INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

ISSN : 0975-4520

STRESS DETECTION FOR WEARABLE DEVICES

P.NETRASRI [1], CHALLAPURAM SPANDANA [2], MOHAMMED MAJID ALI [3], THOTA KUNDAN KUMAR [4]

[1] Assistant Professor, Department of CSE, **ST.PETER'S ENGINEERING COLLEGE** , Hyderabad, Telangana, India.

[2,3,4] UG scholars, Department of CSE, **ST.PETER'S ENGINEERING COLLEGE**, Hyderabad, Telangana, India.

ABSTRACT:

The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions. For providing appropriate results And making effective decisions on data, some advanced data boosting techniques are used .In this study, An Effective and Efficient Heart Disease Prediction Model using Distributed High Performance light GBM is developed for predicting the risk level of heart disease. The system uses 15 medical parameters such as age, sex, blood pressure, cholesterol, obesity etc for prediction. The model predicts the likelihood of patients getting heart disease. It enables significant knowledge. E.g. Relationships between medical factors related to heart disease and patterns, to be established. We have employed the light gradient boosting algorithm as the training algorithm. The obtained results have illustrated that the designed diagnostic system can effectively predict the risk of heart diseases.

Key words: GBM, boosting algorithm, Heart performance.

I INTRODUCTION

Among all fatal disease, heart attacks diseases are considered as the most prevalent. Medical practitioners conduct different surveys on heart diseases and gather information of heart patients, their symptoms and disease progression. Increasingly are reported about patients with common diseases who have typical symptoms. In this fast moving world people want to live a very luxurious life so they work like a machine in order to earn lot of money and live a comfortable life therefore in this race they forget to take care of themselves, because of

ISSN : 0975-4520

this there food habits change their entire lifestyle change, in this type of lifestyle they are more tensed they have blood pressure, sugar at a very young age and they don't give enough rest for themselves and eat what they get and they even don't bother about the quality of the food if sick the go for their own medication as a result of all these small negligence it leads to a major threat that is the heart disease. The term 'heart disease' includes the diverse diseases that affect heart. The number of people suffering from heart disease is on the rise (health topics, 2010). The report from world health organization shows us a large number of people that die every year due to the heart disease all over the world. Heart disease is also stated as one of the greatest killers in Africa. Data mining has been used in a variety of applications such as marketing, customer relationship management, engineering, and medicine analysis, expert prediction, web mining and mobile computing. Of late, data mining has been applied successfully in healthcare fraud and detecting abuse cases.

MAIN OBJECTVE:

Data analysis proves to be crucial in the medical field. It provides a meaningful base to critical decisions. It helps to create a complete study proposal. One of the most important uses of data analysis is that it helps in keeping human bias away from medical conclusion with the help of proper statistical treatment. By use of data mining for exploratory analysis because of nontrivial information in large volumes of data. The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions for providing appropriate results and making effective decisions on data, some data mining techniques are used to better the experience and conclusion that have been given.

II LITERATURE SURVEY

Several studies have reported the development of heart disease diagnosis based on machine learning models with the aim of providing an HDPM with enhanced performance. Two publicly available heart disease datasets, namely Statlog and Cleveland, have been widely used to compare the performance of prediction

ISSN : 0975-4520

models among researchers. For Statlog dataset, a heart disease clinical decision support system based on chaos fireflfly algorithm and rough sets-based attribute reduction (CFARS-AR) was developed by Long et al. (2015) . Through sets were used to reduce the number of attributes while the chaos fireflfly algorithm was used to classify the disease. The developed model was then compared with other models such as NB, SVM and ANN.

The combination of rough sets-based attributes selection and BPNN (RS-BPNN) was proposed by Nahato et al. (2015) . With the selected attributes, the proposed RS-BPNN achieved accuracy of up to 90.4%. Dwivedi (2018) compared six machine learning models (ANN, SVM, LR, k-nearest neighbor (kNN), classifification tree and NB) with various performance metrics. The results showed that LR performed better than the other models by achieving up to 85%, 89%, 81%, and 85 for the accuracy, sensitivity, specificity, and precision, respectively.

Amin et al. (2019) performed comparison analysis by identifying significant attributes and applying machine learning models (k-NN, DT, NB, LR, SVM, Neural Network (NN) and a hybrid (voting with NB and LR)). The experiment results revealed that the hybrid model (voting with NB and LR) with selected attributes achieved the highest accuracy (87.41%). Cleveland heart disease dataset has been widely used by researchers to generate predictive models.

Verma et al. (2016) developed a hybrid prediction model based on correlation feature subset (CFS), particle swam optimization (PSO), K-means clustering and MLP. The results showed that the proposed hybrid model achieved accuracy of up to 90.28%.

Haq et al. (2018) [6] performed a comparative study on a hybrid model based on various feature selection techniques (relief, minimal-redundancy maximal-relevance (mRMR), least absolute shrinkage and selection operator (LASSO)) and machine learning models (LR, kNN, ANN, SVM, DT, NB, and RF). Their study revealed that the features reduction affects the performance of the models. The study

ISSN : 0975-4520

concluded that a combination of Relief-based feature selection and LR-based machine learning algorithm (MLA) provides higher accuracy (up to 89%) as compared with other combinations used in the study.

Saqlain et al. (2019) proposed a technique based on mean Fisher score feature selection algorithm (MFSFSA) and SVM classification model. The selected features are based on the higher Fisher score than the mean score. Then, SVM used the selected feature subset to learn and calculate the MCC through a validation process. The study revealed that the combination of FSFSA and SVM generates accuracy, sensitivity, and specifificity of up to 81.19%, 72.92%, and 88.68%, respectively.

Latha and Jeeva (2019) proposed a hybrid model with majority voting of NB, BN, RF, and MLP. The proposed model achieved an accuracy of up to 85.48%. Ali et al. (2019) [5] proposed two stacked SVMs to improve the diagnosis process.

The first SVM was used to remove the non-relevant features and the second to predict heart disease. The results revealed that the proposed model achieved better performance than other models and previous study results. Mohan et al. (2019) introduced a hybrid RF with a linear model (HRFLM) to enhance the performance of the HDPM. They found that the proposed method achieved accuracy, precision, sensitivity, f-measure and specificity of up to 88.4%, 90.1%, 92.8%, 90%, and 82.6%, respectively.

Recently, Gupta et al. (2020) developed a machine intelligence framework consisting of factor analysis of mixed data (FAMD) and RF-based MLA. The FAMD was used to find the relevant features and the RF to predict the disease. The experimental results showed that the proposed method outperformed other models and previous study results by achieving the accuracy, sensitivity, and specificity of up to 93.44%, 89.28%, and 96.96%, respectively

III EXISTING SYSTEM

Many researchers are working in the field of heart disease prediction. Shen et al. initially, proposed a self applied questionnaire (SAQ) based study to predict heart disease. This study is based on the analysis of the common risk features of the

ISSN : 0975-4520

disease and other data collected in SAQ. Dundee rank factor score is used to validate their study. This study is based on statistically 3 risk factors (blood pressure, smoking, and blood cholesterol) together with sex and age to determine the risk of having heart disease.

NORMA LATIF FITRIYANI et al. (2020) performed a study on various feature selection techniques and machine learning models. Their study revealed that the features reduction affects the performance of the models. The study concluded that a combination of Relief-based feature selection and XGBoost based machine learning algorithm . Initially, 13 clinical features i.e. age, cholesterol, chest pain type, exercise, induced angina, max heart rate, fasting blood sugar, number of vessels colored, old peak, resting ECG, sex, slope, thal, and trestbps are identified for prediction. Finally, training of neural networks is performed using XG Boosting to evaluate the heart disease prediction system. Nearly 94% accuracy is achieved on testing set for heart disease prediction by the given system.

IV PROPOSED SYSTEM

In this system we are implementing effective heart attack prediction system using Light gradient boosting algorithm. We can give the input as in CSV file or manual entry to the system. After taking input the algorithms apply on that input that is Light gradient boosting. After accessing data set the operation is performed and effective heart attack level is produced. The proposed system will add some more parameters significant to heart attack with their ecg, age and the priority levels are by consulting expertise doctors and the medical experts. We added outlier detection and validation methods for each attribute. The heart attack prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.

METHODOLOGY:

The LightGBM framework supports different algorithms includingGBT, GBDT, and RF. LightGBM has many of XGBoost's advantages, including sparse optimization, parallel training, multiple loss functions,

ISSN : 0975-4520

regularization, bagging, and early stopping. A major difference between the two lies in the construction of trees. LightGBM does not grow a tree level-wise row by row as most other implementations do. Instead it grows trees leaf-wise. It chooses the leaf it believes will yield the largest decrease in loss. Besides, LightGBM does not use the widely-used sorted-based decision tree learning algorithm, which searches the best split point on sorted feature values, as XG Boost or other implementations do. Instead, Light GBM implements a highly optimized histogram-based decision tree learning algorithm, which yields great advantages on both efficiency and memory consumption. The Light GBM algorithm utilizes two novel techniques called Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) which allow the algorithm to run faster while maintaining a high level of accuracy.

V RESULTS EXPLANATION

To check the output we have to open the anaconda prompt and we have to go the environment Which we had created and we have to go to the location where the heart disease prediction file is present, After going to the file location we have to run the python file after running the python file it shows a server address as show in fig .We have to copy the server address and we have to paste in the address section in any kind of the browser after pasting the address and clicking on it redirects to a web page as shown in the figure.

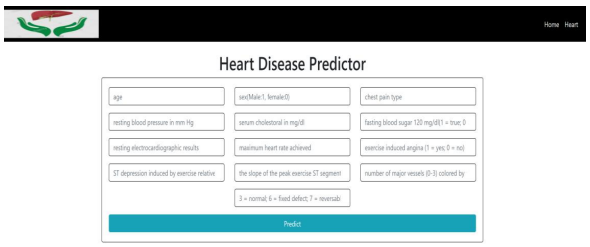

Fig.5.1. heart disease prediction page.

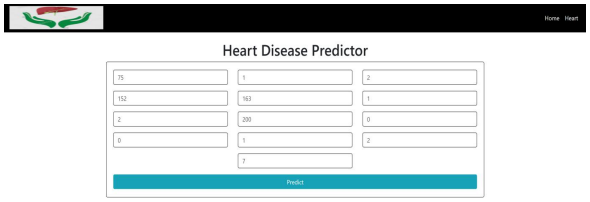

Fig.5.2. Values enter here.

ISSN : 0975-4520

Fig.5.3. Predicting Output.

This project is to know whether the patient has heart disease or not. The records in the datasets are divided into training set and test sets. After preprocessing the data, data mining classification technique namely Light Gradient Boosting was applied. This section shows the results of those classification model done using Python Programming. The results are generated for both training datasets and test data sets.

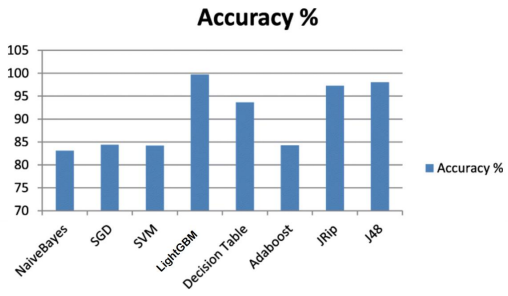

Fig.5.4. Accuracy.

CONCLUSION

In this paper, two supervised data mining algorithm was applied on the dataset to predict the possibilities of having heart disease of a patient, were analyzed with classification model namely Light gradient boosting classification. These two algorithms are applied to the same dataset in order to analyze the best algorithm in terms of accuracy. The model has predicted the heart disease patient with an accuracy level of 98% a. Thus I conclude this project by saying light gradient boosting algorithm is best and better for handling medical data set. In the future, the designed system with the used machine learning classification algorithm can be used to predict or diagnose other diseases. The work can be extended or improved for the automation of heart disease analysis including some other machine learning algorithms. Various techniques were adopted to preprocess the data to suite the requirement of analysis. Feature selections were made to optimize the performance of machine learning algorithms. Ensemble prediction gave better accuracy when combined using Random forest algorithm as combiner. Better feature

ISSN : 0975-4520

selection techniques can be applied to further improve the accuracy.

REFERENCES

[1]. Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., & Lamgunde, A. (2017, June). Heart disease prediction using data mining techniques. In 2017 International Conference on Intelligent Computing and Control (I2C2) (pp. 1-8). IEEE.

[2]. Gandhi, Monika, and Shailendra Narayan Singh. "Predictions in heart disease using techniques of data mining." In 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525. IEEE, 2015.

[3]. Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-560). IEEE.

[4]. Aldallal, A., & Al-Moosa, A. A. A. (2018, September). Using Data Mining Techniques to Predict Diabetes and Heart Diseases. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 150- 154). IEEE.

[5]. Prasadu Peddi (2016), Comparative study on cloud optimized resource and prediction using machine learning algorithm, ISSN: 2455-6300, volume 1, issue 3, pp: 88-94.

[6].Al Essa, Ali Radhi, and Christian Bach. "Data Mining and Warehousing." American Society for Engineering Education (ASEE Zone 1) Journal (2014).

[7].Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil. "Diabetes disease prediction using data mining." In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5. IEEE, 2017.

[8]. Methaila, Aditya, Prince Kansal, Himanshu Arya, and Pankaj Kumar. "Early heart disease prediction using data mining techniques." Computer Science & Information Technology Journal (2014): 53-59.

[9]. Dewan, Ankita, and Meghna Sharma. "Prediction of heart disease using a hybrid technique in data mining classification." In

ISSN : 0975-4520

2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 704-706. IEEE, 2015.

[10] Prasadu Peddi (2017) “Design of Simulators for Job Group Resource Allocation Scheduling In Grid and Cloud Computing Environments”, ISSN: 2319-8753 volume 6 issue 8 pp: 17805-17811.