided, as well as the diseased stage of the heart is maximum likelihood on the modal produced. Author then take precautions premised on the patient's condition. The suggested method is effective in predicting a victim's heart illness. The HDP concept proposed in this view is a one-of-a-kind technique that could be used within the category of heart disease[69].

Norma et al.,(2020) suggests an effective HDPM for a CDSS that includes DBSCAN to identify as well as avoid outliers, a hybrid SMOTE-ENN to balance the training data distribution, as well as XGBoost to forecast heart disease. The 2 datasets (Statlog and Cleveland) were utilized to construct the design as well as analyze the outcomes with those of other designs NB, LR, MLP, SVM, DT, and RF & of preceding research findings. The suggested framework surpassed other designs as well as previous study outcomes, accomplishing accuracies of 95.90% and 98.40% for the Statlog and Cleveland databases, in both[70].

Savitha et al.,(2020) offers a web-based device for predicting HD utilizing ML methodologies with a high level of accuracy when contrasted to previous works. It employs an ensemble classification algorithm to estimate heart disease, as ensemble techniques outperform individual classifiers like SVM or RF[71].

Sanchayita et al.,(2018) introduce an effective model for estimating HD utilizing ML approaches . As a result, a hybrid design for HP was suggested, utilizing RF classifier as well as simple k-means method ML algorithms. The database is also analyzed and compared by 2 ML approaches, such as the J48 tree classifier as well as the Naive Bayes classifier. The findings acquired using the Random forest classifier as well as the corresponding confusion matrix demonstrate the robustness of methodology[72].

Halima et al.,(2020) suggested a clinical support models for detecting HD to assist clinicians in making better diagnostic decisions. ML approaches like NB, KNN, SVM, RF, as well as DT are used to forecast Heart Disease based on risk factor information collected from medical files. Various investigations have been carried out to forecast HD using the UCI data set, and the results show that Nave Bayes outperforms both cross-validation as well as train-test split methods, with accuracy of 82.17 percent and 84.28 percent, respectively[73].

Fen Miao et al.,(2018) Using an enhanced random survival forest, researchers created a comprehensive risk method for analyzing heart failure mortality with high accuracy (iRSF). By employing a novel split rule as well as stopping criterion, the suggested iRSF was able to recognize high accurate predictors to distinguish survivors as well as nonsurvivors and thus enhance discrimination ability. Depend on the public MIMIC II clinical dataset of 8059 patients, 32 risk factors, such as demographics, clinical, laboratory data, as well as medications, were evaluated and utilized to create the risk design for patients with heart failure . The experimental outcomes demonstrated that the developed risk outperforms previous studies as well as the conventional random survival forest-based design, with an out-of-bag C-statistic value of 0.821. As a result, the advanced iRSF-based risk design can be a useful tool for clinicians in predicting heart failure mortality[74].

Obasi et al.,(2019) applied a ML-based system that could identify and forecast heart disease in patients utilizing patient medical records. The suggested approach is focused on current methods such as RF Bayesian Classification as well as LR, so it offers a DSS for medical professionals to identify as well

as estimate heart diseases and heart attacks in humans & individuals utilizing heart disease risk factors. After preprocessing, the database used in suggested framework contains 18 features (risk factors) as well as 1990 observations. Then it was divided into 80 percent train sets and 20% test sets. A field experiment was carried out utilizing real medical records of patients to analyze the quality and accuracy of the suggested scheme. The scheme, which forecasts the risk of heart disease in patients, was built in the RStudio framework. The contrasted outcomes revealed that the device production as well as accuracy are adequate, with HDC accuracy of 92.44 percent for RF, 61.96 percent for NB Classifier, and 59.7 percent for Logistic Regression, in both[75].

Hasan et al.,(2018) Using the info gain feature selection method & eliminating unnecessary features various classification algorithms like KNN, DT, GNB, LR, as well as RF are utilized for accurate evaluation on a HD database. To formulate the performance of the classification methods, various performance measurement factors like accuracy, ROC curve, precision, recall, sensitivity, specificity, and F1-score are used. Logistic Regression outperformed the others, with a classification accuracy of 92.76 percent[76].

Shan et al.,(2017) concentrating on developing a more accurate approach risk prediction device focused on DM approaches in order supply auxiliary medical services. The method consists of 4 parts in order to be used in healthcare industries for gathering and examining patient data: data interface, data preparation, FS, as well as classification. Data interface response is needed to acquire actual information from hospitals; data preprocessing is required for data integration, data cleaning, and rating mapping, among other things. To avoid dimensionality, key factors were extracted using CFS Subset Formulation in conjunction with the Best-First-Search approach. A previous trial in the CVD risk prediction area used random forest as a simple classifier to detect risk level. The CHDD & PKU People's Hospital's Cardiology inpatient set of data were both evaluated for validity and feasibility. In the CHDD test, proposed system outperforms other methodologies with a considerably maximum accuracy of 91.6 percent. In the People's Hospital database experiment, it achieved the accuracy of 97 percent, which is best as compared to the other classifier except SVM (98.9 percent), but RF takes half time that as compared to SVM [77].

Sarah Saud et al.,(2020) focused on building a predictive design for CAD diagnosis The theory was established utilizing computational machine learning methods such as Random Forest (RF) as well as NB. The research will conduct use of an open source database called Z-Alizadeh sani to aid in the diagnostic procedure. Here, in the literature survey, NB outperformed the other experiments with 100 percent sensitivity as well as a negative predictive rate of 100 percent. With 13 features, NB outperformed RF with an accuracy of 83 percent. The outcomes obtained encourage the use of the built designs as supporting software in the treatment method[78].

Indu et al.,(2017) To estimate exactly the occurrence of HD in a specific patient, the researchers used a combination of ensemble approaches (BT, RF,& AdaBoost) as well as a feature subset selection technique - PSO. The test findings demonstrate that the Bagged Tree and PSO reached the best accuracy[79].

Mohamed et al.,(2020) contributes to the study and growth of an intelligent device for HDP depend on the LSTM method . A comparison of MLP as well as LSTM methodologies in terms of accuracy as well as other predictive variables for heart disease is described. The primary goal is to create an intelligent program focused on the LSTM method for predicting heart disease so that an appropriate decision can be made to protect as well as regulate heart disease and stroke. It has higher effectiveness as compared to the MLP method, LSTM is shown to be the most effective approach for solving aforementioned problems[80].

Aishwarya et al.,(2019) configured a Dynamic LSTM risk method using algorithms as well as produced a training or test model that produces a more accurate result than the previous LSTM risk algorithm The suggested methodology has the benefit of being able to work on dynamic dataset documents as well as create a more accurate classified outcome for HDP[81].

Samiul et al.,(2019) suggest a method to estimate CVD risk factors using attention module-focused LSTM, which has nearly 95% accuracy & 0.90 MCC scores; stronger as compared to any other previously suggested approaches. Furthermore, present a novel Intelligent Healthcare Platform for prolonged data collection as well as patient monitoring. Initially, the suggested system is utilized to collect information, as well as extract the best suitable datasets for use with several ML approaches .The findings outcomes demonstrate that the attention module-based LSTM strong other statistical ML techniques for CVD prediction and identifies significant risk factors, that could help CVD patients improve their health[82].

Pandiaraj et al.,(2021) intends to diagnose heart disease using a data analytic process that incorporates support vector machine (SVM) as well as GA. The findings outcomes indicate that the suggested method outperforms for predicting heart disease as compared to other existing techniques[83].

Kanika et al.,(2017) suggested using random forest as well as NB to forecast HD. In addition, a method for selecting features prior to classification is suggested in order to enhance model performance. SVM-RFE and gain ratio techniques are analyzed in this section for feature selection, which outcomes in the assignment of weight to each feature. This method aids in the improvement of accuracy as well as the reduction of computational time. The experimental outcomes show that the suggested method of choosing features improves the accuracy of both models[84].

CHAPTER-3

OBJECTIVES AND METHODOLOGY

3.1 Research Gap

According to reports, half of all Americans have at least one symptom of chronic disease. Unexpectedly, this outcomes in chronic disease therapy accounting for 80% of US healthcare spending. The impact of these illnesses grows in lockstep with the rise in population. The United States itself has spent nearly $2.7 trillion per year on corresponding therapies. The United States is not the only country in which large amounts of money are expended on chronic illnesses. According to reports, chronic illnesses kill the bulk of citizens in China, accounting for more over 85 percent of all deaths in the world's fastest growing country. Obviously, early detection are important, not only to save money, but also to save lives and enhance quality of care. The goal of the health-care device is to correctly estimate how a person is at risk of heart attack. This forecast would be made using machine learning techniques on training data provided by us. After the user enters the requested data, the technique is utilized as well as the result is produced . When the medical information is missing, the accuracy is predicted to reduce. Developer apply the predictive models to real-world medical records. In the traditional model, a convolutional neural network approach was developed as a disease risk prediction model utilizing structured as well as possibly unstructured patient data. But CNN has some drawbacks so that we need to update exiting system with new algorithm and as per the literature study recurrent neural network (RNN) more preferred for data classification.

3.2 Objectives

- To implement deep learning and machine learning algorithm for HDP.
- To improve the accuracy of HD prediction with hybrid of Bi-LSTM and Random forest
- To perform comparative analysis of proposed and conventional system

3.3 Proposed Methodology

1. First step read the heart disease prediction dataset from excel in MATLAB software using **xlsread** function
2. Second step is pre-processing of dataset in phase we have analyze the dataset and separate data into two category input attributes and target.
3. To initialize the bidirectional LSTM layer, training option like number of epochs, learning rate etc.

4. Fourth step is training of bidirectional LSTM and prediction of score for heart disease patient
5. Next is step to initialize machine learning algorithm random forest and perform training of machine learning algorithm
6. Last step predict the score value of random forest and combine the both prediction score value
7. Evaluate performance parameter.

CHAPTER-4

RESULTS

Heart disease is one of the leading origin of death in the globally . Cardiovascular disease forecasting is a complicated issue in CDA. ML as well as deep learning have been shown to be useful in guiding in decision making as well as estimation from the huge amount of information generated via the healthcare industry. The primary aim is to identify a heart disease prediction system in MATLAB utilizing a prior database. The goal of this study is to use datasets that represent actual data to allow the predictive model to draw conclusions from any advanced data.

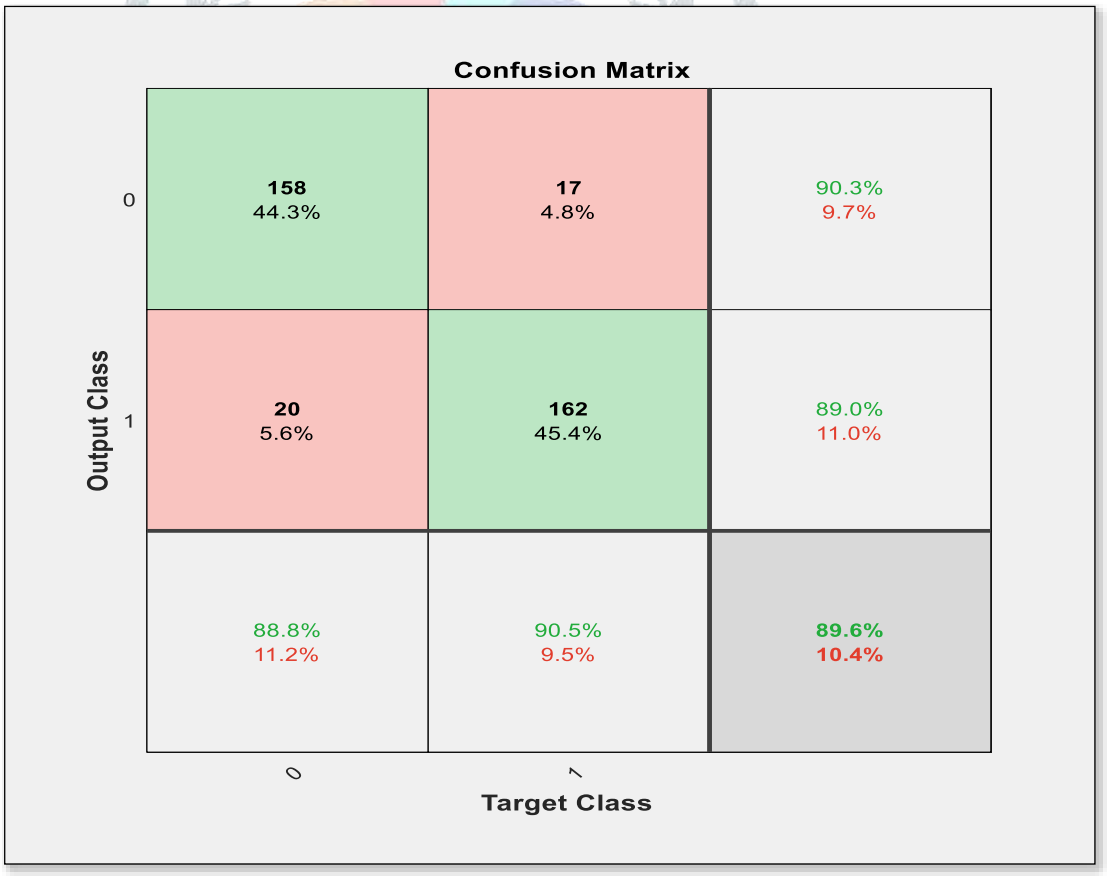

Figure 4.1 : Confusion matrix between output class and Target class

Figure 4.1 depicts the confusion matrix among the output class as well as the target class. CM is utilized to estimate the importance of a classifier's output on a database. The diagonal components indicate the set of points for which the predicted label equals the true label, whereas off-diagonal components are those for which the classifier misidentified.

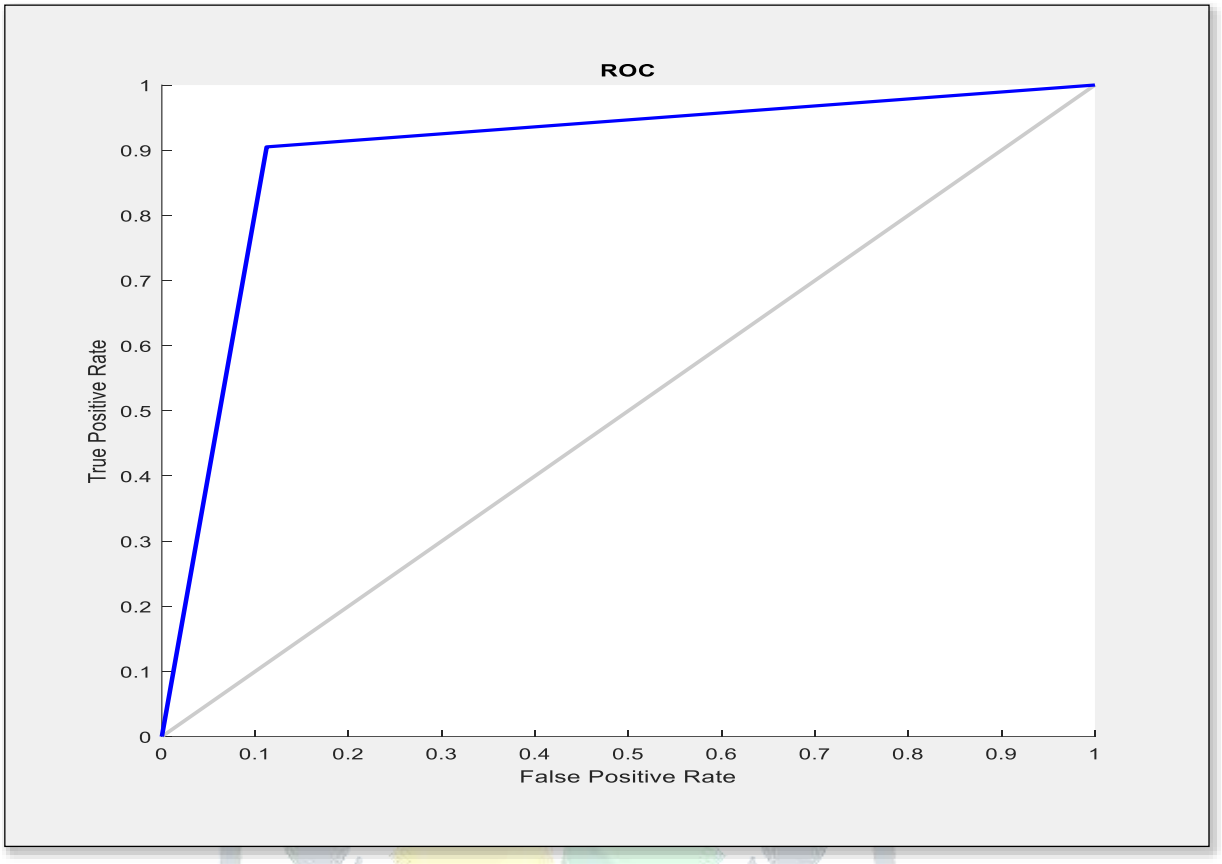

Figure 4.2:Graphical representation of ROC

The receiver operating characteristic (ROC) plot is a widely used metric for assessing classifier performance. The ROC plot is built around two fundamental evaluation metrics: specificity & sensitivity. Specificity is a metric of the negative part's performance, while sensitivity is a metric of the positive part's performance. In relation to predicted labels, the majority of machine learning techniques generate a kind of score. These scores could even take the form of discriminant values, posterior probabilities, etc. Moving threshold values across the scores yields model-wide evaluation metrics.

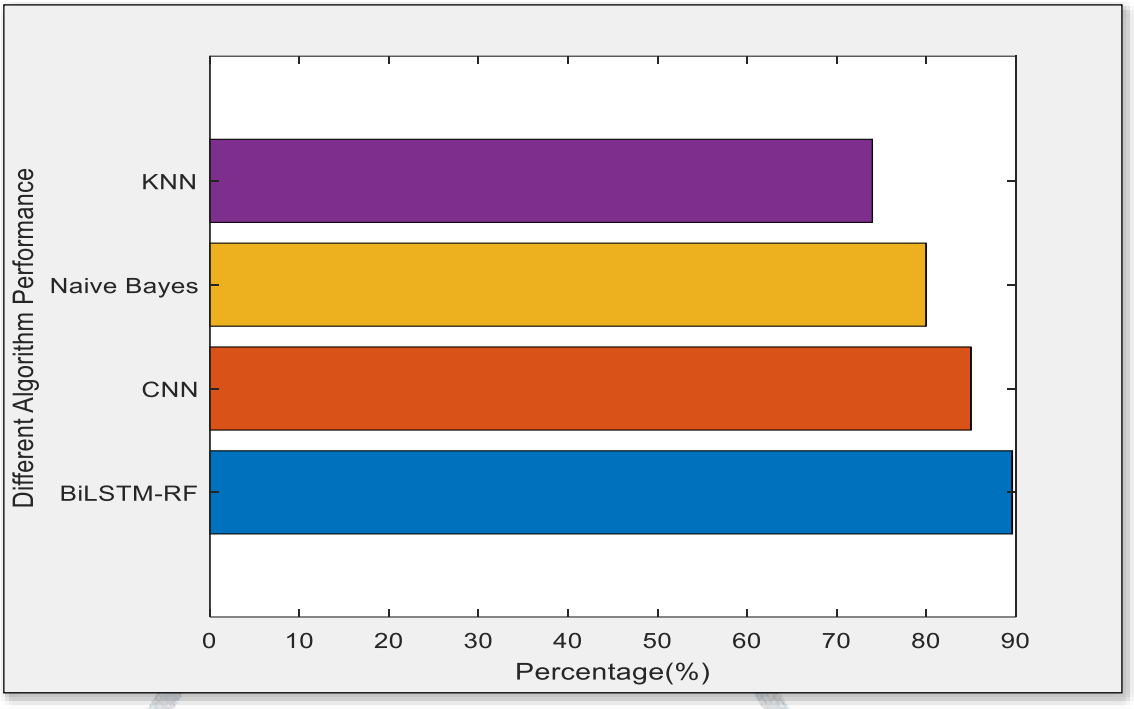

Figure 4.3: Accuracy by using various algorithms

The accuracy readings used in evaluating research methods are used to generate the results. The following table compares multiple metrics among multiple methodologies.

Table 4.1 Comparsion between various algorithms in terms of Accuracy

S. No	Algorithms	Accuracy
1.	KNN	73 %
2.	Naïve Bayes	80 %
3.	CNN	85 %
4.	Bi-LSTM-RF	90 %

Table 4.1 compares the forecast precision of ML methods (RF, SVM, or KNN), DL designs (LSTM & GRU), as well as suggested methodology. This Hybrid approach maintained an accuracy of 90 percent, which outperformed the estimated accuracy of each individual design. This method may be quite valuable in assisting doctors in investigating patient cases in way to support their instruction.

CHAPTER-5

CONCLUSION N AND FUTURE WORK

5.1 Conclusion

Heart diseases are a serious threat that usually occur when the arteries that pass o_2 as well as blood to the heart become completely blocked or narrowed. There is a huge volume of data generated in medical groups that is not being used properly. As a result, an effective prediction method is able for early detection. A few DL as well as ML technologies have been implemented to enhance predictive accuracy.

Heart disease has recently become primary cause of death in men's as well as women's. As a result, heart disease forecasting is regarded as an important component of clinical statistical analyses. In past studies, standard machine learning approaches such as SVM, Nave Bayes, as well as other deep learning approaches were used to forecast heart disease. Due to a lack of test data, these methodologies are incomplete for efficient heart disease prediction. BiLSTM with RF Method has been presented in this thesis to enhance the precision of HDP. For effective analysis, the input medical results were recorded bidirectional. The BiLSTM-RF system has been evaluated on the Cleveland data - set to evaluate the behavior as well as contrasted with existing approaches. The outcomes demonstrate that the suggested BiLSTM-RF strong as compared to the current models for calculating HD. The suggested BiLSTM-RF has a significantly higher accuracy as contrasted to existing methodologies.

5.2 Future Work

This finding suggests the strategies for HDP, as well as the creative BILSTM-RF, which is a machine learning and deep learning approach. In the future, the study intended to estimate heart diseases utilizing DL algorithms approaches and larger datasets. DL will then optimize as well as improve the forecast procedure in means of velocity.

REFERENCES

[1] Deepthi, S & Ravikumar, A., "Computation Methods for the Diagnosis and Prognosis of Heart Disease", International Journal of Computer Applications, vol. 95, no. 19,2014.

[2] Soni, J, Ansari, U, Sharma, D & Soni, S., "Intelligent and effective heart disease prediction system using weighted associative classifiers", International Journal on Computer Science and Engineering, vol. 3, no. 6, pp. 2385-2392,2011.

[3] Zhiyong Wang, Xinfeng Liu, J. G., "Identification of metabolic biomarkers in patients with type-2 diabetic coronary heart diseases based on metabolomic approach", 6(30), 435–439,2016.

[4] Gawande, N., & Barhatte, A., "Heart diseases classification using convolutional neural network", 2nd International Conference on Communication and Electronics Systems (ICCES),2017.

[5] P. de Chazal et al., "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," IEEE transaction biomedical engineering, vol. 51, no. 7, pp. 1196–1206, Jul. 2004

[6] Lee, J., Jung, J., Lee, J., & Kim, Y. T. "Acute myocardial infarction detection system using ECG signal and cardiac marker detection", IEEE SENSORS ,2014.

[7] Zhen Yang, Dao Min Zhou, "Cardiac markers and their point-of-care testing for diagnosis of acute myocardial infarction", Clinical Biochemistry, Volume 39, Issue 8, pp.771-780,Aug 2006.

[8] Sowmiya, C & Sumitra, P. , 'Comparative study of predicting heart disease by means of data mining', vol. 5, no. 12, pp. 19580- 19582,2016.

[9] Brijain, M., Patel, R., Kushik, M. and Rana, K., "A survey on decision tree algorithm for classification", Journal of Machine Learning Research , Vol. 11, No. 2,pp. 649–672,2014.

[10] Tamilarasi, R & Porkodi, R, "A study and analysis of disease prediction techniques in data mining for healthcare", International Journal of Emerging Research in Management &Technology , vol. 4, no. 3, pp. 76-82,2012.

[11] Karaolis, M. A., Moutiris, J. A., Hadjipanayi, D., & Pattichis, C. S., "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees", IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 3, pp. 559–566,2010.

[12] Lakshmi, KR, Krishna, MV & Kumar, SP. , "Performance comparison of data mining techniques for predicting of heart disease survivability", International Journal of Scientific and Research Publications, vol. 3, No. 6, pp. 1-10,2013.

[13] Patil, RR. , "Heart disease prediction system using naïve bayes and jelinek-mercer smoothing", International Journal of Advanced Research in Computer Science and Communication Engineering, vol. 3, no. 5, pp. 6787-6789,2014.

[14] Venkatalakshmi, B & Shivsankar, MV., "Heart disease diagnosis using predictive data mining", In 2014 IEEE International Conference on Innovations in Engineering and Technology (ICIET'14), Tamil Nadu, India, Vol. 3, No. 3, pp. 1873-1877,2014.

[15] Goularas, D., & Kamis, S., " Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data", International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML),2019.

[16] Suzuki, Y., Iwashita, S., Sato, T., Yonemichi, H., Moki, H., & Moriya, T., " Machine Learning Approaches for Process Optimization", International Symposium on Semiconductor Manufacturing (ISSM),2018.

[17] Srinivasan, K., & Fisher, D., "Machine learning approaches to estimating software development effort", IEEE Transactions on Software Engineering, Vol. 21, No. 2, pp. 126–137.,1995.

[18] M. Svensén and C. M. Bishop, "Pattern recognition and machine learning," New York : Springer, 2007.

[19]Y. Guo, et al., "An extensive empirical study on semi-supervised learning" , in IEEE International Conference on Data Mining, pp. 186-195,2010.

[20] Z.-H. Zhou and M. Li, "Semi-supervised learning by disagreement", Knowledge and Information Systems, vol. 24, pp. 415-439, 2010.

[21] X. Guo, et al., "Deep learning for real-time Atari game play using offline MonteCarlo tree search planning", Advances in Neural Information Processing Systems, pp. 3338-3346,2014.

[22] L. Wang, "Support vector machines: theory and applications", Springer Science & Business Media, Vol. 177, 2005.

[23] L. Shi, et al., "The research of support vector machine in agricultural data classification,", in International Conference On Computer And Computing Technologies in Agriculture, pp. 265-269,2011.

[24]R. C. Barros, et al., "Evolutionary design of decision-tree algorithms tailored to microarray gene expression data sets", IEEE Transactions on Evolutionary Computation, vol. 18, pp. 873-892, 2013.

[25] V. Podgorelec, et al., "Decision trees: an overview and their use in medicine", Journal of Medical Systems, Vol. 26, pp. 445-463, 2002.

[26] G. Biau, "Analysis of a random forests model", Journal of Machine Learning Research, Vol. 13, pp. 1063-1095, 2012.

[27] E. E. Tripoliti, et al., "Automated diagnosis of diseases based on classification: dynamic determination of the number of trees in random forests algorithm", IEEE Transactions on Information Technology in Biomedicine, Vol. 16, pp. 615-622, 2012.

[28] S. Taheri and M. Mammadov, "Learning the naive Bayes classifier with optimization models", International Journal of Applied Mathematics and Computer Science, Vol. 23, pp. 787-795, 2013.

[29] Y. Hamid, "An Improvised k-NN Respecting Diversity of Data for Network Intrusion Detection", International Journal of Intelligent Engineering and Systems, Vol. 10, pp. 409-417, 2017.

[30] I. Aleksander, et al., "A brief introduction to Weightless Neural Systems", in European Symposium on Artificial Neural Networks, pp. 299-305,2009.

[31] Kusuma.S1 , Divya Udayan.J, "Machine Learning and Deep Learning Methods in Heart Disease (HD) Research", International Journal of Pure and Applied Mathematics, Vol. 119 , No. 18 , 1483-1496,2018.

[32] Lesole Kalake, Wanggen Wan, Li Hou, "Analysis Based on Recent Deep Learning Approaches Applied in Real-Time Multi-Object Tracking: A Review", IEEE Access , Vol. 9, pp. 32650 – 32671, Feb 2021.

[33] Palangi, H., Ward, R., & Deng, L., "Distributed Compressive Sensing: A Deep Learning Approach", IEEE Transactions on Signal Processing, Vol. 64, No. 17, pp. 4504–4518, 2016.

[34] She, X., & Zhang, D., "Text Classification Based on Hybrid CNN-LSTM Hybrid Model", 11th International Symposium on Computational Intelligence and Design (ISCID),2018

[35] Maragatham, G., & Devi, S., "LSTM Model for Prediction of Heart Failure in Big Data", Journal of Medical Systems, Vol. 43, No. 5 , 2019.

[36] Manohar M., Alok Kumar P., Pankaj K., "A Prediction Technique for Heart Disease Based on Long Short Term Memory Recurrent Neural Network" International Journal of Intelligent Engineering and Systems, Vol.13, No.2, 2020.

[37] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems with Applications", Vol. 72, pp.221-230, 2017.

[38]L. Breiman, "Random forests", Machine Learning, vol. 45, pp. 5-32, 2001.

[39] Jabbar, M. A., Deekshatulu, B. L., & Chandra, P., "Prediction of Heart Disease Using Random Forest and Feature Subset Selection", Innovations in Bio-Inspired Computing and Applications, 187–196,2015.

[40] Huaiyu Wen, Sufang Li., Amin Ul Haq, Rajesh Kumar, "Elimination of Irrelevant Features and Heart Disease Recognition by Employing Machine Learning Algorithms using Clinical Data", 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP),2020.

[41] Rahul Deo Sah, Sibo Prasad , Neelamadhab, Nagesh Salimath, "Diabetics Patients Analysis Using Deep Learning and Gradient Boosted Trees", 8th International Conference on Computing for Sustainable Global Development (INDIACom), Mar 2021.

[42] Sharma, V., Rasool, A., & Hajela, G., "Prediction of Heart disease using DNN", Second International Conference on Inventive Research in Computing Applications (ICIRCA),2020.

[43] Xiao, C., Li, Y., & Jiang, Y., "Heart coronary artery segmentation and disease risk warning based on a deep learning algorithm. IEEE Access, 2020.

[44] Kefaya Qaddoum; Azmi Alazzam; Rula Al Azawi, "A Deep Neural Network heartbeat classification approach for arrhythmia detection", Seventh International Conference on Information Technology Trends (ITT),2020.

[45] Adeen, P. Sondhi, "Random Forest Based Heart Disease Prediction", International Journal of Science and Research (IJSR), Volume 10 Issue 2, February 2021.

[46] Ankita Dewan, Meghna Sharma, "Prediction of heart disease using a hybrid technique in data mining classification", 2nd International Conference on Computing for Sustainable Global Development (INDIACom),2015.

[47] Md. Touhidul I.,; Sanjida R., Md. Golam K., "Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means", 23rd International Conference on Computer and Information Technology (ICCIT),Dec 2020.

[48] Sayali Ambekar, Rashmi Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network", Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),2018.

[49] M. Kavitha; G. Gnaneswar; R. Dinesh; Y. Rohith Sai; R. Sai Suraj, " Heart Disease Prediction using Hybrid machine Learning Model", 6th International Conference on Inventive Computation Technologies (ICICT),Jan 2021.

[50] Purushottam, Saxena, K., & Sharma, R., "Efficient heart disease prediction system using decision tree", International Conference on Computing, Communication & Automation,2015.

[51] Repaka, A. N., Ravikanti, S. D., & Franklin, R. G., " Design And Implementing Heart Disease Prediction Using Naives Bayesian", 3rd International Conference on Trends in Electronics and Informatics (ICOEI),2019.

[52] Jayshril S. Sonawane D. R .Patil, "Prediction of heart disease using multilayer perceptron neural network", International Conference on Information Communication and Embedded Systems (ICICES2014), Feb 2014.

[53] Srinidhi H., S. G M, Kiran M H M & K G Srinivasa, "A comparative study and analysis of LSTM deep neural networks for heartbeats classification", Health and Technology , Vol. 11, pp. 663–671 , 2021.

[54] Senthilkumar Mohan, Chandrasegar T., G., Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access , Volume: 7 ,pp. 81542 – 81554, Jun 2019.

[55] Fatma Zahra A., Menaouer Brahami, Nada Matta, "A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques", International Conference on Smart Homes and Health Telematics, pp 299-306, volume 12157,Jun 2020.

[56] Abhay K., Ajay K., Karan S., Maninder P., Y.Hambir, "Heart Attack Prediction Using Deep Learning", International Research Journal of Engineering and Technology (IRJET), Vol. 05, No. 04 , Apr-2018.

[57] Shelda Sajeev, Anthony Maeder Stephanie , "Deep Learning to Improve Heart Disease Risk Prediction", International Workshop on Machine Learning and Medical Engineering for Cardiovascular Healthcare, pp. 96-103,Oct 2019.

[58] Kathleen H. Miaoa, , Julia H. Miaoa, "Coronary Heart Disease Diagnosis using Deep Neural Networks", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 9, No. 10, 2018.

[59] Joon-Myoung K. , Kyung-Hee , J. Park, "Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography",IEEE,2018.

[60] Shadab H., Susmith B., Shadab A., Md Suaib, "Novel Deep Learning Architecture for Heart Disease Prediction using Convolutional Neural Network",Machine learning, May 2021.

[61] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R., "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Mobile Information Systems, pp. 1–21,2018.

[62] Saboji, R. G., "A scalable solution for heart disease prediction using classification mining technique", International Conference on Energy, 2017.

[63] P. Sujatha; K. Mahalakshmi, "Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease", IEEE International Conference for Innovation in Technology (INOCON), Oct 2020.

[64] Mohamed Aslam , Jaisharma K., "Hierarchical Random Forest Formation with Nonlinear Regression Model for Cardiovascular Diseases Prediction", International Conference on Computer Communication and Informatics (ICCCI),Jan 2021.

[65] Divya Krishnani , Anjali Kumari, Akash Dewangan, Aditya Singh, Nenavath Srinivas Naik, "Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms", IEEE Region 10 Conference (TENCON),2019.

[66] Chunyan Guo, Jiabing Zhang; Yang L., Yaying Xie, Zhiqiang Han, Jianshe Yu, "Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform", IEEE Access , Vol. 8 , pp. 59247 – 59256,March 2020.

[67] Pranav M., Ankita D., G. Suganya, M Premalatha, "Cognitive Approach for Heart Disease Prediction using Machine Learning", International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE),2020.

[68] Yuepeng L., Mengfei Z., Zezhong F., Yinghan C., "Heart disease prediction based on random forest and LSTM", 2nd International Conference on Information Technology and Computer Application (ITCA), Dec 2020.

[69] Shaik Farzana, D.Veeraiah, "Dynamic Heart Disease Prediction using Multi-Machine Learning Techniques", 5th International Conference on Computing, Communication and Security (ICCCS),Oct 2020.

[70] Norma Latif Fitriyani, Muhammad S., Ganjar A., Jongtae Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System", IEEE Access , Vol. 8, pp. 133034 – 133050, July 2020.

[71] Savitha Kamalapurkar; Samyama Gunjal G H., "Online Portal for Prediction of Heart Disease using Machine Learning Ensemble Method(PrHD-ML)", IEEE Bangalore Humanitarian Technology Conference (B-HTC),2020.

[72] Sanchayita Dhar, Krishna Roy; Tanusree Dey; Pritha , Ankur Biswas, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases", International Conference on Computing Communication and Automation (ICCCA),Dec 2018.

[73] Halima El., Saïd Boujraf, Nour El Houda, M. Maaroufi, "A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques", 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sept 2020.

[74] Fen Miao; Yun-Peng Cai; Yu-Xiao Zhang; Xiao-Mao Fan; Ye Li, "Predictive Modeling of Hospital Mortality for Patients With Heart Failure by Using an Improved Random Survival Forest", IEEE Access , Vol. 6 , pp. 7244 – 7253, Jan 2018.

[75] Thankgod Obasi, M. Omair Shafiq, "Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases", IEEE International Conference on Big Data (Big Data), Dec 2019.

[76] S. M. M. Hasan, M. A. Mamun, M. P. Uddin , M. A. Hossain, "Comparative Analysis of Classification Approaches for Heart Disease Prediction", International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2),Feb 2018.

[77] Shan Xu, Zhen Zhang , Daoxian Wang, Junfeng Hu., Xiaohui Duan, "Cardiovascular risk prediction method based on CFS subset evaluation and random forest classification framework", IEEE 2nd International Conference on Big Data Analysis (ICBDA), Mar 2017.

[78] Sarah Saud , Yasmeen A., S., Nida Asalam , "Automated prediction of Coronary Artery Disease using Random Forest and Naïve Bayes", International Conference on Advanced Computer Science and Information Systems (ICACSIS),2020.

[79] Indu Yekkala; Sunanda Dixit; M. A. Jabbar, "Prediction of heart disease using ensemble learning and Particle Swarm Optimization", International Conference On Smart Technologies For Smart Nation (SmartTechCon), Aug 2017.

[80] Mohamed Djerioui, Youcef Brik, Bilal A., "Heart Disease prediction using MLP and LSTM models", International Conference on Electrical Engineering (ICEE), Sep 2020.

[81] Aishwarya Mishra , Dayashankar S., "Heart Disease Predictions Using Numerous Classification Techniques and Dynamic LSTM Model", International Conference on Communication and Electronics Systems (ICCES), July 2019.

[82]Samiul I., Haider M. Umran, Mohammed K., "Intelligent Healthcare Platform: Cardiovascular Disease Risk Factors Prediction Using Attention Module Based LSTM", 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), May 2019.

[83] A. Pandiaraj, S.Lakshmana, Prakash, "Effective Heart Disease Prediction Using Hybridmachine Learning", Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV),Feb 2021.

[84] K. Pahwa, Ravinder K., "Prediction of heart disease using hybrid technique for selecting features", 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), Oct 2017.