Prediction of Cardiovascular Disease using Hybrid Machine Learning Methods

Devarasetty Purna Kishore[1], I.PHANI KUMAR[2]

#1Student, Dept of CSE, VelagaNageswaraRao College Of Engineering, Ponnur(Post),Ponnur(Md)Guntur(D.T)A. Andhra Pradesh.

#2Assoc. Professor, Dept of CSE, VelagaNageswaraRao College Of Engineering, Ponnur(Post),Ponnur(Md)Guntur(D.T)A. Andhra Pradesh

ABSTRACT_ Today, heart disease ranks high on the list of leading killers worldwide. Clinical data analysis has a major problem in the domain of cardiovascular disease prediction. Since there is so much data generated by the healthcare business, machine learning (ML) has been proven to be an excellent tool for helping to make judgments and predictions. Recent innovations in various domains of the Internet of Things have also made use of ML approaches (IoT). Predicting cardiac disease using ML approaches is only partially explored in the available research. In this study, we offer a new strategy for predicting cardiovascular illness that relies on machine learning approaches to identify meaningful information. We introduce the prediction model by utilising a number of feature combinations and well-established categorization strategies. Through the use of the hybrid random forest with the linear model, we are able to get a higher performance level with an accuracy level of 88.7 percent in our heart disease prediction model (HRFLM)

1.INTRODUCTION

Multiple risk factors, including diabetes, hypertension, abnormal cholesterol levels, and an irregular heart rate, make it challenging to diagnose heart disease. Human heart disease severity has been determined using a number of data mining and neural network approaches. Various methods, such as the K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic Algorithm (GA), and Naive Bayes (NB), are used to categorise the severity of the condition [11], [13]. Because of its complexity, heart disease requires meticulous management. Failure to do so may result in cardiovascular problems or even death. Finding the many forms of metabolic syndrome requires a medical perspective and the application of data mining techniques. Predictions of cardiac illness and data investigations benefit greatly from data mining with classification. Heart disease-related events can also be accurately predicted using decision trees [1]. The prediction of cardiovascular disease has been the subject of a number of knowledge-abstraction strategies that make use of established data-mining techniques. A large number of sources have been read in order to develop a prediction model that integrates elements of multiple methods. Hybrid approaches [14] refer to methods that combine existing ones with new ones.

The time stream from a cardiac monitor is used to introduce neural networks. Left bundle branch block (LBBB), right bundle branch block (RBBB), atrial fibrillation (AFIB), normal sinus rhythm (NSR), sinus bradycardia (SBR), atrial flutter (AFL), premature ventricular contraction (PVC), and second degree block (BII) are some of the clinical records that are used for prediction in this method. If you're using a radial basis function network (RBFN) to do your classification, you should know that 70 percent of the data is used for training and only 30 percent is used for classification [4, 15]. We also provide an overview of the Computer Assisted Decision Support System (CADSS) and its applications in the scientific and medical communities. Previous research has demonstrated that data mining techniques can be used in the healthcare business to forecast disease with more accuracy in less time [16]. We recommend using the GA to identify cardiac issues. The new proposed fitness function is the consequence of this method's application of effective association rules inferred with the GA during tournament selection, crossover, and mutation. The Cleveland dataset, built from a UCI machine learning repository, is used for experimental validation. Later, we'll look at how our findings stand out in comparison to other supervised learning methods [5, 17]. Here we propose Particle Swarm Optimization (PSO), the most powerful evolutionary method, and produce some principles for cardiovascular illness. Rules have been applied arbitrarily using encoding methods, leading to an increase in accuracy [2]. Pulse rate, sex, age, and many other factors are used as predictors of cardiovascular disease. Here we introduce the ML approach using Neural Networks, which, as we have seen in [8], [12], produces more accurate and trustworthy results. Predicting the onset of serious illnesses like cardiovascular disease and neurological disorders is a strong area for neural networks. We utilise a 13-factor model to predict cardiac issues. Performance is shown to be improved above that of previously published approaches [3, for example]. In recent times, CAS (stenting of the carotid artery) has also emerged as a common method of treatment. Major adverse cardiovascular events (MACE) are more likely to occur in older people with heart disease when the CAS is present. Their judgement becomes crucial. Our work is generated using an ANN, which has shown promising results in the prediction of cardiac issues [6, 18]. In this article, we introduce neural network methods, which integrate not just posterior probabilities but also predicted values from a number of prior approaches. When compared to prior works, this model's accuracy of up to 89.01% is impressive. Using a Neural Network NN, we increase the performance of heart disease as observed in [9], [19] across all tests using the Cleveland heart dataset. Recent advances in machine learning (ML) approaches applied to the Internet of Things (IoT) have also been observed [43]. Accurate identification of networked IoT devices has been demonstrated using ML algorithms applied to network traffic data. Meidan et al. gathered network traffic data from nine different Internet of Things devices, personal computers, and mobile phones, and labelled the data. In this study, they trained a multi-stage meta classifier using supervised learning. Initially, the classifier can tell the difference between traffic from IoT and non-IoT sources. In the second phase, a category for each type of IoT device is assigned. Using deep learning to sift through massive amounts of raw sensor data from IoT devices in challenging settings [44, 47] is an exciting prospect. Deep learning's layered architecture is well-suited to the distributed nature of edge computing [48], [49]. Specifically, we provide a method called the Hybrid Random Forest with Linear Model (HRFLM). The primary aim of this study is to enhance prediction accuracy for cardiovascular disease. Many investigations have led to limitations on feature selection for algorithms. HRFLM, on the other hand, makes use of all functionalities without any limitations on feature selection. In this work, we employ a hybrid approach, conducting tests to determine the algorithm's features. The experimental results demonstrate that our suggested hybrid strategy outperforms state-of-the-art technologies in its capacity to predict cardiovascular disease.

2.LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company Traffic Redundancy Elimination, once these things are satisfied, then next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support.

This support can be obtained from senior programmers, from book or from websites. Before building the system we have to knownthe below concepts for developing the proposed system.

The main aim of this research work is to classify the emotional expression from the mouth region of the human face. As the initial task is to extract the mouth region from the facial image, a survey on various existing research works to segment the face expression images is reviewed and discussed. 2.1 A. L. Bui, T. B. Horwich, and G. C. Fonarow, ''Epidemiology and risk profile of heart failure,'' Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011.

Heart failure (HF) is a major public health issue with a current prevalence of over 5.8 million in the USA and over 23 million worldwide.[1,2] Every year in the USA, more than 550,000 individuals are diagnosed with HF for the first time, and there is a lifetime risk of one in five of developing this syndrome.[1,3] A diagnosis of HF carries substantial risk of morbidity and mortality, despite advances in management.

Over 2.4 million patients who are hospitalized have HF as a primary or secondary diagnosis, and nearly 300,000 deaths annually are directly attributable to HF.[1]

From the 1970s to 1990s, a dramatic increase in the prevalence of HF and number of HF hospitalizations was observed,[4–6] and an epidemic was declared.[7,8] Most of the HF burden is borne by individuals aged ≥65 years, who account for more than 80% of the deaths and prevalent cases in the USA and Europe.[6,9] The growing prevalence of HF might reflect increasing incidence, an aging population, improvements in the treatment of acute cardiovascular disease and HF, or a combination of these factors. Promising evidence from national databases as well as community-based cohorts, such as those based in Framingham and Olmsted County,[3,10–16] indicates that the incidence of HF seems to be stabilizing, if not decreasing, for women, and that the length of survival in patients with HF is increasing. Such trends may have resulted from demographic shifts, changes in the prevalence of risk factors, or improvements in the availability and application of HF treatments.[17,18] Furthermore, awareness of and appreciation for HF and preserved left ventricular ejection fraction (LVEF) is increasing. HF and preserved LVEF now represents >50% of HF cases and can have outcomes as poor as those associated with HF and reduced LVEF, but it does not yet have a proven effective management strategy.[19–21] In this Review, we describe the epidemiology of HF, highlighting trends in overall prevalence, incidence, and mortality of HF as a whole and in subgroups. We also highlight how identified risk factors influence both incidence and severity of HF and discuss the impact of HF on the utilization of health services.

3.PROPOSED WORK

Here, the author compares and contrasts the accuracy of several popular heart disease prediction algorithms, including support vector machine, naive bayes, and logistic regression. All of these algorithms are great at making predictions, but they aren't very accurate. Author combines Linear Model and Random Forest classification methods to create new algorithm called Hybrid Machine Learning to improve prediction accuracy of cardiac dataset. The Voting classifier will be constructed from a Linear Model and a Random Forest within the hybrid algorithm, and the classification voting algorithm will compare the two algorithms' prediction accuracy and select the one with the highest score. Consequently, a more accurate algorithm for predicting cardiovascular disease can be developed by constantly employing a hybrid model.

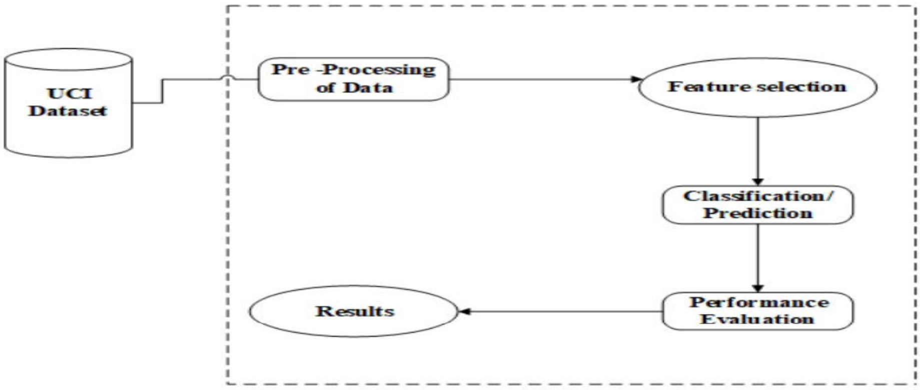

Fig 1:Architecture

4.RESULTS AND DISCUSSION

Fig 2:In above screen click on 'Upload Heart Disease Dataset' button to upload heart dataset

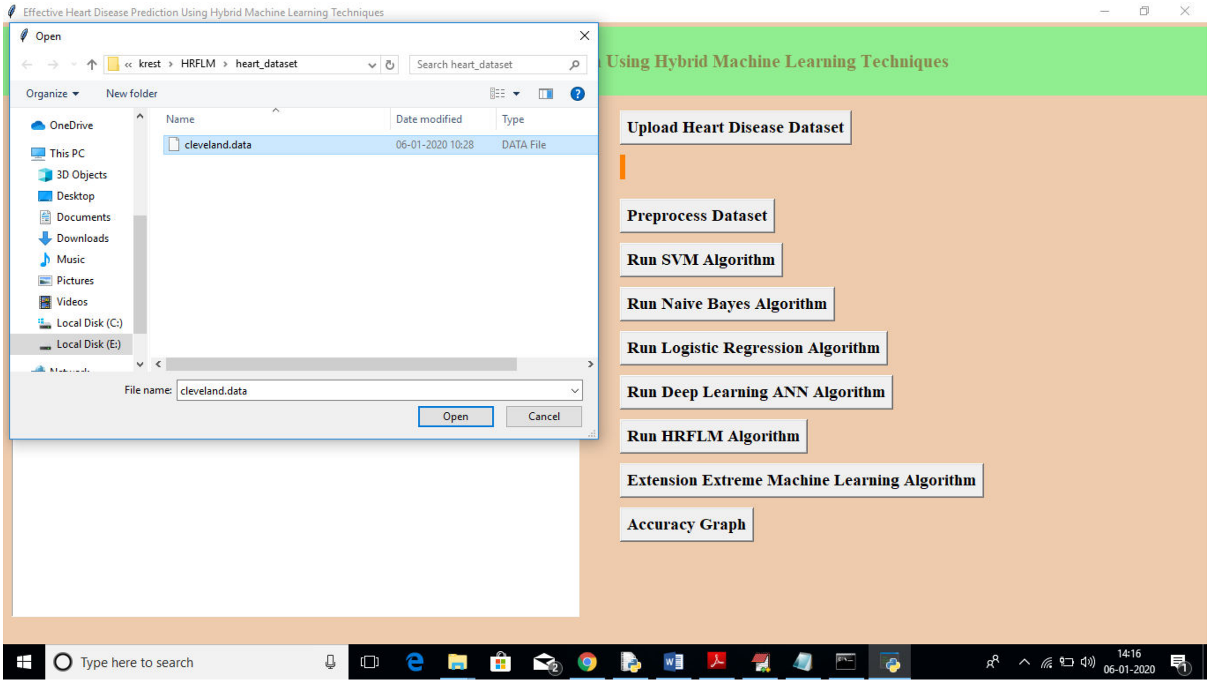

Fig 3:In above screen I am uploading 'cleveland.data' dataset, after uploading dataset will get below screen

Fig 4:In above screen we can see dataset contains total 303 records, now click on 'Pre-process Dataset' button to apply pre-processing technique to remove out all non-numeric data.

Fig 5:In above screen after applying pre-processing dataset size reduced to 297 records and we can see application randomly splitted complete dataset in to tow parts called train and test. For training application using 237 records and for testing application using 60 records. Application will choose random 60 records so always accuracy of same algorithm will be different as records for testing are randomly chooses.

Now click on 'Run SVM Algorithm' button to generate SVM model on train dataset and to apply test data to get SVM classification accuracy.

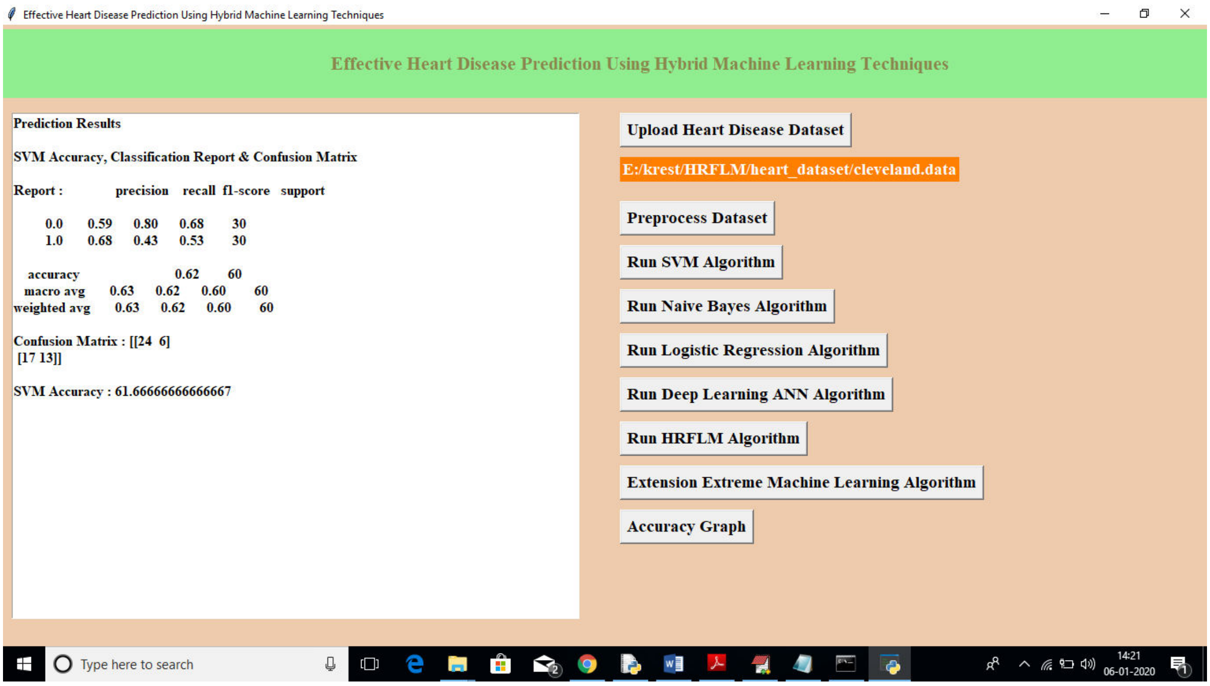

Fig 6:In above screen SVM got 62% accuracy, now click on 'Run Naïve Bayes Algorithm' button to get its accuracy

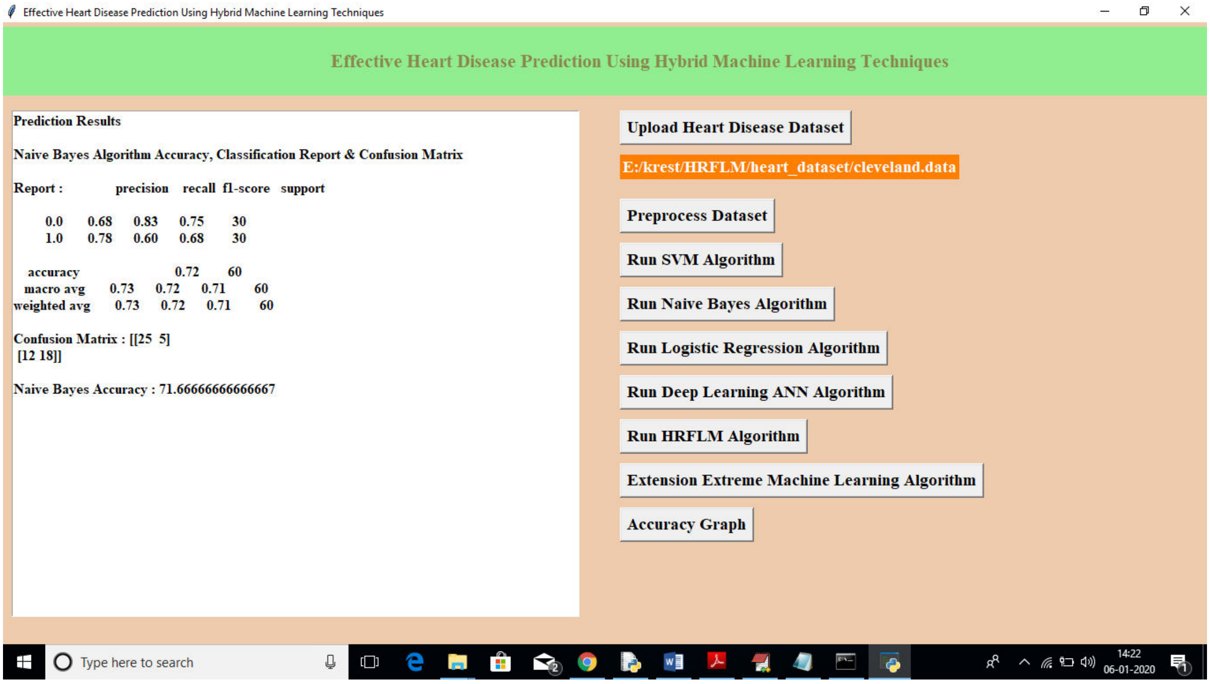

Fig 7:In above screen we can see Naïve Bayes got 72% accuracy, now click on 'Run Logistic Regression Algorithm' to get its accuracy

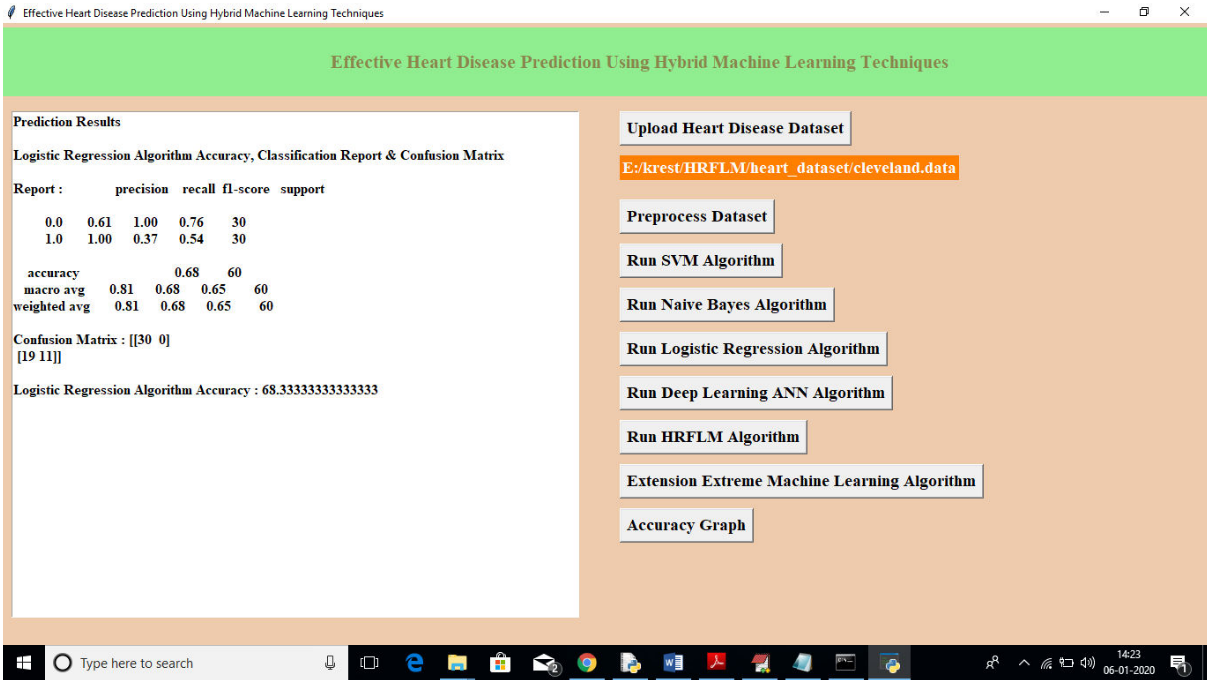

Fig 8:In above screen logistic regression got 69% accuracy, now click on 'Run Deep Learning ANN Algorithm' button to get its accuracy

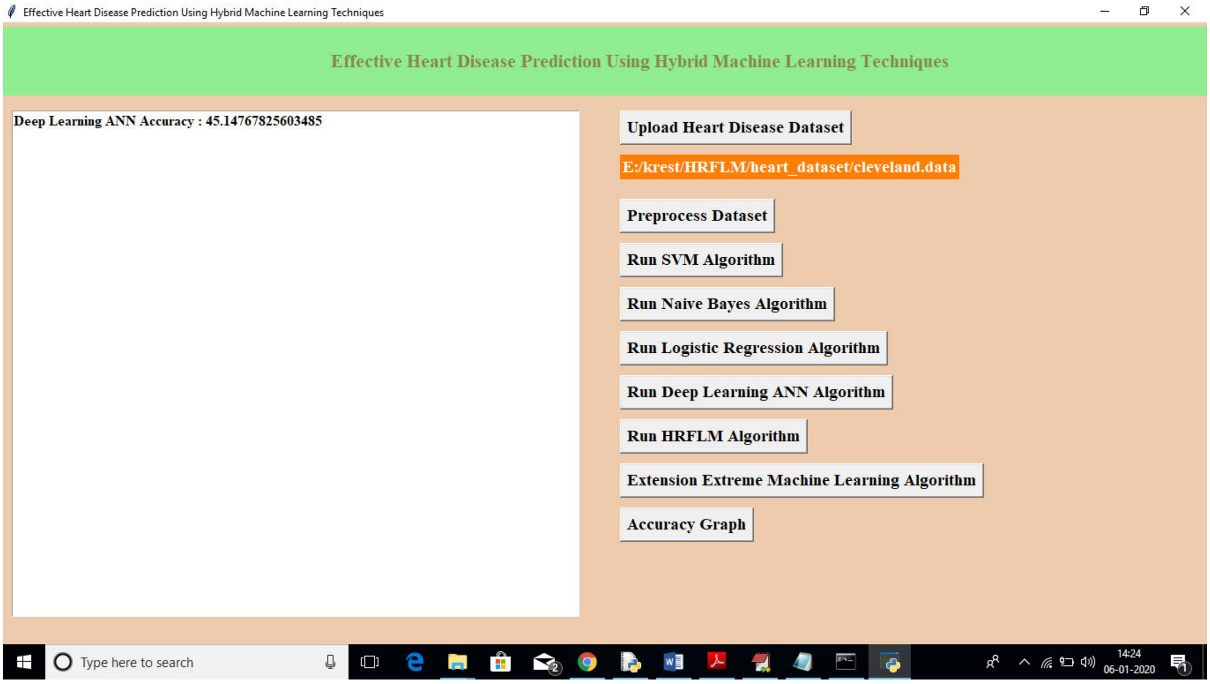

Fig 9:In above screen we can see ANN got 46% accuracy, now click on 'Run HRFLM Algorithm' button to get propose work accuracy

Fig 10:In above algorithm we can see HRFLM got 84% accuracy, now click on 'Extension Extreme Machine Learning Algorithm' button to check EML extension accuracy

Fig 11:In above graph x-axis represents algorithm names and y-axis represents accuracy of that algorithm. In all algorithms propose HRFLM and extension algorithm got better accuracy

5.CONCLUSION

The long-term goal of saving lives and detecting irregularities in heart problems early through the identification of the processing of raw healthcare data of heart information can be achieved through this identification. In this study, we employed machine learning methods to analyse raw data and draw unique conclusions about cardiovascular disease. Predicting the onset of heart disease is a major challenge for the medical community. But if the condition is caught early and preventative steps are taken, the mortality rate can be significantly reduced. It would be ideal to broaden the scope of this research to include real-world datasets rather than only theoretical methods and simulations. The suggested hybrid HRFLM method combines the best features of both the Random Forest (RF) and Linear Method approaches (LM). HRFLM's predictive abilities for cardiovascular disease were shown to be high. Different combinations of machine learning methods can be used to improve prediction methods in the future of this study. To further improve the efficacy of heart disease prediction, novel featureselection methods can be created to obtain a wider understanding of the important features.

REFERENCE

[1] A. S. Abdullah and R. R. Rajalaxmi, ''A data mining model for predicting the coronary heart disease using random forest classifier,'' in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ''Using PSO algorithm for producing best rules in diagnosis of heart disease,'' in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306– 311.

[3] N. Al-milli, ''Backpropogation neural network for prediction of heart disease,'' J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ''Analysis of neural networks based heart disease prediction system,'' in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.

[5] P. K. Anooj, ''Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,'' J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27–40, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002.

[6] L. Baccour, ''Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets,'' Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.

[7] C.-A. Cheng and H.-W. Chiu, ''An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-widedatabase,'' in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566–2569.

[8] H. A. Esfahani and M. Ghazanfari, ''Cardiovascular disease detection using a new ensemble classifier,'' in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014.

[9] F. Dammak, L. Baccour, and A. M. Alimi, ''The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains,'' in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1–8.

[10] R. Das, I. Turkoglu, and A. Sengur, ‘‘Effective diagnosis of heart disease through neural networks ensembles,’’ Expert Syst. Appl., vol. 36, no. 4, pp. 7675–7680, May 2009. doi: 10.1016/j.eswa.2008.09.013.

[11] M. Durairaj and V. Revathi, ‘‘Prediction of heart disease using back propagation MLP algorithm,’’ Int. J. Sci. Technol. Res., vol. 4, no. 8, pp. 235–239, 2015.

Author’s Profiles

Devarasetty Purna Kishore pursing M. Tech in Computer Science and Engineering from Velaga Nageswara Rao College Of Engineering, Ponnur. Affiliated to JNTUK, KAKINADA.

I.Phani Kumar, Qualification: (M.Tech,Ph.D), having 11 years experience in teaching field at present he is working as Assoc professor in Velaga Nageswara Rao College Of Engineering, Ponnur(Post),Ponnur(Md)Guntur(D.T),And hrapradesh. PHANIKUMAR148@gmail.com, phone:9110323715