DEEP LEARNING BASED CLINICAL DECISION SUPPORT SYSTEM FOR THE PREDICTION OF HEART DISEASES

Dr. J. B. Shajilin Loret[1], S. Praveeshna[2],
1. Professor, Department of IT , Francis Xavier College of Engineering, India
2. PG Student, Department of IT , Francis Xavier College of Engineering, India

Abstract; Heart disease is the major health issue or challenge faced by the entire world in modern medicine. It has become a crucial factor for increasing the mortality rate. The desperation of heart disease is more vital and can even result in vulnerable consequences if it is not predicted at the initial stage. The methods such as electronic health records, monitoring body are network continuously and diagnosing the patient health condition via the medical sensors projection and wearable device across human bodies. Since the generated data from the human body are continuous and huge in amount, the data mining techniques are utilised for efficient classification of obtained health data. Moreover, the classification of health data is the most critical process as it needs an accurate execution with the early detection of heart disease. As most medical enthusiasts and practicing physicians finding reveals that difficult to diagnose the disease at an early stage is the key for the failures of incurable disease. The hardest thing for the medical community to do is save lives, thus it's critical to diagnose patients as soon as possible. By using an efficient feature selection and classification-based prediction system, this research aims to lower the risk of heart disease. Thus, a major constraint of any current procedure is the earlier and more accurate prognosis of cardiac disease. So, this research attempts to develop an efficient classifier and predicts heart disease at the initial stage with high-performance measures and accuracy. The significant contribution of this research is divided into three parts.

First, an effective method is implemented to predict heart disease by feature selection and classification. The proposed research comprises of optimised MLP-EBMDA (Multi-Layer Perceptron for Enhanced Brownian Motion-based Dragonfly Algorithm) for classification for heart disease prediction and an unsupervised technique for feature selection. The input in this implementation will be obtained from the dataset and performed pre-processing followed by the proposed feature selection technique that efficiently performs the selection of features. Heart disease classification VI is carried out utilizing innovative MLP-EBMDA, which aids in the early prediction of heart disease based on certain traits. This proposed technique may effectively predict cardiac disease as normal or abnormal with an accuracy rating of 94.28 percent.

INTRODUCTION

1.1 AN OVERVIEW OF HEART DISEASE AND RISK

In the current scenario, data are dissipated as Reports, Forms, and Statistics and so on. They are used as inputs for the various kinds of techniques. Since the technology has been booming currently, many techniques have been invented and they are going on with the invention of nullifying the issues occurring in the every field. The technology has become very helpful for identifying the defects in the particular field and it helps a lot for rectifying the issues in a short period. This growing technology plays a major role mainly in the health care domain. This has benefited a lot in case of creating the result in the real-time situation. Despite that, lot of studies and researches have been conducted in different fields, and mainly the medical domain has extended the utilization of technology for accessing the data officially, and estimating the results which could be displayed all over the world [1].

According to a World Health Organization (WHO) report, heart disease, or more specifically cardiovascular disease, is a major contributing factor to the high death rate worldwide. One of the body's components, the heart pumps and circulates blood to every area of the body, including the brain, playing a crucial role for all other sections. If the blood circulation is stopped by the heart to the brain and to different nerves of the body, this causes the death of the nerve system [2] i.e. all nerves and tissues present in the parts of the body will stop working and it will result in the occurrence of death. Therefore, the life of the living being totally depends on the heart. Hence, proper functioning of heart is required for each individual in order to have a healthy 2 life.[3] It is important to identify the disease in the early stage to provide appropriate treatment at the correct time in order to reduce the mortality rate (Beyene et al. 2018). Moreover, the prediction of the heart disease is the major problem in the current situation

While considering the current pandemic situation, the coronavirus disease called COVID-19 has resulted in high rates of mortality and morbidity. This cardiovascular disease is the major cause for increase of mortality and morbidity in the United States, and it is also associated with the unfavorable outcome of a patient suffering from multiple illnesses, such as bacterial pneumonia and

influenza. It is not possible to forecast COVID-19 patients using the recently collected COVID-19 data[4]. In a similar vein, COVID-19 has been linked to a high load of inflammation known as a "inflammatory storm" or "cytokine storm," which induces myocarditis, vascular inflammation, and cardiac arrhythmias and further increases heart damage.

The heart diseases or cardiovascular diseases are classified into various types of diseases which need to be predicted in the earlier stage[5]. This is one of the emerging diseases all over the world and it increases high mortality rate all over the world. Most of the people have lost their life due to this disease. This disease has many risk factors that have to be avoided and the precaution measures have to be undergone in case if the patient has already infected by the heart disease[6]. The patients who are affected by the heart disease or cardiovascular diseases should follow the safety measures and the precaution must be taken as per the doctor's advice in order to reduce the infection rate of the heart disease.

1.1.1 Types of Cardiovascular Disease

Heart disease, commonly referred to as cardiovascular disease, or CVD, is a group of illnesses affecting the heart and blood vessels. Heart disease can be classified into several types, with myocardial infraction—also referred to as angina or a heart attack—being one of them[7]. Coronary heart disease-CHD is also another cardiovascular disease, which is created due to presence of some waxy substance called plaque that develops in the inner part of the coronary arteries[8]. These coronary arteries will be used for supplying the blood with rich content of oxygen to the heart muscle. When the waxy substance plaque starts to grow up in these arteries, this condition is known as atherosclerosis. This growth of plaque will happen inside the artery for many years. When the growth of the plaque is not noticed in the initial stage, it affects the heart to become rupture-break open or harden. The plaque that tends to more hard in nature will make the coronary artery narrow in shape that causes reduction in the flow of rich content of oxygen blood to the heart. If the plaque gets harden, a blood clot will be formed on the surface[9]. The huge size of blood clot can most of the time will block the flow of blood completely via coronary arteries. The blood flow has to retain quickly. If not so, the part of the heart muscle starts to die. The quick treatment must be given when there is an occurrence of heart disease, otherwise it may lead to serious health issues and it can even cause death.

1.1.2 Risk Factors of Heart Disease

The risk factors are the reasons for the increase of blockage. These risk factors are categorized into two: modifiable risk factors and non-modifiable risk factors. The parameters of non-modifiable risk factors are gender, age and heredity. These risk factors cannot be changed and they will be the main resource for the occurrence of the heart disease. Modifiable risk factors are the risk factors, which can be modified by our own efforts. Some of the modifiable risk factors are (1) Habit related (2) Stress related (3) Food related (4) Miscellaneous and bio chemical risk factors. The various types of heart diseases are coronary, atherosclerosis, rheumatic, congenital, myocarditis, angina and arrhythmia.

Figure 1.1 Risk factor of heart disease

1.2 THE NEED FOR HEART DISEASE PREDICTION

Around the world, the estimated death is around 17.5 million and that occurs due to cardiovascular disease. This cardiovascular disease mostly occurs in the middle-income and low-income people and it has resulted in 75 percentage of mortality rate[10]. Moreover, 80 percent of the deaths happens due to cardiovascular disease because of heart attack and stroke. As per the report of WHO, India is one of the countries with increasing number of cardiovascular disease patients added in every year. Since the patient who is affected by heart disease increases every year, two lakh open-heart surgeries are conducted in the each year[12]. In the past few years, patient count increased at a rate of 20%-30% which has become a major concern.

1.3 VARIOUS PREDICTION METHODS FOR HEART DISEASE

It is necessary to get quicker attention from the medical field, in case the heart attack happens, it helps to prevent the damage of heart and to maintain the patient life with a heart attack[13]. The medical professional uses the emerging technology for monitoring the data of the heart patients constantly and providing the continuous advice to the patients in order to heal from the heart disease. Therefore, the usage of data by utilizing computer technology for medicine has been increasing now a day.

The data mining techniques explain the past and predicts the future with the help of analysing the data. These data mining techniques are the combination of various fields like machine learning, database technology

and artificial intelligence[14]. Data mining techniques are used in various types of applications and importantly, it plays a major part in predicting the disease in the earlier stage. The general process for predicting the heart disease is like extracting the significant data from the vast amount of data by feature extraction process[15]. Secondly, the extracted data will be trained with the selected dataset and then, data will further be given to the testing process. This is done with different classification methods. This technique is also called as KDD-Knowledge Discovery of data.

2.RELATED WORK

2.1 INTRODUCTION

According to the report of WHO-World Health Organisation, the major death rate has increased due to Health diseases. This health issue has been caused due to mental stress, workload and many other sources. Accordingly, the treatment of the heart patients is given based on the lab test, history of patients and the answer given by the patients asked by the doctors. The hybrid technique of data mining and machine learning techniques are used to predict the cardiovascular disease. By using these hybrid techniques, medical technicians will get the valuable data about the patients who have cardiovascular diseases, which enhance them to give proper diagnosis for the patient. Various ways to predict the cardiovascular disease have been reviewed in three different phases. In the first phase, the prediction of heart disease is done using machine learning algorithm with the help of classifiers and feature selection techniques. In the second phase, the heart disease prediction is performed by using deep learning technique based on entropy with Random forest classifier and dimensionality reduction is used. In the third phase, the identification of heart disease is done using deep learning techniques. Various investigations have been done to predict the heart disease using different types of machine learning and data mining techniques in healthcare organisation and they are described in the below sections.

2.2 ROLE OF FEATURE EXTRACTION AND SELECTION IN HEART DISEASE PREDICTION

This section elaborates the previous researches related to the selection and extraction of diverse techniques used. The Feature selection plays a vital role in removing irrelevant and redundant features and decreases the training time and cost of the predictive models.

2.2.1 Feature Selection Model for Heart Disease Prediction using Machine Learning Techniques

To detect heart diseases by balancing trained data and XGBoost, a new technique by Fitriyani et al. (2020), called the HDPM heart disease prediction model, was effectively implemented for a clinical decision support system supported by DBSCAN density-based spatial clustering with noise and using SMOTE- ENN (synthetic minority sampling technique). -Nearest Neighbor edited). The results of implementing two separate datasets, statlog and Cleveland, are discussed along with the results of other models including LR (Logistic Regression), Navies Bayes – NB, Decision Tree –DT, Multilayer Preceptor –MLP, SVM –Support Vector Machine. , and the results of previous studies. The results of this study were inferred from these two models, yielding an accuracy of 95.90% for the statlog dataset and 96.40% for the remaining dataset.

The hybrid methods have been introduced by Babu et al. (2018) for identifying various types of diseases using the combination of Grey Wolf optimisation and auto-encoder based on RNN (GWO+RNN).The GWO is used to extract the features and the identification of disease is done by RNN method which have resulted better performance than the existing methods. Various datasets like Cleveland, mammographic, and Hungarian are used and their outputs are compared with the proposed model. The outcome obtained has shown the improvement in the accuracy by 16.82 percent compared to other methods. The improvement can be done by implementing various techniques for attaining more accuracy than the implemented techniques.

The major focus of Gokulnath et al. (2019) is to optimize the function by combining support Vector machine-SVM with Genetic Algorithm-GA to select the features significantly. The results that are obtained by the proposed model are compared with other techniques like CFS, chi squared, Filtered subset, Info gain and consistency subset. The proposed technique has produced 88.34 percent accuracy with the selected features for the prediction of heart disease when comparing with other techniques. Additionally, these techniques have proved that ROC analysis provides good performance in the SVM classifiers.

The main goal of Gokulnath et al. (2019) is feature optimization by combining Support Vector machine-SVM with Genetic Algorithm-GA to meaningfully select features. The results obtained by the proposed model are compared with other techniques such as CFS, chi-square, filtered subset, information gain subset, and consistent subset. The proposed technique provides 88.34% accuracy with selected features for heart disease prediction compared to other techniques. Additionally, these techniques demonstrated that ROC analysis provides good performance in SVM classifiers.

2.2.2 Different Feature Extraction Models for Identification of Heart Diseases

There are number of classifiers used for the prediction of cardiovascular disease. The initial stage of heart disease has been predicted by Nayak et al. (2019), using different types of classification methods like Navies Bayes, Support Vector Machine, Decision tree classifier, and k-NN classifiers. They are used to identify the early stage of disease and for the early prevention. By analysing various techniques for detecting the heart disease with the initial stage of their un-healthiness, the accuracy obtained by Navies Bayes is better than the other classification techniques. This technique has been implemented in R data analytical tool for identifying the

cardiovascular disease after better filtration. By ensemble machine learning, there will be better accuracy as the part of the enhancement.

The utilisation of decision tree algorithm has been explained by Maji et al. (2019) as the classification method that can be used for the prediction of cardiovascular diseases. In this technique, hybridization method has been implemented by combining the decision tree and artificial neural network as a single technique to attain improvement in the prediction of heart diseases with the help of WEKA. This performance has been evaluated using tenfold evaluation test and it is helpful to analyse the heart disease patient dataset that has been taken from UCI dataset. The performance metrics like accuracy, specificity and sensitivity are calculated for both hybrid and individual classifier techniques. These different combinational techniques are introduced to predict the heart disease and it is helpful to the clinical technicians to detect different types of diseases in the initial stage. The output obtained from the hybridization techniques have resulted that the proposed technique shows better performance than the individual technique and it shows good result for the heart disease prediction. In the enhancement, various data mining techniques can be used for predicting different types of diseases earlier.

Generally, the classifiers are used in machine learning techniques to deliver the defects of the person and they are introduced for predicting heart diseases by evaluating different parameters like HbA1c, cholesterol level, ECG results and heart rate. Kumar et al. (2021) have evaluated different types of machine learning techniques and their classifiers depending on the accuracy rate and the computation time over the prediction of heart disease datasets to identify the heart diseases. The classifiers that are used for the identification of heart diseases are LMT-Logistic Model Trees, Random Forest, J48, Random Tree classifiers and Hoeffding tree. After the completion of analysis, the result obtained by the evaluated datasets concludes that the Hoeffding tree has resulted with the better accuracy of 85.1852 percent than the other classifiers and the computation time for predicting the heart disease patient is 0.17s duration.

Generally, the classifiers are used in machine learning techniques to deliver the defects of the person and they are introduced for predicting heart diseases by evaluating different parameters like HbA1c, cholesterol level, ECG results and heart rate. Kumar et al. (2021) have evaluated different types of machine learning techniques and their classifiers depending on the accuracy rate and the computation time over the prediction of heart disease datasets to identify the heart diseases. The classifiers that are used for the identification of heart diseases are LMT-Logistic Model Trees, Random Forest, J48, Random Tree classifiers and Hoeffding tree. After the completion of analysis, the result obtained by the evaluated datasets concludes that the Hoeffding tree has resulted with the better accuracy of 85.1852 percent than the other classifiers and the computation time for predicting the heart disease patient is 0.17s duration.

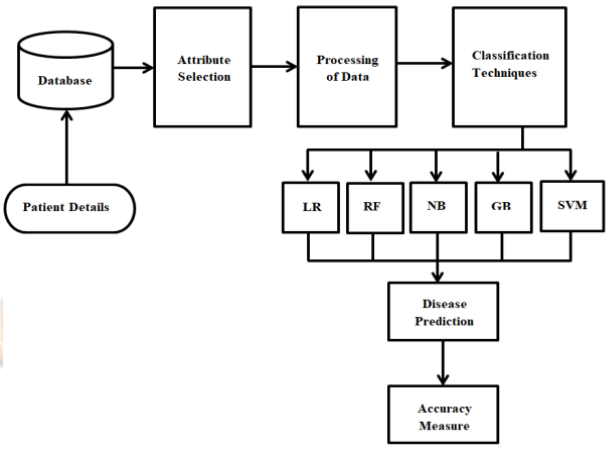


Figure 2.1 Proposed system for heart disease prediction

Uddin et al. (2019) have used supervised machine-learning algorithm for single disease prediction. Two different datasets like Scopus and PubMed are used for the identification of cardiovascular diseases. 48 articles are selected to compare the variants among these supervised machine - learning algorithms for the prediction of diseases. The main objective of this article is to detect the trends among various types of supervised algorithms in machine learning and their performances are evaluated as well as the risk is predicted in the track of particular disease. From the analyses, the Support Vector Machine – SVM algorithm has been applied in 29 studies and followed by Navies Bayes has been applied in 23. From these studies, the conclusion inferred by the authors is that Random Forest is superior in terms of accuracy than the other techniques with 53 percent accuracy rate. Hence, SVM techniques with the accuracy rate of 41 percent from the studies have been followed.

(Source: Javeed *et al.* 2019)

Figure 2.2 Block diagram of diagnostic system

The proposed architecture of Ganesan et al. (2019) has been used to detect the cardiovascular disease in an efficient manner with the help of UCI dataset as well as the sensors are used in healthcare to identify the people who are suffering from heart diseases. The algorithm that is used for classification has been divided into two phases namely training phase and testing phase. The

dataset considered as benchmark is used as the training data for the classifiers in the training phase. Similarly in testing phase, the information that has been collected from the patient is used to identify the disease. The benchmark datasets that are used for testing are J48, LR-Logistic Regression, SVM-Support Vector Machine and MLP-Multilayer perception whose performances are evaluated to find the better among them. The performance metrics used for the evaluation are Accuracy rate, F-score, Precision, kappa 38 value and recall. The result that has been obtained by this study shows that the appropriate algorithm is J48 for the health care prediction model in IoT environment comparing of other techniques.

2.3 ROLE OF CLASSIFIERS IN HEART DISEASE PREDICTION

This section enumerated the past studies regarding to the various classifiers established during the research to diagnose heart diseases precisely and accurately.

Uyar et al. (2017), to diagnose the heart disease prediction, have used genetic algorithm based on the RFNN: Recurrent Fuzzy Neural Networks. The dataset that has been used for this technique is Cleveland dataset based on the University of California Irvine-UCI repository. The total numbers of samples used are about 297 information, 252 are used for training the data and 45 of them are used for testing the data. Through this technique 97.78 percent of accuracy rate has been recorded from the testing dataset. Moreover, the other performance metrics are also evaluated for the comparison of the ANN-Fuzzy AHP approach. The result has displayed that the proposed technique has shown a better performance than the other techniques.

To detect the heart disease in the earlier stage, a new technique has been implemented by Jayaraman et al. (2019) by using artificial gravitational cuckoo search algorithm along with the bee optimising included with the neural network to handle the features for predicting the heart diseases in the initial stage. At the initial stage, for the prediction of heart disease, the related data of heart disease dataset based on the UCI dataset have been collected. Since the collected data are vast in dimension, it will be difficult to process the data and it will result in reduction of efficiency. To overcome this issue, the proposed technique has been implemented for the dimensionality reduction. The proposed technique has attained maximum efficiency when compared with the other traditional techniques. Finally, the authors have suggested that by including the optimization technique in the training phase, the system effectiveness could be improved as a future enhancement.

3.METHODOLOGY

3.1 INTRODUCTION

This Chapter describes Heart disease prediction using FTGM-YOLO with informative entropy based-random forest. According to the World Health Organisation (WHO) report, heart disease is the main reason that causes death. Predicting heart disease early and treating it in a corrective way can reduce the death rate. This is the main reason for introducing this method.

Fast track gram matrix-principal component analysis is termed as FTGM-YOLO. The fast-track gram matric-principal component analysis is mainly used for dimension reduction and fusion in solving the over-fitting issues. It also helps to minimize space and time, improves the performance of the classifier and removes the redundant data. The principal component analysis is a method that is used in linear algebra and simple matrix operation for the calculation of original data into a reduced dimension or to the same number.

The informative entropy based-random forest is termed as IEB-RF. This method is introduced in this system to improve the classifier performance. Though it can flexibly handle high data, it also has high accuracy in the classifier. The random forest can perform both regression and classification. Also, it gives a good prediction, and it is in the form of understanding easily. To interpret the received non-parametric model & high dimensional, it frequently becomes infeasible.

3.2 PROPOSED METHODOLOGY

Due to high cause of death in heart disease, this research is tent to focus on heart-related diseases. To reduce the death rate, predicting the disease earlier is the best way. Though there are various classification models for predicting the heart disease, the accuracy in the classification outcomes is very low. The main issues notified in the existing systems are efficiency, selecting relevant features, increased accuracy and fast dimensionality.

In the introduced deep Convolutional Neural Network (deep CNN), the feature extraction and fusion are carried out only for the selected relevant features. This study is aimed to solve the issues faced in ML (Machine Learning) algorithms for classification and DL (Deep Learning) algorithms for feature extraction based on data mining techniques.

Machine learning is a sub-field of Artificial Intelligence (AI) that allows the software application to predict the results more accurately. Deep Learning (DL) is a type of both artificial intelligence and Machine Learning (ML) and it works as the knowledge of the human. Deep learning is mostly used on the important elements of data that involve predicting models and statistics.

The overall process of the proposed methodology is shown below in Figure 3.1. The flow graph represents the step-by-step process of the introduced method.

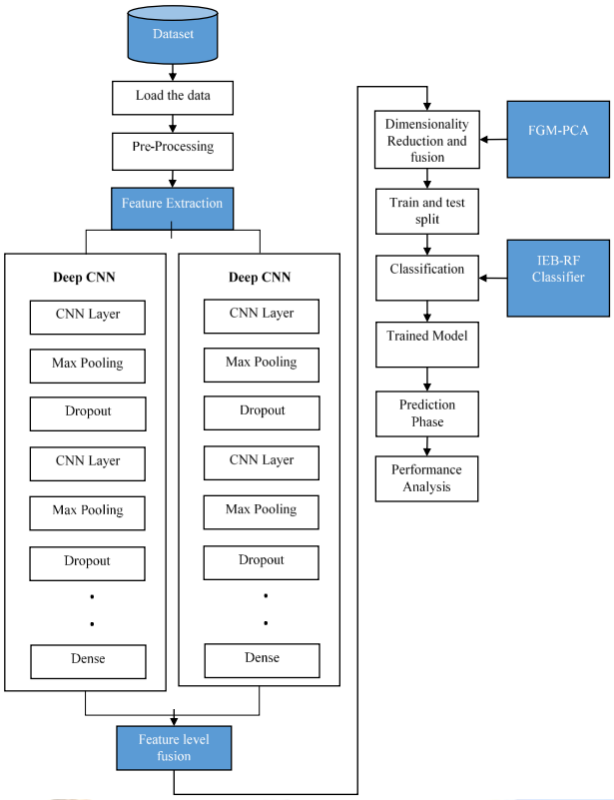


Figure 3.1 Overall view of the proposed system for predicting heart disease

The proposed system consists of various processes that are pre-processing, feature extraction and classification, which are explained below. In the initial step, pre-processing is carried out with the loaded Cleveland dataset of heart disease. The Cleveland heart disease dataset, found in the UCI machine learning source, consists of 14 variables measured from 303 heart disease patients.

The UCI machine learning is a collection of datasets. This process helps the user to get a highly informative dataset comprising data cleaning in which the unwanted datasets are eliminated. As the next step, through the proposed deep CNN model, the fusion and feature extraction are performed. This improves the reliability and classifier efficiency by determining the compact feature and information.

By this process, dimension counts are minimized. Hence, the data storage space is reduced and it also reduces computation time and training. Besides, it helps in the visualization of data. The Principal Component Analysis helps (YOLO) to overcome the over fitting issues of data by decrementing the features. The above-obtained result is fed to test especially for split and train. Then, the proposed IEB-RF classifier is used to perform the classification.

The classification helps to differentiate the data between the suspected patient data and normal person data. Finally, a trained model is used to accomplish the prediction, and according to the performance exploration, the efficiency is assessed.

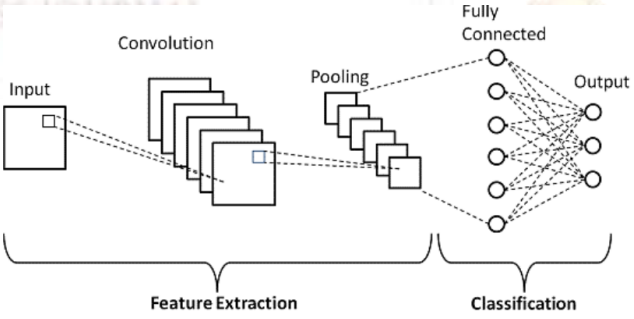
3.2.1 Feature Extraction and Fusion - Deep Convolutional Neural Network

In recent times, DL has been confirmed as a valuable tool for predicting disease, since it has a huge data handling ability. In this study, for feature extraction, two deep CNN models are used. In that, 128 features are extracted from the 1st model, and 246 features are extracted from the 2nd model. The above mentioned models have the ability to learn the filters and to permit latent concept description for pattern recognizing hierarchically by subsequent integration.

Various layers are present in these models, and at each layer, the filter gets reduced. Only the significant features get extracted with an initial layer possessing 4*4*1*64 filters that finally get reduced to 3*3*14*1 filters. The outputs are found by feature maps corresponding to CNN, once the filters are given as input. The feature map at each layer indicates outputs for the specific layer.

Deep Convolutional Neural Network (DCNN) is the most commonly used type to identify the patterns in the video and image. From the traditional artificial neural networks, deep convolutional neural networks have progressed. Deep CNN focuses on applications like image classification, object detection and recommendation system. It is also used for normal language dispensation.

This is because all the extracted features are relevant. The dense indication obtained for 40 features is further reduced through dimensionality reduction and this will lead to computational complexity and better performance.



(Source: Li *et al.* 2018)

Figure 3.2 Framework of the convolutional neural network

Figure 3.2 depicts the several layers of CNN comprising convolution layer and max-pooling layer which assist in effective feature extraction. After the extraction, the classification is done. From the classification, the output is retrieved. A CNN (Convolutional Neural Network) is

an artificial neural network class used to filter up inputs for valuable information by the convolutional layers. The input data are involved in the convolutional 87 operation with a filter (kernel) to form a transformed feature map. This network automatically adjusts to find the best feature on the base of the given task.

3.2.2 Dimensionality Reduction and Fusion-Fast Track Gram Matrix-Principal Component Analysis

A Fast Track Gram Matrix-Principal Component Analysis (FTGM-YOLO) has been proposed in this study and the kernel functions indicate the gram matrices in ML to reduce the dimensionality features in a fast method. Principal Component Analysis (YOLO) helps to overcome the overfitting issues in the given data by decreasing the number of feature. The proposed Fast Track Gram Matrix-Principal Component Analysis (FTGM-YOLO) improves the performance of the dimension reduction by which the classifier performance is increased. Output shows high variance. Hence, visualization is improved.

The linear independence is calculated by Gram Matrix (GM), since it is the significant way of removing the redundant features and this is the main application of GM. The pseudo-code of FTGM is given in Algorithm 4.1. The input of n × n in GM (Gram Matrix) is taken, and as output, it returns GM~ k = CHSWk + CHST in which the decomposition is approximate. The given input value is iterated till the condition gets true. The correlation matrix and the gram matrix are computed through the below-mentioned Algorithm 3.1.

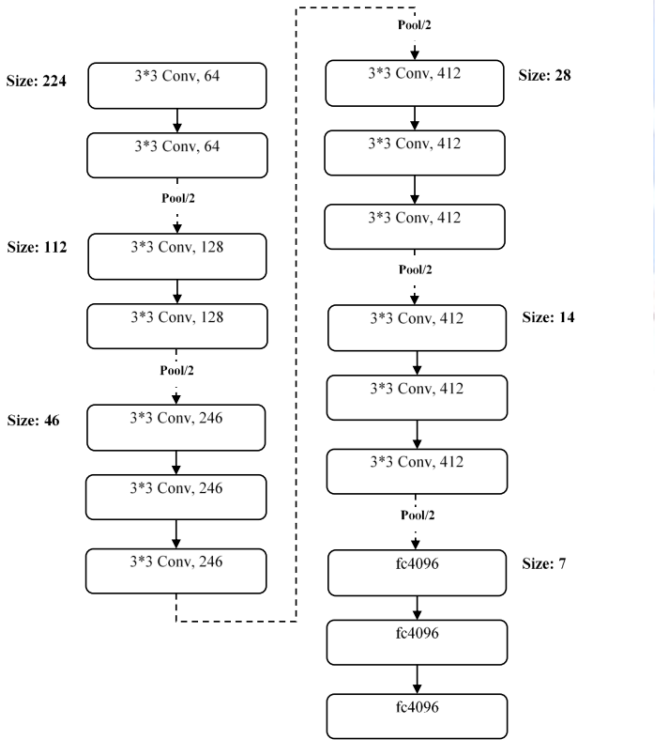


Figure 3.3 CNN architecture for feature extraction

YOLO helps in improving the performance of ML algorithm, and it eliminates the correlated variables, which do not give any contribution to the decision making. This is a technique and by enhancing the low dimension variance, the high dimension data are converted into low dimension data; by this, the dimensionality is reduced. Followed by the above processes, computation is done by eigenvectors with its Cumulative Explained Variance (CEV) and values. The projection matrix is computed when the CEV is > than or the CEV meets the threshold value.

The function is defined as YOLO. The input is given from the d dimension and followed by this, the correlation matrix and gram matrix are calculated. After this, the while statement is applied and from that, sma is assigned as the mean a and mean b to smb. From the achieved values, σab which is the decomposition of Full-YOLO is calculated.

When the statement gets true, the while condition ends. After this, the achieved decompose value is assigned to the eigenvalues and eigenvectors as a magnitude. Cumulative Explained Variance (CEV) is computed. Then, if the condition is applied as CEV which is less than or equal to the threshold, the projection matrix W should be constructed. When the condition is stated to be false, the condition should be terminated. After the termination of the condition, the input is converted to T through W.Through the T ~, the subset features are attained for the dimension k and the value of T ~ is returned

The proposed FTGM-YOLO has been used for dimensionality reduction, and the process involved is shown in Figure 4.4. At the initial stage, the input is taken from the original matrix, and the mean value is determined. Then followed by this, to minimize or eliminate the redundant data, the normalization is performed.

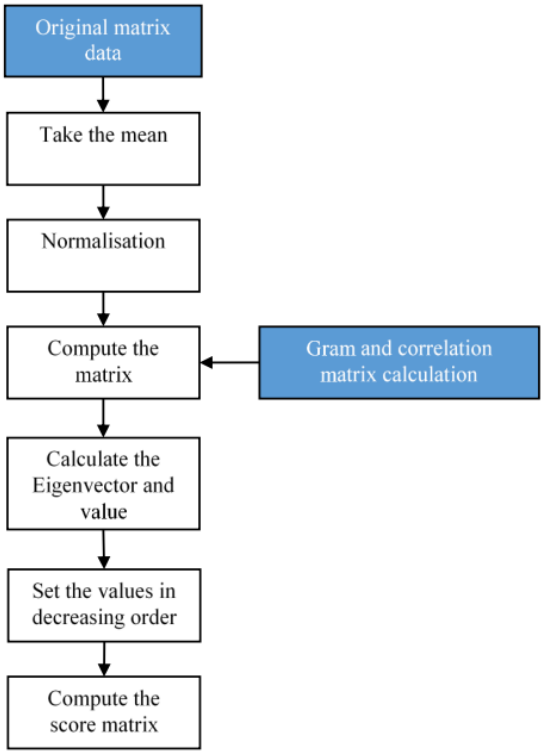
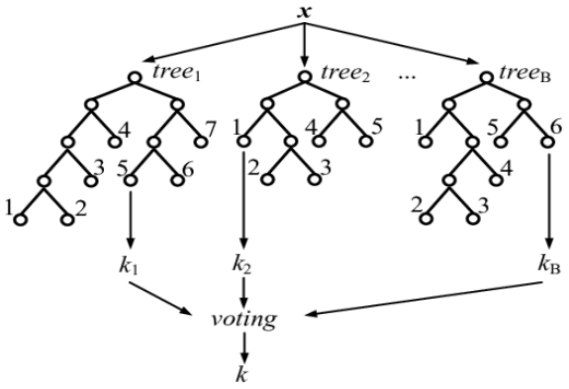


Figure 3.4 Overall process of dimensionality reduction by FTGM-YOLO

3.2.3 Classification-Informative Entropy based-Random Forest

For classification, Informative Entropy-Based Random Forest (IEB-RF) has been proposed in this study. The average information of transferring amount through an event in which all probable results are taken into account for this process information entropy is applied. On the other hand, predictions are effective in RF affords, and it can be understood easily. It can deal flexibly with a huge dataset and has high accuracy in predicting results. Hence, the proposed IEB-RF is used to differentiate the affected patient data and normal person data. In Figure 4.5, the classification of this stepwise process is presented. At first, the nodes of the tree are created. After that, the next split is classification is improved by the attained reduced and relevant features. Figure.4.4. Overall process of dimensionality reduction by FTGM-YOLO 95 built by choosing a subset in training data. This process chooses a subset of a variable. Computation is done for the info gain entropy and Gini index by each of the variables chosen to return the best split and the next split is fed to build. After this process, the computation of prediction error is done and return of predicted value is noted.



(Source: Gelzinis *et al.* 2014)

Figure 3.5 Framework of general random forest The above mentioned

Figure 3.5 represents the architecture of random forest, which is a machine learning algorithm that is supervised and that is mainly used for regression and classification problems. With various samples, the decision trees are built, and the majority of the votes are taken for classification, while average of the tokens is considered for regression.

4.CONCLUSION AND FUTURE WORK

Data mining techniques play an important role in almost every field and help deliver effective results. The proposed study emphasizes the need for machine learning techniques in the process of mining and managing health data. Indeed, heart diseases are considered one of the major concerns in many countries and they even lead to death if not treated appropriately at an early stage. Predicting heart disease in its early stages is considered important and is the subject of various reports on mortality published by the World Health Organization. Therefore, this study focuses on predicting heart disease using machine learning techniques. The main goal of this study is to use machine learning techniques to predict heart disease in its early stages. To achieve this goal, feature selection and classification techniques have been used effectively and efficiently. The effective use of feature selection and classification techniques allows clinical representatives to predict heart diseases in their early stages and help diagnose them with appropriate treatment. This helps save patients' lives, while also leading to reduced mortality rates. To achieve this goal, three different methods were implemented to achieve effective prediction of heart disease. Various data mining techniques have been used to predict heart disease, but they are not effective at the medical representative level. Therefore, to improve the performance of the method, this study focuses on optimizing feature selection and classification techniques. This thesis has been divided into three different methods to analyze their effectiveness as well as find the improved performance of the proposed method. Initially, the research focused on integrating MLP with the improved Motion-Based Dragonfly Algorithm (MLP-EBMDA). This hybrid technique is used in feature selection and classification techniques; An optimized unsupervised method was used to effectively predict heart diseases. The proposed system is evaluated with several parameters such as precision, F1 score, recall and precision rate. The proposed method is compared with other existing methods based on the described parameters and selected experimental

data that helped determine effective prediction of heart disease. The proposed system yields an accuracy rate of about 94.28% in predicting normal and abnormal diseases.

Until now, heart disease prediction has relied solely on machine learning techniques. But this study tried to integrate machine learning with deep learning techniques to effectively predict heart diseases. This research uses deep learning techniques to extract features and machine learning techniques for classification techniques. At this stage, the implementation of deep learning techniques for relevant feature extraction and feature-level fusion helped eliminate irrelevant data through efficient learning. In this implementation, FTGM-YOLO is used for dimensionality reduction purposes, for features and classification, IEB-RF model is used to classify normal and abnormal patient data. The evaluation of this model is compared with other proposed models on meaningful performance measures. The proposed system allows the proposed IEB-RF to perform accurate feature classification while comparing with other existing methods with a rate of 97% for the provided Cleveland Heart Disease dataset.

5. REFERENCE

1. Abdar, M, Książek, W, Acharya, UR, Tan, RS Makarenkov, V & Pławiak, P 2019, 'A new machine learing for an accuate diagnosis of coronary artery disease', Computer Methods and Programs in Biomedicine vol. 179, P. 104992.

2. Abed-Alguni, BH & Barhoush, M 2018, 'Distributed grey wolf optimizer for numerical optimization problems', Jordanian J. Comput. Inf. Technol. (JJCIT), vol. 4, no.

3, pp. 130-149. 3. Alarifi, A, Tolba, A, Al-Makhadmeh, Z & Said, W 2020, 'A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks', The Journal of Supercomputing, vol. 76, pp. 4414-4429.

4. Alkeshuosh, AH, Moghadam, MZ, Al Mansoori, I & Abdar, M 2017, 'Using PSO algorithm for producing best rules in diagnosis of heart disease', in 2017 International Conference on Computer and Applications (ICCA), pp. 306-311.

5. Alsaeedi, AH, Aljanabi, AH, Manna, ME & Albukhnefis, AL 2020, 'A proactive metaheuristic model for optimizing weights of artificial neural network', Indonesian Journal of Electrical Engineering and Computer Science, vol. 20, pp. 976-984.

6. Al-Tashi, Q, Abdulkadir, SJ, Rais, HM, Mirjalili, S, Alhussian, H, Ragab, MG, Alqushaibi & Alawi 2020, 'Binary multi-objective grey wolf optimizer for feature selection in classification', Access, IEEE, vol. 8, pp. 106247-106263.

7. Al-Tashi, Q, Rais, H & Jadid, S 2018, 'Feature selection method based on grey wolf optimization for coronary artery disease classification', in International Conference of Reliable Information and Communication Technology, pp. 257-266.

8. Amin, MS, Chiam, YK & Varathan, KD 2019, 'Identification of significant features and data mining techniques in predicting heart disease', Telematics and Informatics, vol. 36, pp. 82-93.

9. Ayon, SI, Islam, MM & Hossain, MR 2020, 'Coronary artery heart disease prediction: A comparative study of computational intelligence techniques', IETE, Journal of Research, pp. 1-20.

10. Azhar, M & Thomas, PA 2020, 'Heart disease prediction based on an optimal feature selection method using autoencoder', International Journal of Scientific Research in Science and Technology, vol. 7, no. 4, pp. 25-38.

11. Babič, F, Olejár, J, Vantová, Z & Paralič, J 2017, 'Predictive and descriptive analysis for heart disease diagnosis', in 2017 Federated Conferences on Computer Science and Information Systems (Fedcsis), pp. 155-163.

12. Babu, SB, Suneetha, A, Babu, GC, Kumar, YJN & Karuna, G 2018, 'Medical disease prediction using grey wolf optimization and auto encoder based recurrent neural network', Periodicals of Engineering and Natural Sciences (PEN), vol. 6, pp. 229-240.

13. Beyene, C & Kamat, P 2018, 'Survey on prediction and analysis the occurrence of heart disease using data mining techniques', International Journal of Pure and Applied Mathematics, vol. 118, pp. 165-174.

14. Bharti, R, Khamparia, A, Shabaz, M, Dhiman, G, Pande, S & Singh, P 2021, 'Prediction of heart disease using a combination of machine learning and deep learning', Computational Intelligence and Neuroscience, vol. 2021.

15. Buchan, K, Filannino, M & Uzuner, Ö 2017, 'Automatic prediction of coronary artery disease from clinical narratives', Journal of Biomedical Informatics, vol. 72, pp. 23-32, 2017.

16. Burse, K, Kirar, VPS, Burse, A & Burse, R (Eds.) 2019, 'Various preprocessing methods for neural network based heart disease prediction', Smart Innovations in Communication and Computational Sciences, Springer, Singapore.

17. Chantar, H, Mafarja, M, Alsawalqah, H, Heidari, AA, Aljarah, I & Faris, H 2020, 'Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification', Neural Computing and Applications, vol. 32, no. 16, pp. 12201-12220.

18. Chaurasia, V & Pal, S 2014, 'Data mining approach to detect heart diseases', Int J Adv Comput Sci Inf Technol (IJACSIT), vol. 2, pp. 56-66.

19. Choi, E, Schuetz, A, Stewart, WF & Sun, J 2017, 'Using recurrent neural network models for early detection of heart failure on set', Journal of the American Medical Informatics Association, vol. 24, pp. 361-370. 169

20. Dahiwade, D, Patle, G & Meshram, E 2019, 'Designing disease prediction model using machine learning approach', in 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), pp. 1211-1215.