An Effective Heart Disease Detection and Classification using Machine Learning

Abstract: - About 20.5 million people die yearly due to cardiovascular disease (CVD). Early prediction of heart disease and accurate heart severity identification can save the individual life with timely medications. In many cases, most deaths occur due to inaccurate diagnosis of the heart. In computer science, various existing researchers have dealt with the classification of heart disease on synthetic as well as real-time datasets. But still, those systems have a challenge such as low classification accuracy, high error rate and inaccurate heart severity detection. After identifying all these challenges, we proposed an effective heart disease detection and classification using hybrid machine learning techniques. In this article, we describe how various feature extraction and hybrid machine learning classifiers produce accurate severity of heart disease on real-time datasets. In the first phase, we collected a Cleveland dataset from Kaggle and then applied preprocessing and normalization for data balancing. Various feature extraction and selection methods, such as TF-IDF, Co-relation coefficient, N-Gram, and bi-Gram features, are used for practical module training. The different machine learning we used on normalized datasets for classification. The five supervised machine learning algorithms are used, such as Support Vector Machine (SVM), Naïve Bayes (NB), Artificial Neural Network (ANN) and Hybrid Machine Learning (HML) etc. The HML achieves 99.30% accuracy on the Cleveland dataset using Weka 3.8 machine learning framework. As a result, the proposed system compares with various heart disease predictions using machine learning techniques.

Keywords: Heart disease prediction, CVD, data pre-processing, feature extraction, feature selection, supervised machine learning, body sensor network.

I. INTRODUCTION

Due to the imbalanced lifestyles that are common in today's society, severe ailments can easily affect the human heart. As a direct consequence, it is leading to the development of serious illnesses and issues, such as diabetes, stress, and intense smoking. The human heart is adversely affected by all of these elements, and as a result, a range of heart disorders can develop. A lifestyle that includes having hypertension, raised serum triglycerides which lead to higher levels of cholesterol is one of the key risk factors for cardiovascular disease. Smoking increases the amount of blood clotting factor, which can include fibrinogen and other substances like this. Diabetes mellitus is associated with high blood pressure and cholesterol levels. A poor diet that consists of more junk food and meals from outside might lead to more severe forms of heart disease. Diet is very important because it is the major source of human survival. Obesity and a lack of physical activity are major risk factors for heart disease. It is critical for the heart to get regular exercise. There are numerous other factors that can cause more severe forms of heart disease, including having a low socio-economic status, being mentally ill, experiencing high levels of stress exposure, using alcohol, having irregular blood clotting, growing older, and having a cardiovascular disease in the family. Blockage in the coronary arteries, which are the main blood vessels that supply the heart with blood, is another primary cause of heart disease.

In addition to this, the above results in decreased nutrients being delivered to the myocardial cells of the heart muscle. In most cases, there are three major arteries that are responsible for supplying blood to the heart. If any of the arteries become blocked, the patient will experience a cardiac arrest or myocardial infarction. As a consequence of this, this can lead in life-threatening complications, and even death in some situations. The earlier these signs are recognised, the greater the likelihood that prompt treatment will be administered, which in turn increases the likelihood that a human life will be saved. The prediction of cardiac illness with an accuracy level that is sufficient is a very demanding and complex topic; nonetheless, it is possible to achieve this goal by utilising

1Shabeena Naaz Khan
2Dr. Sanjay Y. Azade
3Deepali Sawane
4Sarfaraz Pathan

J. Electrical Systems 20-3s (2024): 869-879

Copyright © JES 2024 on-line : journal.esrgroups.org
more effective machine learning algorithms. Not only can the development of an appropriate ML framework forecast cardiovascular illness with a high level of accuracy, but it can also eliminate the need for human intervention, which is the need of additional medical testing. The death rate and the extent of the illness can be deduced from a quick prognosis.

Previous studies have shown us how common it is for people to suffer from cardiac illnesses which are brought on by leading a sedentary lifestyle and eating an unhealthy diet. At the very least once each year, each and every one of us should have our health checked out so that we are aware of the current state of our various body parts, particularly with regards to heart disease. In the event that elderly people live alone at home and need continual care, there is a need for a device or mechanism that allows for the simple monitoring of required critical parameters of the heart. These parameters are crucial for determining the state of the heart's ability to perform its functions.

According to the research that has been done up to this point, it has been discovered that there are certain machines that can compute or monitor things like an ECG or blood pressure. A glucometer may be used to measure sugar in any form, even random sugar. All of them are distinct computing tools that each compute a unique set of parameters. As a result, what we need is a device that can measure all of the important vital signs of the heart in one location and at one time, and that can be accessed by both the patient and the doctor in the event of an emergency. We also design and develop a heart disease detection and prediction system using IoT and various machine learning techniques. This research we collect data from various sensors and extract various features for classification of machine learning algorithms.

II. LITERATURE REVIEW

Senthilkumar Mohan et al. proposed a hybrid machine learning strategy for accurately predicting heart illness in their study [1]. They developed a new method employing machine learning techniques to pinpoint crucial factors that enhance the precision of cardiovascular forecasting. The prediction model is introduced using a range of feature combinations and established classification algorithms. The research used machine learning techniques to analyze raw data, resulting in a novel insight into heart illness.

Amin Ul Haq et al. [2] used cross-validation, three feature selection methods, and evaluation criteria for classifier performance including accuracy, sensitivity, specificity, execution time, and Matthews’ correlation coefficient. They used seven established machine learning algorithms. Evaluations were conducted on each classifier to assess accuracy and execution time using all features. The classifiers’ performances were tested by using feature selection approaches including LASSO with k-fold cross-validation, mRMR, and Relief on selected features. Researchers have developed a sophisticated algorithm to classify those with healthy hearts and those with heart disease. Li Yang et al. used many strategies in constructing their prediction model. Consistent follow-up was maintained via the use of an electronic health record system. A three-year risk assessment prediction model was provided based on a substantial population in eastern China with a high risk of cardiovascular disease. The heart disease classifier in Youness Khourdifi et al.’s [4] research was improved by eliminating redundant features with Fast Correlation-Based Feature Selection (FCBF). Subsequently, they categorized everything using several categorization methods.

A disease prediction system was developed by researchers using cloud technology as described in Shadman Nashif et al. [5]. An Arduino microcontroller was used to develop a real-time monitoring system that monitors health indicators such as blood pressure, temperature, heart rate, and humidity. The technology can detect cardiac abnormalities using machine learning methods by transferring recorded data to a central server and updating it every 10 seconds.

P. Suresh et al. [6] assessed several prediction models and feature selection strategies before developing an effective prediction model using a genetic algorithm. It outperforms other traditional prediction methods. The various prediction models were validated using real-time data sets and then reassessed with heart disease data sets. The K-Cross validation technique is used to provide a balanced training and testing data set. In the Fahd Saleh Alotaibi [7] study, researchers used the Rapid Miner tool together with several machine learning approaches to enhance the accuracy score and predict heart disease. We analyzed the UCI heart disease dataset. The recommended task improved the precision score compared to the prior one.

Lewlyn L. R. Rodrigues promoted the use of partial least squares in structural equation modeling for data analysis.
They used machine learning to analyze the correlation between body mass index, age, systolic and diastolic blood pressure, daily cigarette consumption, weekly alcohol intake, and the presence of hypertension and coronary heart disease. All factors, save age, SBP, and BMI, showed a significant positive association with coronary heart disease (CHD) and hypertension. The results assisted machine learning researchers and professionals in identifying relationships among these elements.

In Mohd Ashraf et al.'s [9] study, researchers introduced the Deep Neural Network technique to create an automated system for predicting heart attacks. Various datasets were used to evaluate the precision of machine learning methods. The proposed solution introduced an automated preprocessing method to the data and removed systemic anomalies.

Kathleen H. et al. proposed developing a Deep Neural Network (DNN) classification model and a DNN diagnosis model for heart illness in their research [10]. The classifier achieved an accuracy of 83.67% while using an enhanced Deep Neural Network (DNN).

Researchers presented the Talos Hyper-parameter optimization model for predicting cardiac and heart diseases in Sumit Sharma and Mahesh Parmar [11]. Utilizing deep neural networks is essential for enhancing the accuracy of heart classification in cases of cardiac sickness. Various classification approaches such as SVM, Naive Bayes, and Random Forest showed varying performance. Cardiac arrest at UCI The Talos Hyper-parameter Optimization outperformed the classification algorithms previously discussed in the dataset.

Asma Baccouche and colleagues proposed an ensemble-learning architecture using a CNN and either a bidirectional or unidirectional BiLSTM or BiGRU model, achieving a 91% accuracy in predicting different types of heart disease. Feature selection is used in data preparation to improve the classifier's performance. Nathalie-Sofia Tomov and Stanimire Tomov discovered the HEARO-5, a sophisticated five-layer DNN architecture with the highest predictive accuracy. This model utilizes regularization optimization and can automatically identify missing or abnormal data. They used Matthews correlation coefficient (MCC) and K-way cross-validation to assess the consistency of designs.

Shubhanshi Singhal et al. [14] used Convolutional Neural Networks (CNNs) using 13 clinical variables as input. The improved back propagation training strategy achieved 95% accuracy in predicting heart disease while training the CNN.

Kusuma, S. et al. conducted a comprehensive study on methodologies for heart disease research, focusing on prediction and diagnosis utilizing machine learning and deep learning technologies. SVM and machine learning approaches make for 60% of the methodologies used, with deep learning methods accounting for 30%. The majority of the data used consists of clinical datasets.

Joonmyoung Kwon and colleagues [16] proposed authenticating real-time patient data using deep learning modules. This therapy mainly addressed those with prolonged symptoms of cardiac arrest. This research achieved a classification accuracy of around 82% on a dataset including real-time patient data. Researchers used a language model technique in N. Sowri Raja Pillai et al. [17] to forecast high-risk prognosis from patients' diagnostic histories by using deep recurrent neural networks (RNNs) known as Prognosis Prediction using RNN (PPRNN). The proposed PP-RNN utilizes several RNNs to analyze patient diagnostic code sequences and predict the emergence of high-risk diseases. The recommended technique ultimately resulted in enhanced accuracy.

M. Ganesan and Dr. N. Sivakumar developed an innovative healthcare application using IoT and cloud technology to monitor and identify serious medical conditions. The classifier was trained using data from the benchmark dataset in the training phase. Real patient data were used throughout the testing phase to determine the presence of sickness.

In their article [19], AKM Jahangir and Alam Majumder proposed a multi-sensor system with smart IoT technology to give early warnings of illness risk. The system collects data from the user continuously and transfers it to a smartphone via Bluetooth using a Body Area Sensor (BAS) system. The program enabling users to see real-time charts of possible cardiac arrest completed all processing and data analysis. An IoT system using smartphones was developed with a low power consumption communication architecture to regularly monitor body temperatures and heart rates. Sensor data was examined to provide very accurate forecasts of cardiac arrest using
machine learning and signal processing techniques. A wearable system using an ECG and body temperature combination was developed for heart rate detection, based on a smartphone. Analyze heart rate on Android platform to see body temperature and real-time ECG signal graphs.

Mohm Ayoub Khan et al. proposed an Internet of Things (IoT) architecture for predicting cardiac disease by using a Modified Deep Convolutional Neural Network (MDCNN). Health indicators such as ECG and blood pressure were monitored using the patient's wristwatch and a heart monitor device. The system's performance was evaluated by comparing the suggested technique with existing DNNs. MDCNN outperformed other techniques.

Shadman Nashif et al. proposed accurate prediction of heart diseases using machine learning algorithms in WEKA, a java-based Open Access platform for data mining. The proposed method achieves an accuracy of 97.53% using Support Vector Machine (SVM) and 10-fold cross-validation. An Arduino system was developed to monitor patients in real-time by sensing variables like as heart rate, body temperature, humidity, and blood pressure. If a parameter above the threshold, the patient's live video feed is watched, and a designated physical consultant is notified using GSM technology. A smartphone app is developed to store patient and doctor records. A pre-processing technique was proposed to enhance the classification accuracy of ECG data.

According to Deva Priya Isravel et. al. [22], the noise in the collected raw data reduces the classification accuracy. An innovative pre-processing method is used to remove distorted ECG readings. In order to distinguish between normal and pathological cardiac rhythms, classifier algorithms like KNN, Naive Bayes, and Decision Tree are used to assess the classification performance.

Using ML algorithms, Kumar G. Dinesh. et al. [23] suggested Cardiovascular Disease Prediction. This study suggested a prediction approach for determining if a person has heart disease and for alerting or diagnosing them. Here, the rules are applied to each individual result of Gradient Boosting, Random Forest, Naive Bayes Classifier, Logistic Regression, and SVM to compare the prediction accuracy.

A huge data analysis of coronary artery heart disease is presented in Prerna Jain et al.'s [24] study, "Coronary artery disease prediction on large big data." Hospital management is enhanced by employing DM and ML algorithms to analyze a patient's massive quantity of data. It is challenging to evaluate, collect, manage, and store structured and unstructured data in order to leverage big data technologies and tools as the volume of data expands dramatically in every industry.

Jianliang Gao et al. [25] proposed a technique for predicting node-representation learning similarities in their paper, "Similar Disease Prediction with Heterogeneous Disease Information Networks." First, they use data from many sources to integrate semantic score and topological score among disorders. In order to map each illness into a vector with the same spatial dimension and to compute and thoroughly assess disease similarity, the combined scores of each disease with all other diseases were employed. Finally, they conducted out tests based on benchmarks and additional disease nodes not included in the benchmark set. The efficiency of a technique in forecasting comparable illnesses may be shown by comparing several approaches using statistics like average, variance, and variance coefficient as the standard.

### III. RESEARCH METHODOLOGY

A visual representation of the framework for the proposed system is provided in figure 4.1. It is made up of a total of five primary components, including Medical IoT sensors, a dataset on heart illness, patient information, a cloud database, and an artificial intelligence-based heart disease prediction system. Figure 1 is an illustration of the typical block diagram for the work that is being suggested. It is considered that IoT devices include both those that can be worn and those that can be implanted. It is used to collect patient data from various geographically dispersed areas. These precise measurements are compiled as patient data through the utilization of Internet of Things (IoT) devices that are connected to a person’s body.

The IoT based real time cardiac disease dataset is utilized which is created after observation various patients. The dataset on heart disease contains historical logs of medical data that have been compiled from various hospitals and other medical facilities. The patient records are comprised of the patients; historical medical files, which are gathered from various institutions across the country. The cloud is where each of these datasets has been saved. The necessary information will be stored in the cloud where it will be accessible at any time. The purpose of the heart disease prediction system seems to be to forecast the occurrence of heart illnesses by the application of
classification techniques that are based on machine learning. The IoT-based healthcare approach that has been outlined here functions in three stages. The data will be obtained in the first stage through the utilization of IoT devices from the human body, information from benchmark datasets, and clinical documentation. The subsequent step will involve storing all of the acquired information in a database that is hosted on the cloud. The classification of the data is what happens in the final stage, and its what allows for the prediction of cardiovascular disease. In the beginning, the classification algorithm will carry out the training process. This process will make use of the heart disease dataset in order to educate the classifier on how to determine whether or not heart disease is present. After then, the classifier that has been trained is prepared to test the incoming patient’s information in order to correctly determine whether or not the individual suffering from heart disease. At this point, the user will be presented with the test report after it has been generated.

Figure 1: Proposed System Architecture for HDP

The proposed system architecture for a simulated and real-time IoT data processing environment is shown in Figure 1. Two of the data set's twenty-five properties, the patient's age and gender, are used to identify the patient's specifics. This is because only those characteristics have any bearing on the age and gender of the patient. The 12 characteristics are important since they constitute a necessary part of the patient's medical history. When it comes to detecting cardiac disease and determining how serious it is, clinical data are crucial. We compare the RNN's performance to that of various tried-and-true machine learning classifiers. Machine learning was utilised to create a wide variety of classification models that were used here. In order to identify whether model is superior than the RNN, accuracy calculations are performed using the provided dataset. In order to compare numerous outcomes with different cross validation strategies, we employ the three available activation functions: Relu, Sigmoid, and Tanh.

During the HML feature extraction and feature selection processes, the input, hidden, and feedback layers are defined. When it comes time to choose a single feature vector to use for classification, the SoftMax function is used after all others have been eliminated. The following is a description of the SoftMax function:
\[ y'(n) = \text{softmax}(W(n) h(n) + b(n)) \]

For task process \(n\), the extrapolation alternatives are denoted by \(y(n)\), the required knowledge is denoted by \(W(n)\), and the bias time is denoted by \(b(n)\). Our total cost function is a linear representation of the utility function across all possible crossings.

\[ \phi = \sum_{n=1}^{N} \lambda_n L(y'(n), y(n)) \]

Now, \(\lambda_n\) is the respective weight for specific \((n)\) task.

It's important to remember that labelled data for training each task will originate from totally diverse sources. The lessons are then iteratively taught by repeatedly going through the tasks:

Therefore, following the mutual learning stage, we should use a calibrating strategy to further enhance the output with each challenge. Input gate, decoder (ct), forgetting gate (ft), offset gate (ot), and prior hidden (ht) are the Rd vectors defined as LSTM sub-divisions from each stage \(t\). In LSTMs, \(D\) represents the total number of models. It, ft, and ot are the values in \([0, 1]\) that correspond to the gating dimensions. All of the above are the LSTM transformation coefficients:

\[
\begin{align*}
    i_t &= \sigma(W_i x_t + U_i h_{t-1} + V_i c_{t-1}), \\
    f_t &= \sigma(W_f x_t + U_f h_{t-1} + V_f c_{t-1}), \\
    o_t &= \sigma(W_o x_t + U_o h_{t-1} + V_o c_t), \\
    \tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1}), \\
    c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \\
    h_t &= o_t \odot \tanh(c_t),
\end{align*}
\]

Here, \(x_t\) is the time-step source, and LG (logistic regression) denotes fractal dimension and wise element multiplication. The activation function regulates the internal storage state display, the forgotten gate theoretically decides how much each component of the memory module is gone, and the input gate determines how much the other unit is modified. Here is a rundown of the snippet technique that makes learning easier.

**System modules**

**Data Collection:** We collect a data from various source such as Kaggle, machine learning repository and some synthetic datasets. The IoT is another dataset which we used for real time scenario. The IoT module consist various attributes including numerical and categorial attributes.

**Data pre-processing:** It deals with issues such as noisy data, missing information, and so on. When part of the data in the information is missing, several solutions have been used, such as filling in the blanks or ignoring the tuples, depending on the specific situation. There is a possibility that the data will include null values, which are unintelligible to machines. These noisy data might be the consequence of improper data collection, wrong data entry, or any number of other things. In order to solve the problem, methods such as regression, clustering, and the binary approach are used.

**Feature extraction:** This program will return a wide range of characteristics using the data that is supplied. After the features have been retrieved, a feature selection threshold is applied to them, which removes features that are redundant or otherwise not required for the training process. In order to get a wide range of hybrid attributes, the normalized data with relational features are used, and training is performed by choosing an optimization method. The detailed explanation of the procedures of feature extraction, feature generation, and feature evaluation are explained as follows.

**Feature Extraction** - This dimensionality-reduction technique divides the original information into identifiable categories based on their relationships. These massive datasets are characterized by including several parameters that need significant computer capacity for processing. Therefore, feature extraction may effectively choose certain variables and include important ones, thereby reducing the amount of data. Precision and recall metrics
will be used to evaluate the results. PCA is one of the most used techniques for lowering the number of linear dimensions. It is a technique used for self-directed learning.

**Feature generation** - It involves generating novel attributes from existing ones. Managing large datasets is challenging due to the significant variations in dataset sizes. The feature generation process might be quite beneficial in simplifying the work. We use several mathematical formulae and statistical techniques to enhance accuracy and clarity while preventing the creation of irrelevant features. This strategy usually enhances the model's understanding to enhance its accuracy. Enhancing model correctness may be achieved with this strategy. This strategy focuses on discovering significant connections while excluding insignificant ones.

**Feature evaluation** - Prioritizing features and doing feature assessment are key techniques for completing the assignment in an orderly manner. All features are being evaluated to grade them appropriately and then use them as needed. Avoid the irrelevant ones. Feature assessment is essential to provide an accurate final output from the model by reducing bias and missing data.

**Linear and Non-Linear Feature Extraction** - Feature extraction may be categorized into linear and non-linear methods. PCA is an example of linear feature extraction. Primary characteristics of a dataset are integrated in a normalized linear way to create a main component. Principal Component Analysis (PCA) is a method used to extract important factors from a large range of variables in a dataset. Data is often converted using PCA into a lower-dimensional space in order to optimize the data distribution. Anomalies and outliers, often seen as irrelevant or disruptive data within a dataset, may be used for abnormality and anomaly analysis.

**Feature Selection** : For several purposes, feature selection is the most important phase in creating a better model. One is being limiting the amount of features that can be taken into account while developing a model through feature selection involves some level of cardinality reduction. The majority of the time, data is either inaccurate or includes more data than is necessary to solve the model. We have gathered real-time data of twitter, which has several features but it has few features which doesn't obtain any advantage after adding. There are some redundant columns, and using them might affect the model. Feature selection not only boosts the model's performance, but it also speeds up the modeling process. When creating a model, if unnecessary columns are included, more CPU and storage are needed for training, and more memory space is required for the produced model. Even if resources were not a concern, it is still crucial to do feature selection and choose the optimal features because unnecessary columns might harm the model's performance in a number of ways, including identifying useful patterns in distorted or redundant information is more challenging, and most DM methods need substantially bigger training data sets if the set of data is highly dimensional.

The technique actively chooses or excludes features during the feature selection process depending on how valuable it is for evaluation. Obtaining too much information that is of low value or not enough data that is of great value are two issues that feature selection aids in resolving. Finding the smallest number of data source attributes that are important for creating a model is our aim while choosing features.

**Classification**: The system identifies each record as either an attack or normal using a supervised classification approach. We used several machine learning algorithms, including Support Vector Machine (SVM), for supervised categorization. Supervised machine learning is used to train the classifier.

**Algorithm for Module Training using HML**

<table>
<thead>
<tr>
<th>Input:</th>
<th>Training_DB[] as train dataset, set of activation function AF[].</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1:</td>
<td>Initialize the both methods Train_DB[], AF[], Iteration as epoch_size</td>
</tr>
<tr>
<td>Step 2:</td>
<td>Extract_Feat_Set ← Extract_Feat(Train_DB[])</td>
</tr>
<tr>
<td>Step 3:</td>
<td>Select_Feat_Set [] ← optimization(Extract_Feat_Set)</td>
</tr>
<tr>
<td>Step 4:</td>
<td>Training.pkl ← Build_Classifier(Select_Feat[])</td>
</tr>
<tr>
<td>Step 5:</td>
<td>Return Training.pkl</td>
</tr>
</tbody>
</table>

**Output:** Trained module in .PKL file for whole divided dataset

**Algorithm for Module Testing using HML**

<table>
<thead>
<tr>
<th>Input:</th>
<th>Test_DB [] as test instance set, Train Background Knowledge Training.pkl, threshold Th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1:</td>
<td>Read every test records with the help of following equation</td>
</tr>
</tbody>
</table>
test-feature (m) = \sum_{m=0}^{n} (\text{feature}_{\text{Set}}[A[i] \ldots A[n]} \leftarrow \text{Test_DB})

**Step 2:** Retrieve selected features from whole testing record testFeature(m) by using following function.
Extract_Feat_set[x][t…n] = \sum_{t=1}^{n}(t) \leftarrow test_Feature(m)
The feature vector is the collection of extracted hybrid attributes from given input

**Step 3:** Extract every train instance from trained components by using following function
train_Feature (m) = \sum_{m=1}^{n} (\text{feature}_{\text{Set}}[A[i] \ldots A[n]} \leftarrow \text{Train.pkl})

**Step 4:** Input the testing instances or record set to test classification model as testFeature(m) by using following equation
Extract_Feat_Set_x[t…n] = \sum_{t=1}^{n}(t) \leftarrow test_Feature(m)
The whole class labels' feature vectors are included in the Extract_Feat_Set_x[t].

**Step 5:** Validate individually all test instance with every training features
Calculate weight = Calculate_Sim (Feature_Set_x || \sum_{i=1}^{n} Feature_set_y[y] )

**Step 6:** Return calculate_weight

**Output:** Op_Map <forecasted_class_label, Sim_weight> optimized instance recommends by classification model

---

Table 1: Detail description of Cleveland dataset which is used for classification

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Description</th>
<th>Attribute type</th>
<th>Value / Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age of patient</td>
<td>Numeric</td>
<td>29 to 77</td>
</tr>
<tr>
<td>Sex</td>
<td>Gender</td>
<td>Nominal</td>
<td>0 → Female, 1 → Male</td>
</tr>
<tr>
<td>Cp</td>
<td>Type of chest pain</td>
<td>Nominal</td>
<td>1 → typical angina, 2 → atypical angina, 3 → non-angina pain, 4 → asymptomatic</td>
</tr>
<tr>
<td>Trestbps</td>
<td>BP when admitted in hospital</td>
<td>Numeric</td>
<td>94 to 200</td>
</tr>
<tr>
<td>Chol</td>
<td>Serum cholesterol mg/dl</td>
<td>Numeric</td>
<td>126 to 564</td>
</tr>
<tr>
<td>Fbs</td>
<td>Blood sugar in fasting if &gt; 120 mg/dl</td>
<td>Nominal</td>
<td>0 → False, 1 → True</td>
</tr>
<tr>
<td>Restecg</td>
<td>Result of electrocardiographic test</td>
<td>Nominal</td>
<td>0 → normal, 1 → ST-T wave abnormality, 2 → definite left ventricular hypertrophy by Estes criteria</td>
</tr>
<tr>
<td>Thalach</td>
<td>Heart rate max value</td>
<td>Numeric</td>
<td>70 to 205</td>
</tr>
<tr>
<td>Exang</td>
<td>Exercise induces angina</td>
<td>Nominal</td>
<td>0 → No, 1 → Yes</td>
</tr>
<tr>
<td>Oldpeak</td>
<td>ST depression induced by exercise relative to rest</td>
<td>Numeric</td>
<td>0 to 6.2</td>
</tr>
<tr>
<td>Slope</td>
<td>The slope of the peak exercise ST segment</td>
<td>Nominal</td>
<td>1 → upsloping, 2 → flat, 3 → downsloping</td>
</tr>
<tr>
<td>Ca</td>
<td>Number of major vessels colored by fluoroscopy</td>
<td>Nominal</td>
<td>0 → 4</td>
</tr>
<tr>
<td>Thal</td>
<td>The heart status</td>
<td>Nominal</td>
<td>3 → normal, 6 → fixed defect, 7 → reversible defect</td>
</tr>
<tr>
<td>Target</td>
<td>Prediction attribute</td>
<td>Nominal</td>
<td>0 → no risk of heart disease, 1 to 4 → risk of heart disease</td>
</tr>
</tbody>
</table>
In order to test how well the suggested categorization method works, the dataset would be used. Machine learning repository at UCI has been accessed using the dataset. There is a hierarchy of significance for the qualities, with basic, intermediate, and significant being the most basic levels.

IV. EXPERIMENT DESIGN

In this experiment, the performance of the HDP using a hybrid machine learning model. The IoT real time data set includes records for 480 patients with a variety of parameters, including age, systolic BP, diastolic BP, body temperature, oxygen saturation, pulse rate etc. To increase the performance of the dataset, preprocessing and normalization are applied. The entire data set is split into two sections after the pre-processing of the dataset is finished. One is testing dataset and other is train dataset. The data set is split into three parts in a 3:1 ratio.

After building the hybrid machine learning technique using HDP, comparative study is done with machine learning methods. The following Table 2 displays the accuracy of classification of different ML algorithms utilizing HDP in the Weka framework.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>HML</td>
<td>0.9930</td>
<td>0.982</td>
<td>0.981</td>
<td>0.993</td>
</tr>
<tr>
<td>ANN</td>
<td>0.923</td>
<td>0.913</td>
<td>0.893</td>
<td>0.901</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.892</td>
<td>0.882</td>
<td>0.932</td>
<td>0.912</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.871</td>
<td>0.861</td>
<td>0.921</td>
<td>0.893</td>
</tr>
<tr>
<td>J48</td>
<td>0.903</td>
<td>0.922</td>
<td>0.853</td>
<td>0.881</td>
</tr>
</tbody>
</table>

Consider the following figure 5.2 which depicts the accuracy of prediction of HDP using machine learning techniques.

Figure 2: Accuracy of machine learning algorithm using HML for heart disease detection and classification

Figure 2 shows precision of various machine learning algorithms like HML, Naïve bayes, J48, RF, Artificial neural network is 0.99, 0.91, 0.88, 0.86, 0.92 respectively using various learning algorithm for HDP. Consider the following figure 4.4 which depicts the recall of prediction of HDP using machine learning. As a result, it can be concluded from the experimental results that support vector machines outperform other machine classifiers for effective detection of heart disease.
V. CONCLUSION

In the experiment of forecasting HDP using hybrid deep learning classification model, several DL classifiers like artificial neural network, deep neural network and recurrent neural network were used. From the experimental findings, we observed that among all conventional machine learning technique, recurrent neural network obtains highest classification accuracy of 95.4%. In the last experiment which is proposed methodology of HDP using HML classifier, we have used the real-time huge IoT dataset. Data has been pre-processed using data filtration and noise removal using sum rule method. On this filtered data various feature extraction techniques like N-gram, correlation c-occurrence, bi-gram, tri-gram dependency features were applied to obtain the best classification accuracy. We performed three experiments using HML, tan h and sigmoid function and observed that HML function obtained best accuracy of 99.30% for 15-fold cross validation as compared to Tan h and sigmoid function. Thus, our proposed method performs better as compared to traditional learning techniques used for predicting the academic performance of HDP. In future work improvement in research since the majority of the indicators that are beneficial for determining the condition of the heart artery stiffness are not taken into account. Some of these factors include augmentation index, arterial stiffness, and augmentation pressure.

REFERENCES


