International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2390222

Using an Adapted Hybrid Intelligent Framework to Make Predictions Regarding Heart Diseases

Mr. Sumit Kumar Soni[1], Dr. Kalpana Rai[2], Dr. Harsh Mathur[3]
[1]Research Scholar, Madhyanchal Professional University, Bhopal, Madhya Pradesh, India
[2]Research Guide, Madhyanchal Professional University, Bhopal, Madhya Pradesh, India
[3]Associate Professor, Rabindra Nath Tagore University, Bhopal, Madhya Pradesh, India

ARTICLEINFO

Article History:

Accepted: 01 April 2023
Published: 07 April 2023

Publication Issue
Volume 10, Issue 2
March-April-2023

Page Number
214-232

ABSTRACT

The effects of heart disease on a person's life can be devastating, making it one of the world's most serious health problems. Patients with heart disease have a compromised ability of the heart to pump blood throughout the entire body. A proper and prompt diagnosis of cardiac disease is the first step in preventing and treating heart failure. Diagnosing heart illness has a long history of being fraught with difficulty. Machine learning-based noninvasive technology can accurately and quickly distinguish between healthy people and those with heart disease. In the proposed research, we used heart illness datasets to develop a machine-learning-based detection system for predicting cardiovascular disease. In order to measure the efficacy of our machine learning algorithms, feature selection algorithms, and classifiers in terms of metrics like accuracy and specificity, we employed cross-validation. Our method allows for quick and easy differentiation between those with heart illness and healthy people. Analysis of the receiver optimistic curves and area under the curves for each classifier was performed. Classifiers, feature selection algorithms, preprocessing techniques, validation strategies, and performance metrics for classifiers have all been discussed in this work. The performance of the suggested system has been evaluated using both the full set of features and a subset. The results include a comparison of recall, F1 score, and false positive rate. Decreases in the number of features used to make a classification have a notable effect on both the classifier's accuracy and the time it takes to run. The anticipated machine-learning-based decision support system would help doctors make more precise diagnoses of cardiac illness.

Keywords : Heart Diseases, Classifiers, Feature Selection Algorithms, Preprocessing Techniques

I. INTRODUCTION

Many studies have focused on cardiovascular disease because it is a major killer. Automated cardiac prognosis improves the odds of a successful treatment, which is a major task. The severity of a patient's condition is evaluated in light of signs and symptoms of cardiovascular disease. Cardiovascular disease is more likely to occur in people who already have a predisposition for it due to variables including smoking, obesity, and hypertension.

Healthcare facilities, such as hospitals, are concerned with providing effective treatment at a reasonable price.

1 In order to provide great care, it is crucial to provide an accurate diagnosis and administer the appropriate treatment. Data about cardiac disease might be numerical or categorised. It is necessary to clean and filter these entries first, as doing so will remove any unnecessary information from the database and allow for more efficient processing.

2 It is possible to unearth previously unknown information by analysing a database of past cases of heart disease for patterns and associations. Further, it can help doctors make more informed decisions in the clinic by answering difficult diagnostic questions about heart disease. This approach proved to be highly efficient at accomplishing the predetermined mining goal

Confusion Matrix

The selection of a performance metric is often determined by the nature of the business problem being solved. Let's pretend you have 100 examples that you fed into your model and received a categorization for. The discrepancy between the predicted and observed labels can be displayed graphically in a table called a confusion matrix.

The above table displays the findings, both positive and bad. These two results define the "classes" to which these examples belong. Due to there being just two possible categories, the model used to generate the confusion matrix is a binary classifier. The use of a binary classifier in the spam detection process is illustrative. Similarly to how everything in life is either a hot dog or not, all emails are either spam or not spam. To make sense of the data, we compute several metrics, including F1, accuracy, and False positive rate.

Accuracy

An immediate indicator of whether or not a model is being trained appropriately and of its potential performance in the wild is its accuracy. However, it provides no specifics on how to apply it to the issue at hand.

Accuracy as the primary performance metric has issues when there is a large disparity between classes. We'll use the data from the aforementioned confusion matrix. Supposing the negatives represent legitimate purchases and the positives represent fraudulent ones. You are correct in all classes 99 percent of the time, according to accuracy.

However, we can see that you are only right 50% of the time for the fraud class (positive), which indicates you will lose money. You'd be right 98% of the time if you made a strict rule stating that all transactions were typical. However, that is not a very good model or criterion for evaluation. Therefore, you may respond, "It's complicated," when asked by your superior, "How accurate is that model?"

Knowledge of precision, recall, and f1 scores will aid in providing a more complete response. Find Out How to Use AI in Games »

Precision

What is the success rate of the model when it makes a positive prediction?

Accuracy is especially important when the cost of getting something wrong is high. Just imagine a world where it's difficult to spot skin cancer in its early stages. Low-precision models will lead to incorrect diagnoses of melanoma in many patients. There will be more checks and worries. When there are too many false positives, the people responsible for monitoring the results will become accustomed to disregarding them.

Recall

Recall is useful when the cost of false negatives is high. Can oncoming nuclear missiles be detected and avoided? A false negative has potentially catastrophic results. The slightest misstep will be our final one. To be hit by that which you are attempting to avoid is a direct result of a high number of false negatives. A false negative would be to ignore the snap of a twig in the night and risk getting eaten by a bear. A chipmunk is the source of the sounds you heard while laying awake in your tent, perspiring and listening to every movement in the forest. It's not fun being here. You'd want to get rid of your model if it had a flaw that allowed nuclear missiles through. You would also get rid of a model if chipmunk noises kept you up all night. If you want to avoid being eaten by a bear or staying up all night thinking about chipmunk alarms, you should aim for an evaluation metric that assesses precision and recall together. The F1 score entry box is open.

F1 Score

In order to get the F1 score, precision and recall are added together and multiplied, much like when cooking. You can accurately identify major threats without being bothered by false alarms because to the low rates of both false positives and false negatives. If the F1 score is 1, the model is perfect; if it's 0, the model is a total bust.

Remember that while every model has its flaws, some of them may prove useful in certain situations. To rephrase, every model will have some erroneous positive results and some erroneous negative results. Even while a model can be fine-tuned to eliminate either false negatives or false positives, you'll typically have to pick one or the other. It's important to zero in on the metrics that mean the most to you when trying to solve a problem.

By extension, recall measures the proportion of true positives that were retrieved (found), or the proportion of accurate hits that were unearthed. How many of the results returned were actually right answers is what you mean by "precision" (your formula is erroneous).

Once again, it is that time of year!There is encouraging evidence that more and more authors are mixing these two approaches in their work.

The filter method is one of many approaches that may be used to reduce the number of features that make up a product. Before employing any kind of learning procedure, variables are filtered in order of importance. In doing so, it determines which characteristics are the most important overall. Without a classifier to rely on, it generates a large variety of potential outcomes for the forecast. The use of these techniques makes for more rapid and effective implementation. As a result, wrapper methods are favoured when dealing with massive datasets. One of its drawbacks is that it may fail to capture the most valuable characteristics since it does not account for the interaction between classifiers.

It is common knowledge that a huge list of features can be whittled down using the filter strategy. Before any sort of learning method is carried out, variables are picked using a filter strategy. In order to determine what qualities are most important, it uses an assessment rubric. Because the classifier is independent of its implementation, it cannot be relied upon to produce reliable results in foresight. These methods allow for quicker and more effective outcomes. This makes them more appropriate for extremely large data sets. The difficulty is that they fail to account for the interdependence of numerous characteristics, which can cause them to overlook the most "helpful" characteristics.

An internal feature selection mechanism used by the learning process frequently directs the search. The "usefulness" of feature subsets is evaluated during the learning process using the nested subset method. The efficiency of the learning algorithm can be improved with the use of a special learning algorithm. By not separating the data into training and validation sets, this technique makes better use of the available data and yields a solution more quickly. Wrapper methods increase the likelihood of overfitting since they need more processing power. Wrapper methods are more complex computationally than embedding approaches. There is a fundamental flaw in this approach since it puts all the decision-making weight on the classifier. If the classifier is used with a different classifier, the predicted properties may change.

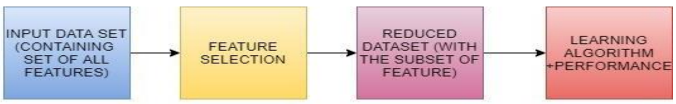

Figure 1 : Data set is input with all features, then features are picked, reduced with subset of characteristics, and Learning Algorithm is used on it. This is what Feature Selection depicts.

II. Analysis of the Literature

Among the leading causes of death today, S. Mohan, C. Thirumalai, and G. Srivastava all point to cardiovascular disease. Clinical data analysis for the prediction of cardiovascular disease is exceedingly challenging. Machine learning (ML) can be used to make inferences and predictions from the mountains of data collected by healthcare institutions. There have been a lot of recent improvements in the IoT thanks to the implementation of machine learning methods (IoT). Only a small number of research have demonstrated the potential of ML for predicting cardiovascular disease [1]. We've come up with a new approach to using machine learning to discover crucial characteristics that boost prediction precision for cardiovascular disease. Several feature sets and standard methods of categorization are used to build the prediction model. Researchers have shown that a hybrid random forest-linear model can better predict cardiovascular illness (HRFLM).

This report asserts that heart disease ranks among the most complicated illnesses due to the sheer number of people it affects and the variety of approaches needed to treat it. For cardiologists to be successful, they need to be able to diagnose heart disease quickly and accurately. In this article, we use machine learning techniques to provide a reliable and efficient method for identifying cardiac problems. [2] Techniques of classification are used in the system's construction. Support vector machines, logistic regression, artificial neural networks, K-nearest neighbours, and Nave bays are just a few of the algorithms that have been used to eliminate superfluous or irrelevant features. Other algorithms include Relief, Minimal redundancy, Maximal relevance, Least absolute shrinkage, and local learning. To address the issue of feature selection, we devised a novel conditional mutual information feature selection technique. In order to improve classification accuracy and decrease classification system execution time, feature selection algorithms are employed to choose the best features. The best strategies for model evaluation and hyperparameter tweaking have been found using subject-free cross-validation. The effectiveness of the classifiers can then be assessed using the metric for measuring performance. Features were selected using feature selection approaches, and the classifier's performance was evaluated. The proposed feature selection technique (FCMIM) and the resulting high-level intelligent system for diagnosing heart disease were implemented using a support vector machine (SVM) classifier. Results showed that the FCMIM-SVM diagnostic strategy outperformed other methods suggested at the time (FCMIM-SVM). The detection of coronary artery disease with this approach is also rather straightforward, making it useful in healthcare settings.

The authors of this research, H. Wang, Z. Huang, D. Zhang, J. Arief, T. Lyu, and J. Tian, argue that successful treatment of Kawasaki disease hinges on prompt and accurate identification of intravenous immunoglobulin-resistant individuals. Data-driven methods can use the intricate structures present in real-world data to identify potentially dangerous people. To overcome the limitations of limited clinical data and the interpretability gap in machine learning models, a multi-stage approach is used to predict intravenous immunoglobulin resistance in clinical settings. The clustering of patients and clinical variables enables (a) feature selection based on the availability of data and (b) patient subgroup-specific prediction models. Finding unique risk factors for a selected group is another common application of Group Lasso. For the third step, we use the Explainable Boosting Machine, a generalised additive learning approach, to predict each patient subgroup. The suggested framework for predictive modelling outperformed a set of benchmark approaches in trials that used real-world Electronic Health Record data. Through highlighting the integration of various methodologies via the use of co-clustering and supervised learning methods, this study advocates for data-driven approaches to develop predictors and effective algorithms for healthcare decision-making.

Co-authored by Fitriyani, Syafrudin, Alfian, and Rhee, this research suggests that diagnosing cardiac disease early can improve patients' health and length of life. With the help of a clinical decision support system (CDSS), heart disease can be detected at an early stage. In this paper, we introduce a powerful HDPM (Heart Disease Prediction Model) for a CDSS by using DBSCAN to detect and remove outliers, a hybrid Synthetic Minority Oversampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to equalise the distribution of training data, and XGBoost to predict heart disease. [4] The model was constructed using a number of different types of models, such as naïve bayes (NB), logistic regression (LR), multilayer perceptrons (MLP), support vector machines (SVM), and decision trees (DT). Existing models and previous study findings were surpassed by the suggested model, which achieved accuracies of 95.90 percent and 98.40 percent on the Statlog and Cleveland datasets, respectively. The Heart Disease Clinical Decision Support System (HDCDSS) prototype can be used by doctors and medical professionals to assess a subject's or patient's risk for developing heart disease. Early detection and treatment of heart disease could save many lives.

According to the research of P. Ghosh et al., cardiovascular disease (CVD) is one of the world's leading causes of death and disability. In the case of cardiovascular disease (CVD), this can be accomplished by early diagnosis and treatment. Machine learning techniques can be used to pinpoint potential danger factors. Better heart disease prediction requires more complex models, and we'd like to offer one. Our proposed model is able to provide accurate results because it was trained using quick data collection, data pre-processing, and data transformation techniques. For this analysis, we used a single dataset (Cleveland, Long Beach VA, Switzerland, Hungarian and Stat log). Relief and the Least Absolute Shrinkage and Selection Operator are two techniques used for feature selection (LASSO). Classical classifiers are combined with bagging and boosting methods to produce hybrid classifiers like the Decision Tree Bagging Method (DTBM), Random Forest Bagging Method (RFBM), K-Nearest Neighbors Bagging Method (KNNBM), AdaBoost Boosting Method (ABBM), and Gradient Boosting Boosting

Method (GBBM) (GBBM). [5] Our model's FPR and FN, along with its accuracy, sensitivity, error rate, precision, and F1 score (F1), are all determined via machine learning techniques (FNR). In order to facilitate comparisons, the outcomes are presented in several distinct columns. Based on the data, our proposed model performed best when combining RFBM and Relief feature selection techniques (99.05 percent).

According to this research by M. A. Khan and F. Algarni, IoT can be useful in many different fields. The IoMT has expanded as a result of the widespread implementation of wearable devices into the healthcare monitoring infrastructure (IoMT). IoMT has a significant effect on death rates due to its ability to detect disease at an early stage. [6] Predicting cardiac issues from medical records is crucial work. Using machine learning, we seek to identify the most important factors in predicting cardiovascular disease. There have been a lot of attempts to figure out how to properly identify heart disease. Researchers have developed an IoMT framework that combines MSSO with an adaptive neuro-fuzzy inference system with the aim of increasing prediction accuracy (ANFIS). When used in conjunction with the Levy flying strategy, the MSSO-search ANFIS's capabilities are greatly enhanced. Gradient-based learning can get trapped in local minima when utilising ANFIS, which significantly affects the typical learning process. The performance of ANFIS can be improved by using MSSO to fine-tune the parameters of its learning process. Information from medical records, such as blood pressure (BP), age, gender, and chest discomfort, can be used to predict the risk of heart disease. Using MSSO-ANFIS, we can classify sensor data and determine heart state. MSSA-ANFIS has been shown to reliably predict disease through simulation and analysis. The simulation results show that the MSSO-ANFIS prediction model outperforms the alternatives. The proposed MSSO-ANFIS prediction model has a higher accuracy (96.54) than the alternatives.

Heart disease, according to this study, is the leading cause of mortality in the United States. Predicting cardiovascular illness is a significant challenge in the field of clinical data analysis. To efficiently determine and forecast a significant amount of data offered by the health business, researchers have found that deep learning (DL) is a useful tool. In this research, heart disease detection is accomplished using Recursion Enhanced Random Forest with Improved Linear Model (RFRF-ILM).

Heart disease is the leading cause of death in China, according to research published by C. Guo, J. Zhang, Z. Han, Y. Xie, and J. Yu. Predicting cardiovascular disease using clinical data presents a significant challenge. The healthcare industry has provided data that deep learning (DL) has successfully analysed and forecasted. In this research, heart illness detection was accomplished using an RFRF-ILM (Recursion Enhanced Random Forest with an Improved Linear Model). To better forecast cardiovascular illness, this article employs machine learning techniques to identify the most important factors. Multiple features and tried-and-true classification techniques are incorporated into the prediction model. The heart disease prediction model comes highly suggested because of its efficacy and precision [7]. Potential causes of cardiovascular disease are discovered in this investigation. The IoMT platform analysed the data by contrasting crucial components. This supports the idea that there is an age-related increase in the likelihood of developing coronary artery disease. One theory puts the spread of the disease attributable to hypertension. Prevention is key, because diabetes adds another risk factor for coronary heart disease, with a 96.6 percent accuracy ratio, a 96.7 percent stability ratio, and a 95.6 percent F-measure ratio. Type 2 diabetes needs to be considered.

Joint Work by J. Zhang et al. Intelligent technology has recently been used to reduce healthcare expenditures and burdens while simultaneously improving patients' quality of life. Using a rapid

Fourier transformation as part of a machine learning ensemble model makes it simpler to forecast short-term illness risk in patients with chronic heart disease. Using the patient's medical history, the model can then decide whether or not the patient needs to be tested the following day. From patient time series data, frequency information can be extracted using the fast Fourier transform [8]. The future state of the patient is predicted using ensemble modelling, and a final suggestion is reached based on this information. An ensemble framework is created using three different classifiers (artificial neural network, least squares support vector machine, and naive bayes). Researchers can use real-time telehealth data from patients with chronic cardiac disease. Preliminary research shows that the suggested approach is effective, and that daily body checks can lower the risk of erroneous suggestions and ease patients' responsibilities. These results imply that the proposed approach has the capacity to analyse medical time series data and provide accurate, trustworthy recommendations for individuals with chronic heart diseases.

Artificial intelligence (AI) and massive amounts of data are required for the development of a reliable cardiovascular disease prediction model (see also G. Joo, J. Park, H. Im, and Y. Song) (CVD). Despite the fact that race and ethnicity are strong indicators of CVD risk, most previous research has primarily focused at populations in the United States and Europe. To complement existing research, we analysed the Korean National Health Insurance Service-National Health Sample Cohort to compare the predictive abilities of ML and big data for CVD risk (KNHSC). Machine learning allows for the prediction of CVD, including atrial fibrillation, heart failure, coronary artery disease, and strokes (ML). [9] Prediction models were developed using KNHSC data on co-morbidities, medication use, and the outcomes of routine medical examinations. Logistic regression, deep neural networks, random forests, and Light generative adversarial networks (GBM) were utilised to develop ML-based prediction models that were then evaluated using a variety of criteria. Multiple studies have found that ML models used to evaluate 10-year CVD risk are superior to the standard method created by the ACC/AHA. In addition, whether or not we used historical drug information as a feature, the ML models in our investigation performed similarly well in terms of accuracy. All prediction models fared better when doctor medication utilisation data was added.

S.J. Pasha and E.S. Mohamed will give a presentation together. Machine learning (ML) and data mining (DM) are used in today's healthcare systems to derive useful information from massive amounts of raw data. Repeated research demonstrates that medical professionals make incorrect diagnoses at a rate of about 12% of the time. When evaluating the precision of disease risk prediction, AUC accuracy is increasingly being considered by medical researchers. So far as we're aware, the AUC's function has never been established. By coordinating ML and DM algorithms, NFR[10] is a revolutionary feature reduction model built to improve accuracy and efficiency. Accurate and reliable illness risk prediction can be made by combining the AUC and accuracy. In the first approach, we evaluate the performance of the prediction by reducing the number of features using the area under the curve (AUC) and the improvement in accuracy as assessment measures. After taking into account all of the features, the ones with the highest AUCs and the lowest amount of variation are selected as the most relevant ones to employ for classification. For this reason, we utilise publicly available cardiac data from UC Irvine (UCI), and our analysis reveals promising signs of progress. The proposed NFR model achieves the highest levels of accuracy and AUC while having 41.67 percent less features (95.52 percent and 99.20 percent, respectively). The algorithm is around 4.22

percentage points more precise than its predecessors, and it also runs 25% faster.

The authors of this work are S. A. Ali and colleagues. Poor diet, emotional stress, genetic predisposition, and inactivity all likely contribute to the rising rates of heart disease. Recent studies have proposed a number of different automated diagnosis methods for the prediction of heart disease, with the majority of these studies concentrating on improving the accuracy of features prior to processing, selection, or merging. Overfitting and underfitting are avoided, and network setup and optimization issues are addressed, among other things, in this comprehensive investigation. OCI-DBN-based deep belief networks are intended to improve the system's performance by solving these issues. [11]
With Ruzzo-Tompa, we were able to boost system performance by deleting unused parts. Our approach combines two genetic algorithms to get the best DBN possible. As a means of gaining insight into the system, we examine an RBM and DBN training. Accuracy, sensitivity, specificity, precision, F1 score, and Matthew's correlation coefficient were the six metrics utilised to evaluate the proposed approach (Mathematics). Experimental results show that OCI-DBN performs better than state-of-the-art alternatives. The validation study showed that using the proposed method improved the accuracy of heart disease predictions by 94.61%.

All three of us (C. Xiao, Y. Li, and Y. Jiang) had posters up at the conference. This method, which is based on a refined three-dimensional U-net convolutional neural network deep learning algorithm for predicting sickness risk, may be used with a wide variety of data sets, either without or with the centerline present. Extraction of features from ventricular data was performed using a deep belief network, and novel local features were also implemented. Tissue in both ventricles might be regressed using these attributes as inputs. Regression networks that combine cheap computation with deep neural networks can accomplish this. Both datasets' segmentation results were compared using the dice coefficient. It is the pre-processing of centerlines, rather than the raw data, that has the most impact on model training[12]. Our data suggested that a coefficient of 0.8291 on the dice was optimal. The extended test data had a negative impact on the experiment's results. The curve predicts that as coronary artery segmentation accuracy rises, so too will the quality of training data. This is great news for both doctors and patients because it will lead to more precise diagnosis and recommendations. Intended for the benefit of working professionals

The authors of this study note that death from heart failure is widely recognised as a severe public health problem worldwide. Despite advances in medical technology, diagnosing heart failure remains challenging even in developed countries. Therefore, several high-tech technologies have been created to detect the symptoms of failing hearts. Recent methods tend to suffer from overfitting, increasing heart failure detection accuracy on testing data while decreasing it in training. It's called "heart overfitting" when this happens in the medical field. This causes the generated models to inappropriately fit the validation data. Our findings allow for the creation of a diagnostic system that can examine both test and training data. Instead of using a traditional linear regression model, the proposed diagnostic method employs RSA and a random forest model to select features and predict the likelihood of heart failure. Utilizing a grid search, the proposed diagnostic method is optimised. In order to determine how accurate the proposed method is, it is tested in two separate settings. Both a standard random forest model and an RSA-based random forest model are built in the first experiment. The investigations leverage the Cleveland dataset, a publicly available online data collection dedicated to the study of heart failure. [13] Accuracy is improved by 3.5% when using this method compared to using a random forest model with only 7 features. Furthermore, it outperformed five other state-of-the-art machine learning methods. We enhanced the training accuracy to 93.33 percent after achieving a classification accuracy of that level. Eleven other methods for heart failure detection have been proposed, but the current method has been shown to be the most effective.

According to A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, the prognosis for cardiovascular and cerebrovascular events in hypertensive patients is essential for halting the development of heart illness. The current clinical gold standard techniques for predicting risks of vascular events continue to have shortcomings. We assess the potential of a machine learning strategy based on HRV to identify hypertensive persons at increased risk of vascular events. HRV features were extracted from all patients' datasets using time-domain, frequency-domain, non-linear, and fragmentation metrics, and the 24-hour cycle study was split into four segments: late night, early morning, afternoon, and evening. The features were gleaned through this analysis. Each factor was analysed using a period-by-period one-way analysis of variance (ANOVA). Chi-square analysis was used to determine the most advantageous features to use for both demographics and HRV. Then, we created the RUSBOOST model, which combines decision trees with random under-sampling boosting (RUSBOOST). During the afternoon, the maximum accuracy of the trained model reached 97.08 percent while using all available features. Patients at high risk were identified with an F1-score of 86.67 and a precision of 81.25. The model had a good degree of sensitivity and specificity, as measured by the AUC of 0.98. Hypertensive individuals can now have their vascular events predicted using heart rate variability and machine learning. It provides a straightforward method for doctors to implement that nonetheless produces accurate and reliable forecasts in comparison to alternative methods.

Approximately 50% of them will perish after five years, as stated by L. Ali et al. Machine learning-based methods have been developed for HF detection and to improve cardiologist diagnosis in recent years. High fructose corn syrup predictions are described in detail in a paper[15] that details an expert system that employs two SVM models. Due to its linearity and L1-regularity, this SVM model is the most basic. This method can be used to reduce to zero the coefficients of features that aren't important. L2 regularisation is applied to the second SVM's residuals to make them more manageable. It has widespread application in the field of predictive modelling. Using a hybrid grid search strategy, we can optimise both models in parallel (HGSA). Some measures used to assess a system's efficacy include accuracy, sensitivity, specificity, the Matthews correlation coefficient, receiver operating characteristic (ROC) plots, and area under the curve (AUC). These indicators can be used to gauge the success of the suggested approach (AUC). Applying the proposed method to SVM models yields a 3.3% increase in performance. Eight different methods have been tried before, with results ranging from 57% to 91% accuracy. When compared to conventional approaches, the new strategy performs far better. This method is so much better than the alternatives that it can be applied to machine learning ensemble models.

There have been previous research that have built ANN-based automated decision support systems for the diagnosis of heart disease (ANN). To the contrary, the vast majority of these methods merely permit feature preprocessing. This research not only focuses on fixing common issues with predictive models like underfitting and overfitting, but also on improving feature reduction. The model's performance on both the training and testing datasets improves when overfitting and underfitting are avoided. Incorrect network configuration or extraneous features could lead to overfitting the training data. Combining a thorough search strategy with a 2-statistical model, we can filter out extraneous characteristics (DNN). To evaluate the 2-DNN model, we compared its performance to that of other state-of-the-art machine learning models and previously published methods for the prediction of heart disease. The proposed model has a high rate of success (93.33%). The outcomes look good compared to competing methods. The study's findings suggest that doctors can accurately predict their patients' risk of heart disease by using the proposed diagnostic approach.

According to research by J. Wang and coworkers, invasive coronary arteriography (CAG) is highly accurate for detecting coronary heart disease (CHD). Invasiveness of the annual physical exam precludes CHD detection. Motivated by the widespread acceptance of ML across many fields, we intend to put to the test the efficacy of feature selection techniques utilising individual clinical data typically collected during routine annual checks. In this investigation, we stacked two levels on top of one another: level 1 was the basis, and level 2 was the meta-level. Meta-level classifiers are fed predictions from lower-level classifiers. At first, we calculate the Pearson correlation coefficient and the maximum information coefficient to see which classifier has the lowest correlation. After then, it's put to use to find out which classifier combinations produce the best results. There are 303 CAG-verified cases in the Z-Alizadeh Sani coronary heart disease dataset. The experimental results show that the proposed model can identify CHD with a 95.43 percent accuracy, a sensitivity of 95.84 percent, and a specificity of 94.44 percent. New diagnostic tools now available to cardiologists allow for accurate differentiation between patients with healthy coronary arteries and those with coronary heart disease.

Patients with arrhythmias are more likely to have sudden cardiac death (SCD) than patients without arrhythmias, however the authors of this research believe that early detection and treatment of arrhythmias can boost survival rates in the event of an unexpected SCD. Accordingly, in this research, we introduce a computer-based strategy for reliably forecasting sudden cardiac death (SCD) from ECG parameters. Conduction and repolarization time markers, as well as repolarization interval ratios (such the ratio of the conduction-to-repolarization time (TpTe) to the QT interval), are investigated in the context of cardiac arrhythmia. The QRS and T waves of an ECG are used to determine each one (ECG). After the markers have been created, the information is automatically classified into normal and SCD risk categories using machine learning classifiers such k-nearest neighbour (KNN), DT, Naive Bayes, SVM, and random forest (RF). The proposed method is tested using an electronic heart-rate (ECG) database consisting of readings from 28 people with SCD and 18 healthy controls. [18] In less than a second, the automated technique can predict SCD using the set of five arrhythmic risk markers with an average accuracy of 98.91 percent (KNN), 99.99 percent (SVM), 97.46 percent (NB), and 99.49 percent (RF) in the 30 minutes before to the occurrence of SCD. Furthermore, a simple and useful SCD index is proposed to be created using the student's t-test (SCDI) (SCDI). The difference between healthy individuals' and SCD patients' SCDIs is 1.2058 0.0795, with a p-value of 6.5061 10-35. Up to 30 minutes in advance, SCD can be predicted by both the automated classifier and the integrated SCDI. These predictions may be more practical and efficient if implemented on portable smart devices with real-time requirements, such as those found in hospitals or homes.

In this article, researchers including A. Mdhaffar, I. Bouassida Rodrguez, K. Charfi, L. Abid, and B. Freisleben describe a novel approach to predicting cardiac failure. Statistical methods and complex event processing (CEP) technology are used in this system. One way in which a CEP engine analyses health data is by running threshold analysis algorithms. An automated statistical method has eliminated the need for human intervention in the threshold-setting process. In terms of speed, accuracy, and memory recall, our method is superior.

Heart disease is the main cause of death and disability in the United States, as reported by A. Ishaq et al. Survival prediction in patients with heart disease is a significant challenge in clinical data analytics. Huge amounts of data are produced in the healthcare industry, which can be leveraged for decision improvement. The effectiveness of machine learning models has been demonstrated to depend critically on a variety of factors. Heart failure patients who managed to avoid hospitalisation are the primary focus of this study of 299 patients. The goal of this research is to utilise data mining methods in order to better predict a patient's chance of survival from cardiovascular disease. Decision Tree (DT), AdaBoost, Logistic Regression, and Random Forest are just some of the classification models used here (SGD). Other classifiers include the Gradient Boosting classifier, the Extra Tree Classifier (ETC), and the Gaussian Naive Bayes (G-NB) (SVM). The problem of underrepresentation can be addressed with the help of a technique called SMOTE (Synthetic Minority Oversampling Technique)[20]. In order to properly prepare machine learning models for training, RF isolates the most crucial details. Here, we evaluate the results of this experiment against those produced by a machine learning algorithm. Test results show that ETC is superior to other models in predicting heart patients' survival, and reaches an accuracy value of 0.9262 with SMOTE.

Researchers M. Alkhodari, H. F. Jelinek, N. Werghi, L. J. Hadjileontiadis, and A. H. Khandoker conducted a study to identify the optimal fit for measuring LVEF from 24-hour ECG recordings and to compare it to the optimal fit found by two gold-standard guidelines. Methods: The Intercity Digital ECG Alliance (IDEAL) project used support vector regression (SVR) models derived from heart rate variability (HRV) data to predict LVEF in patients with preserved, mid-range, or reduced LVEF. The most accurate estimates of LVEF were obtained by a step-by-step process of feature selection. By doing many studies, it was shown that the best RMSE values occurred between 3 and 4 am, 5 and 6 am, and 7 and 8 pm (RMSE). Our findings suggest these windows of opportunity may be optimal for therapeutic intervention. [21] When evaluating LVEF in the mid-range, ACCF/AHA recommendations are more accurate. Future research utilising HRV features to estimate LVEF percentages as an early predictor of disease progression for CAD patients can build upon the foundation laid by this work.

Evidence suggests that when a person is found to be afflicted by any illness or health concern, a diagnosis is made. The diagnosis of an illness might be simple in some cases and complex in others. Tools that can discover patterns and predict the future in large volumes of data are limited. The old methods of disease severity assessment, which take too much time and are prone to mistakes, must be replaced. Predictive methods based on artificial intelligence (AI) can be used to achieve self-diagnosis and lower detection error rates. We've compiled a database of research published on the topic between January 2009 and December 2019. More than a hundred articles were found in the eight most used databases, the study found. Following a thorough literature review [22], we now know which AI methods are most commonly used in clinical diagnosis. Fuzzy Logic, Machine Learning, and Deep Learning are just few of the AI methods discussed, along with a number of different disorders. The study's overarching goal is to shed light on the current and historical applications of artificial intelligence (AI) in medical research, with a focus on their use in the prediction of cardiovascular illness and brain disease. The research also highlights some challenges and gaps in the development of AI-based diagnostic tools.

Myocardial infarction is the leading cause of death in many industrialised countries, as reported by D. Tay, C. L. Poh, E. Van Reeth, and R. I. Kitney (MI). Lifesaving preventive measures could be implemented if MI episodes could be identified sooner. Risk prediction algorithms based on machine learning allow enable early diagnosis of disease. Risk prediction models should be flexible enough for clinicians to use the one they think would work best for their patients (for example, by changing the prediction range). [23] Because of this, we develop models for predicting the risk of MI and assess how factors like sample age and prediction resolution affect these models' accuracy. Therefore, this investigation relied on data from the cardiovascular health study. The SVM algorithm-based prediction model has a sensitivity of 95.3%, a specificity of 84.8%, and a balanced accuracy of 90.1% during a 6-year period. Research involving patients aged 65 and up found that MI risk prediction models performed similarly regardless of sample age or prediction resolution. Because of this, risk prediction models may be built with a wide range of sample ages and prediction resolutions. Incorporating the many clinical assessments needed by these models into a computer-aided screening tool is a useful way to determine an individual patient's risk of MI (such as tests of cognitive function or physical function or electrocardiography or changes in general health or lifestyle). It's possible that this approach might be used to identify persons at risk of getting MI in the near future and prescribe early medical intervention.

According to W. Chang, Y. Liu, X. Wu, Y. Xiao, S. Zhou, and W. Cao, the rise in hypertension cases over the past few decades can be traced back to modifications in lifestyle and rising material standards. Since the prevalence of hypertensive heart disease has been on the rise, there is an international emergency in terms of making correct predictions about the condition. This research employed the XGBSVM hybrid model to provide hybrid methods for the 3-year prediction of hypertensive heart disease. This approach allows hypertension patients to learn their three-year risk of developing heart disease, allowing for targeted preventative medicine and a lessening of the impact on the patients' psychological, physiological, and monetary health. This study, for instance, proves that machine learning can be used fruitfully to the biomedical sector, where it would have both practical and theoretical significance.

Researchers R. Ferdousi, Mamdouh A. Hossain, and A. E. Saddik propose integrating computational and communication capabilities into the fundamentals of Cyber-Physical Systems (CPS) to enable remote monitoring and control of physical processes. Health CPS is a subset of CPS that utilises the Internet of Things to collect, process, and analyse real-time data from health sensors. Patients with or at risk for acquiring non-communicable diseases can benefit from the use of these tools (NCDs). Only by embedding AI methods into the heart of health CPS will NCDs like heart disease and diabetes be detectable. The introduction of machine learning into CPS is proving useful for the classification and detection, monitoring, and prediction of a wide range of noncommunicable diseases. It's a problem that people don't realise there can be serious consequences to making premature diagnoses of disease. It is possible to use information gathered from Internet of Things (IoT) sensors worn by a person, such as glucose monitors, to make early predictions about the likelihood of developing noncommunicable diseases (NCDs) like diabetes. The system was trained using an actual diabetes dataset, and then tested using synthetic sensor data. Following extensive testing of numerous machine learning strategies, it was determined that the Random Tree method, which takes only 0.01 seconds to create a model and has an accuracy of 94%, is the most effective for early diabetes prediction.

Choice of Indicators:

Feature Selection is a method for reducing the dimensionality of a dataset. In medical diagnosis, it is crucial to isolate the most important factors associated with illness risk. Identifying essential features in the dataset can help speed up and enhance the accuracy of disease detection.

Using the inputs and outputs of the model, predictions can be made, and then those predictions may be applied to testing data to get classification results. In an effort to aid in the identification of chronic disease, numerous classification algorithms have been tried out on disease datasets with encouraging results. The diagnosis of chronic diseases can be sped up and made easier with the use of a unique classification technique.

Due to the current data explosion, new medical records are being created and updated on a regular basis. The EHR contains the clinical notes, test results, medications, and pharmacy records of the patient. The EHR also incorporates social media content, such as blog postings and tweets. There is a need for a parallel data processing method that can store and analyse massive numbers of medical records.

Monitoring and controlling a chronic illness may benefit from the use of CDD systems or similar tools. It aids doctors in keeping tabs on their patients around the clock.

Selection of feature characteristics, also known as Variable Selection, is an important data preprocessing approach in data mining for decreasing dataset size by eliminating irrelevant and unused attributes [3]. Improved prediction accuracy is just one side effect of this principle; other advantages include a deeper comprehension of the data and less time spent training learning algorithms.

Health care is only one of several fields where techniques for identifying relevant features might be put to use. It is possible to pick variables in a variety of ways, some of which include filters, wrappers, ensembles, and embedding techniques. There has been a rise in recent years in the application of hybrid methods to feature selection in the scientific community.

It is recommended practise to clean the data of any inconsistencies or noise before applying a model to it. As real-world applications grow more sophisticated, the need to streamline their data storage strategies becomes more pressing. Using only the most important features also has the added benefit of keeping things simple.

Several feature-recognition algorithms have been applied to healthcare datasets in recent years for the purpose of gaining access to more relevant data. Clinical studies make use of feature selection techniques.

III. Proposed Model

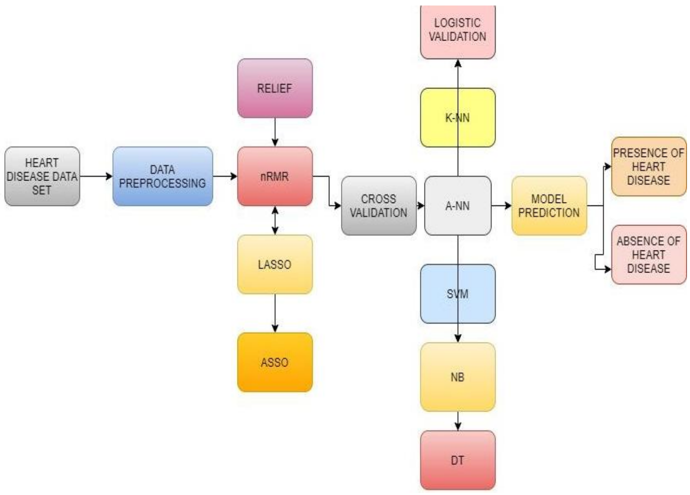

Figure 2 : Proposed model

Figure 2 displays the suggested model in which a heart disease dataset is preprocessed using four techniques (likely Relief, Rmr, LASSO, and ASSO) and then cross-validated using machine learning classifiers (such K-NN, A-NN, SVM, NB, and DT).

Implementation

A. AN ASSORTMENT OF MACHINEVLEARNING DATBASES

The model is constructed in Python utilising libraries like Panda,, Pyplot, and Scikit-learn in Jupiter notebook.

B. INFORMATIONAL SET

Accurate results from machine learning algorithms depend on a wealth of data. UCI's machine learning repository provided the dataset used in this study. These five data sets cover the areas of Cleveland, Hungary, Switzerland, VA Long Beach, and Statlog. Combining them in this study allowed us to get more precise results. They have compiled a textbook with over 1190 examples and 14 distinguishing features from their database. There is a diagnostic input for each of the other 12 attributes, and a diagnostic output for each of the other 12 attributes. A patient's age, sex, BP at rest, BG at fasting, CP type, and EC tracings at rest were all present or nearly present in all medical records. Figure 3 displays the ranges of values we've assigned to a number of features, such as age, gender, cp, trestbps, chol, and more.

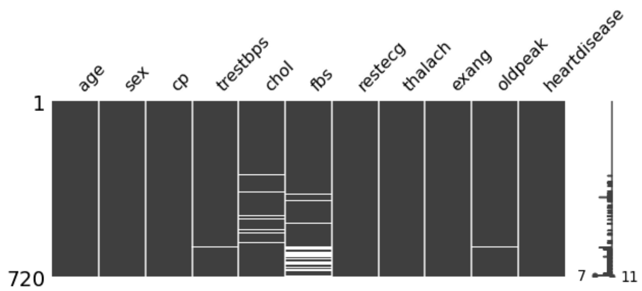

Figure 3: Different Attributes of Data set

C. REVIEW OF DATA PREPARATION AND CLEANING TECHNIQUES. Obtaining large amounts of information in the modern era may involve using the internet, surveys, or experiments. Unfortunately, most useful data has errors, outliers, and other imperfections that prevent it from being fully utilized. For this reason, some information is absent or unreachable in the dataset used for this analysis. To deal with missing data, imputation or deletion is frequently used.

D. METHODS FOR CHOOSING PARTICULAR FEATURES

To determine the most useful characteristics for data classification, machine learning can use a wide range of approaches. This is a bonus that contributes to the efficiency of the process. We've narrowed our options down to two main strategies: the Least Absolute Shrinkage and Selection Operator, and selecting relief features. By using this strategy, you can narrow in on the most effective pain treatments for you.

E. TO COMBINE MACHINE LEARNING METHODS

More accurate classification results can be obtained using an ensemble technique because it combines multiple Decision Tree classifiers. To boost the model's accuracy and precision, an ensemble of weak learners can be combined into a single strong learner. Figure 5 displays the three most common reasons for the discrepancy between the actual and predicted outcomes. All machine learning ensembles function in this way. The application of an ensemble method offers a potential solution to several of these problems.

F. PROPOSED METHOD OF CATEGORIZATION

In this article, we will examine the machine learning approaches currently being used to create a reliable method for predicting heart illness.

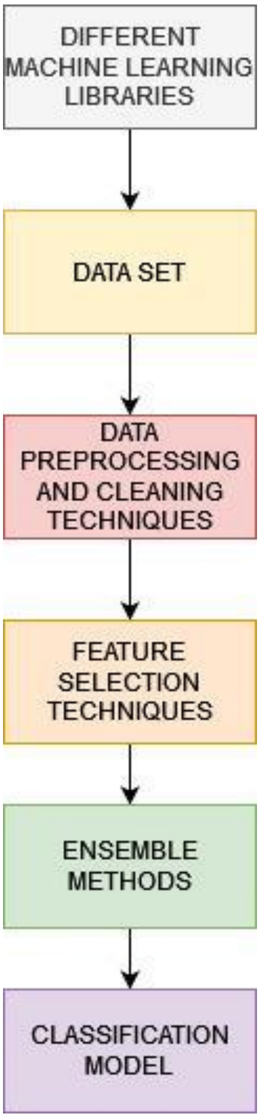

Figure 4: Implementation Techniques

Figure 4 shows how to put everything together, from machine learning libraries and data sets to data preprocessing and feature selection tools, ensemble and classification models, and more.

A. VARIOUS MEANS OF MACHINERY EDUCATION

The model is constructed in Jupiter notebook's Python programming language utilising Panda, Pyplot, and Scikit-learn as simple libraries.

B. DATASET

In order to achieve correct results from machine learning algorithms, data is considered the first and most fundamental step. UCI machine learning repository is the source of the dataset that is used in the application. It is possible to use a variety of different datasets: the VA Long

A. A SUMMARY OF DATA PROCESSING AND PURIFICATION METHODS

More accurate classification results can be obtained with an ensemble strategy because it combines multiple Decision Tree classifiers. To boost the model's accuracy and precision, an ensemble of weak learners can be integrated into a single strong learner. Figure 5 displays the three most common reasons for the discrepancy between the actual and predicted outcomes. All machine learning ensembles function in this way. The application of an ensemble method offers a potential solution to several of these problems.

IV. Consequences and Outcomes

In results the comparison of some parameters done in paper [26] and some are added in this work and compared with existing work. Classification of Heart Diseases by Gender, Age, sex etc. is shown in figure 5.

Methodology Evaluation in Light of Memory Recall

The recall or sensitivity score is an important performance metric because accurate diagnosis of heart disease is so important. FIGURE 5: Recall Scores for Different Methods and Feature Sets. KNN only correctly identified 84% of the first 13 characteristics, while RFBM got the highest possible recall score (92%) possible. Recall scores of 89% were achieved by ABBM, KNNBM, RF, and GBBM using all 11 LASSO features; recall scores of 86% were achieved by RF using only 10 features; and recall scores of 89% were achieved by RFBM and GBBM using all 11 features. When trained on the 10 Relief characteristics, all of the aforementioned classifiers and hybrids achieved recall rates of around 98%. When using RFBM, participants had the highest recall of all 10 Relief features.

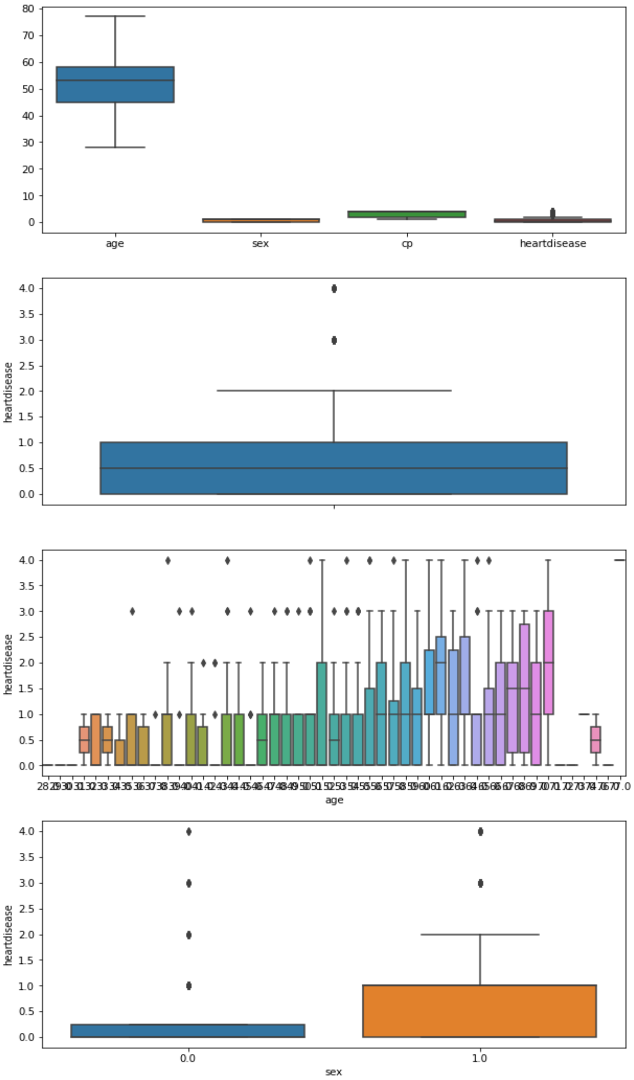

Figure 5: Classification of Heart Diseases

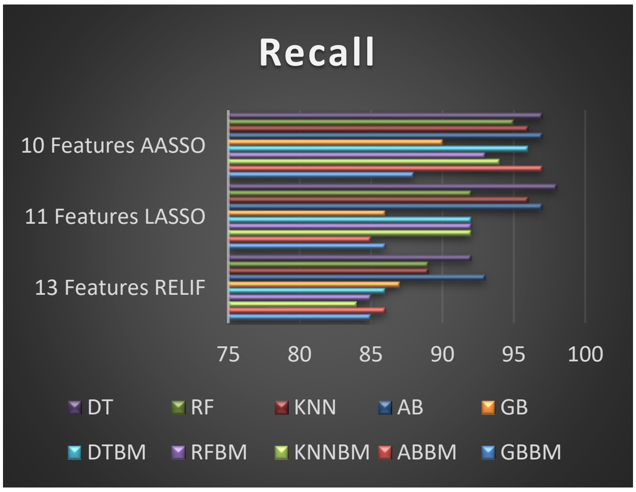

Figure 5: Comparison of Recall

B. Evaluation of Various Strategies Using the F1 Score

This measure consists of two parts: accuracy and memory. The RFBM outperformed all other algorithms with 13 features by achieving the highest F1-score (about 92%). KNN had the worst F1 score (84 percent) when evaluating 13 features, followed by DT (87 percent) and the GBM (90 percent) (88 percent). The F1-score improved when the number of features was decreased. The best performance was seen in classifiers that had F1 scores higher than GBBM's for 11 attributes. For the ten Relief characteristics evaluated, this means that KNNBM and GBBM now have F1 ratings of around 98%, while DT, RF, and AB now have F1 ratings of 90%, 988%, and 93%, respectively. The RFBM model produced the highest F1-score (99%), while the KNNBM model produced the second highest F1-score (89%). For ten of the most common characteristics, the DTBM model scored the lowest. The F1 scores are shown in Figure 5.

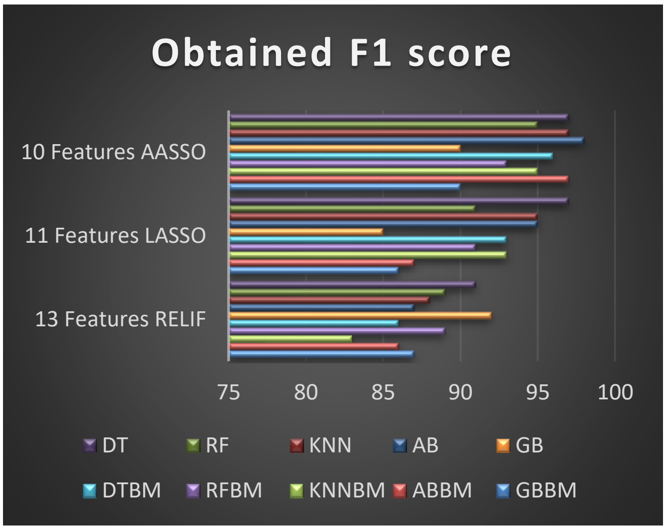

Figure 6: Comparison of F1 Score

C. Comparison of Different methods based on False Positive rate

Several algorithms' false-positive rates are displayed both before and after feature selection. False-positive rates varied widely between methods, with the RFBM having the lowest at 2.05%. Classifiers and hybrid algorithms have far greater false-positive rates when they don't use the Relief or LASSO feature selection techniques. Results for FPR ranged from about 3.5%

to 3.7% for RF, KNNBM, GBBM, and others. Of the 13 attributes considered, RFBM's FPR was the lowest, while KNN's was exceptionally high. The FPRs are shown in Figure 6.

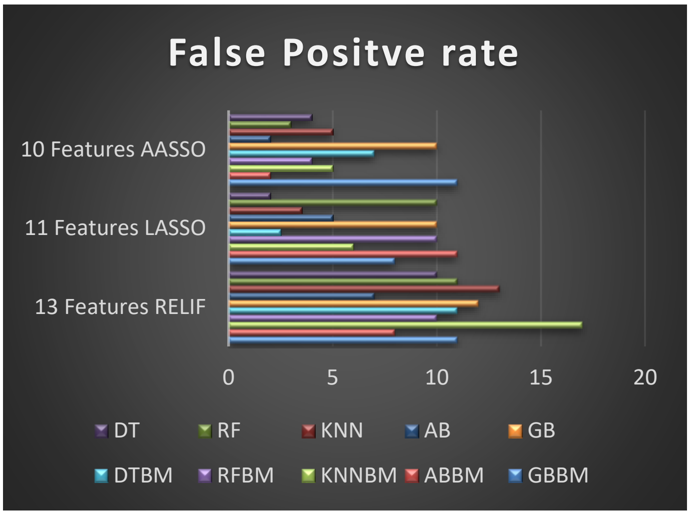

Figure 7 : Comparison of False positive rate

V. Conclusion

If heart disease risk could be anticipated with sufficient precision, long-term mortality rates might be dramatically decreased for people of all socioeconomic and cultural backgrounds. Early detection of the disease is crucial to achieving this objective. Several studies have already used machine learning to foresee cardiac disease. This research takes a similar approach, albeit with some key differences, including the use of a more advanced methodology and a more extensive dataset for model training. This research demonstrates that the Relief feature selection strategy can generate highly connected feature sets that can be used in a variety of machine learning methods. Deep Learning algorithms represent yet another future possibility. The primary objective of this research was to contribute novel information to the existing body of knowledge. In the future, we hope to make the model more generic so that it can be used to a variety of feature selection strategies and data sets, and so that it may be practically applicable and straightforward to implement.

VI. REFERENCES

[1]. S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019. doi: 10.1109/ACCESS.2019.2923707

[2]. J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in IEEE Access, vol. 8, pp. 107562-107582, 2020. doi: 10.1109/ACCESS.2020.3001149

[3]. H. Wang, Z. Huang, D. Zhang, J. Arief, T. Lyu and J. Tian, "Integrating Co-Clustering and Interpretable Machine Learning for the Prediction of Intravenous Immunoglobulin Resistance in Kawasaki Disease," in IEEE Access, vol. 8, pp. 97064-97071, 2020.

[4]. N. L. Fitriyani, M. Syafrudin, G. Alfian and J. Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," in IEEE Access, vol. 8, pp. 133034-133050, 2020. doi: 10.1109/ACCESS.2020.3010511

[5]. P. Ghosh et al., "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," in IEEE Access, vol. 9, pp. 19304-19326, 2021. doi: 10.1109/ACCESS.2021.3053759

[6]. M. A. Khan and F. Algarni, "A Healthcare Monitoring System for the Diagnosis of Heart Disease in the IoMT Cloud Environment Using MSSO-ANFIS," in IEEE Access, vol. 8, pp. 122259-122269, 2020. doi: 10.1109/ACCESS.2020.3006424

[7]. C. Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han and J. Yu, "Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform," in IEEE Access, vol.

8, pp. 59247-59256, 2020. doi: 10.1109/ACCESS.2020.2981159

[8]. J. Zhang et al., "Coupling a Fast Fourier Transformation With a Machine Learning Ensemble Model to Support Recommendations for Heart Disease Patients in a Telehealth Environment," in IEEE Access, vol. 5, pp. 10674-10685, 2017. doi: 10.1109/ACCESS.2017.2706318

[9]. G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," in IEEE Access, vol. 8, pp. 157643-157653, 2020. doi: 10.1109/ACCESS.2020.3015757

[10]. S. J. Pasha and E. S. Mohamed, "Novel Feature Reduction (NFR) Model With Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction," in IEEE Access, vol. 8, pp. 184087-184108, 2020. doi: 10.1109/ACCESS.2020.3028714

[11]. S. A. Ali et al., "An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo–Tompa and Stacked Genetic Algorithm," in IEEE Access, vol. 8, pp. 65947-65958, 2020. doi: 10.1109/ACCESS.2020.2985646

[12]. C. Xiao, Y. Li and Y. Jiang, "Heart Coronary Artery Segmentation and Disease Risk Warning Based on a Deep Learning Algorithm," in IEEE Access, vol. 8, pp. 140108-140121, 2020. doi: 10.1109/ACCESS.2020.3010800

[13]. A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," in IEEE Access, vol. 7, pp. 180235-180243, 2019. doi: 10.1109/ACCESS.2019.2952107

[14]. M. Alkhodari, D. K. Islayem, F. A. Alskafi and A. H. Khandoker, "Predicting Hypertensive Patients With Higher Risk of Developing Vascular Events Using Heart Rate Variability and Machine Learning," in IEEE Access, vol. 8, pp. 192727-192739, 2020. doi: 10.1109/ACCESS.2020.3033004

[15]. L. Ali et al., "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," in IEEE Access, vol. 7, pp. 54007-54014, 2019. doi: 10.1109/ACCESS.2019.2909969

[16]. L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed and J. A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}}$ Statistical Model and Optimally Configured Deep Neural Network," in IEEE Access, vol. 7, pp. 34938-34945, 2019.

[17]. J. Wang et al., "A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease," in IEEE Access, vol. 8, pp. 37124-37133, 2020. doi: 10.1109/ACCESS.2020.2975377

[18]. D. Lai, Y. Zhang, X. Zhang, Y. Su and M. B. Bin Heyat, "An Automated Strategy for Early Risk Identification of Sudden Cardiac Death by Using Machine Learning Approach on Measurable Arrhythmic Risk Markers," in IEEE Access, vol. 7, pp. 94701-94716, 2019. doi: 10.1109/ACCESS.2019.2925847

[19]. A. Mdhaffar, I. Bouassida Rodriguez, K. Charfi, L. Abid and B. Freisleben, "CEP4HFP: Complex Event Processing for Heart Failure Prediction," in IEEE Transactions on NanoBioscience, vol. 16, no. 8, pp. 708-717, Dec. 2017. doi: 10.1109/TNB.2017.2769671

[20]. A. Ishaq et al., "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," in IEEE Access, vol. 9, pp. 39707-39716, 2021. doi: 10.1109/ACCESS.2021.3064084

[21]. M. Alkhodari, H. F. Jelinek, N. Werghi, L. J. Hadjileontiadis and A. H. Khandoker, "Estimating Left Ventricle Ejection Fraction

Levels Using Circadian Heart Rate Variability Features and Support Vector Regression Models," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 3, pp. 746-754, March 2021. doi: 10.1109/JBHI.2020.3002336

[22]. S. Kaur et al., "Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives," in IEEE Access, vol. 8, pp. 228049-228069, 2020. doi: 10.1109/ACCESS.2020.3042273

[23]. D. Tay, C. L. Poh, E. Van Reeth and R. I. Kitney, "The Effect of Sample Age and Prediction Resolution on Myocardial Infarction Risk Prediction," in IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 3, pp. 1178-1185, May 2015. doi: 10.1109/JBHI.2014.2330898

[24]. W. Chang, Y. Liu, X. Wu, Y. Xiao, S. Zhou and W. Cao, "A New Hybrid XGBSVM Model: Application for Hypertensive Heart Disease," in IEEE Access, vol. 7, pp. 175248-175258, 2019. doi: 10.1109/ACCESS.2019.2957367

[25]. R. Ferdousi, M. A. Hossain and A. E. Saddik, "Early-Stage Risk Prediction of Non-Communicable Disease Using Machine Learning in Health CPS," in IEEE Access, vol. 9, pp. 96823-96837, 2021. doi: 10.1109/ACCESS.2021.3094063

[26]. Manoj D. Patil, Dr. Harsh Mathur, "Prediction of Cardiovascular Disease on Different Parameters Using Machine Learning", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 5, pp. 52-61, SeptemberOctober 2021. Available at doi : https://doi.org/10.32628/IJSRSET218486 Journal URL : https://ijsrset.com/IJSRSET218486

Cite this article as :

Sumit Kumar Soni, Dr. Kalpana Rai, Dr. Harsh Mathur, "Using an Adapted Hybrid Intelligent Framework to Make Predictions Regarding Heart Diseases", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 2, pp.214-232, March-April-2023. Available at doi : https://doi.org/10.32628/CSEIT2390222

Journal URL : https://ijsrcseit.com/CSEIT2390222