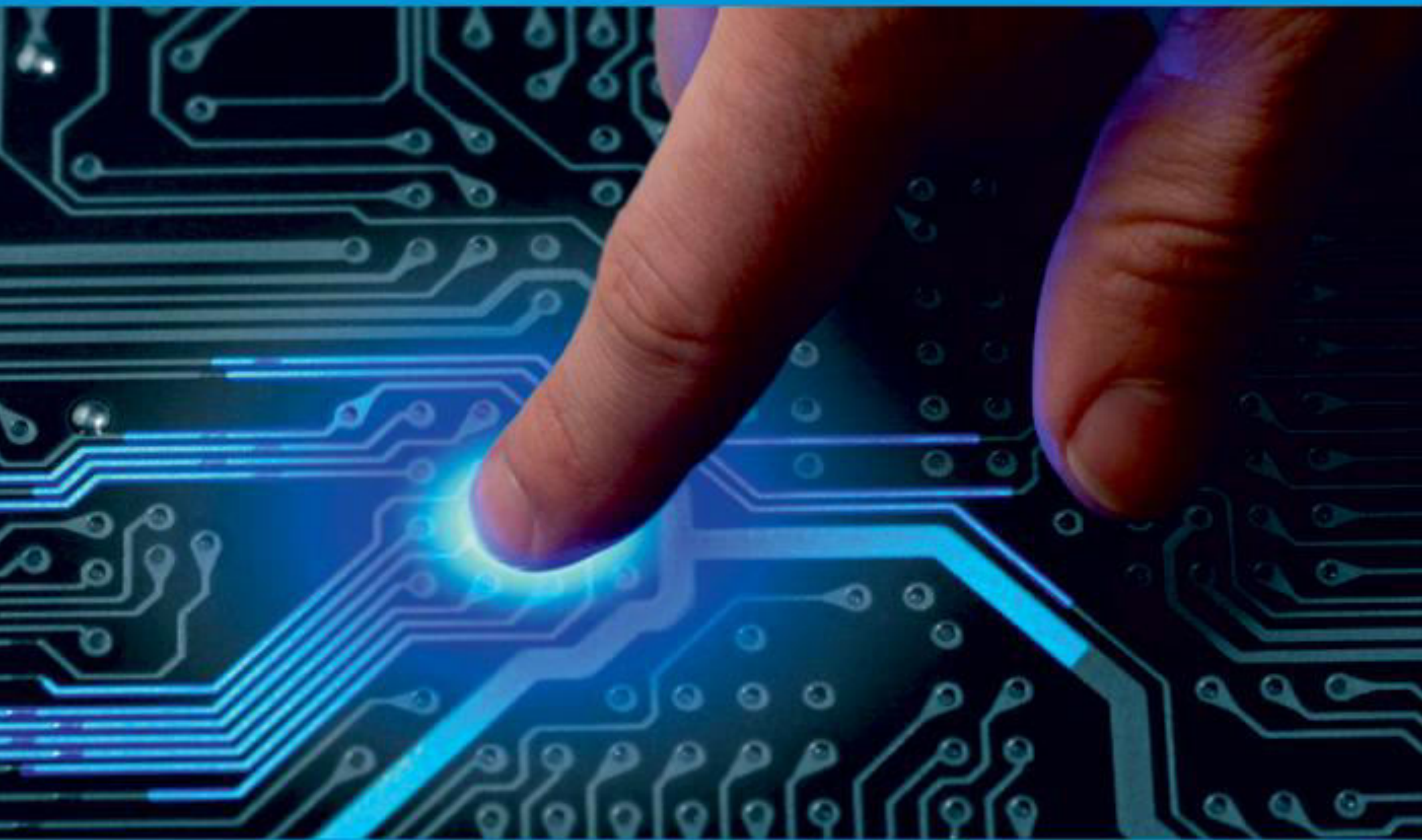
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104077 |

Performance Analysis of Cardiovascular Disease Pattern Prediction Techniques for Healthcare Monitoring Systems

Subhasini SV, S. Raja Mohamed

M.E. Computer Science and Engineering, KIT- Kalaignar Karunanidhi Institute of Technology, Coimbatore, TN, India

Head, Computer Science and Engineering, KIT- Kalaignar Karunanidhi Institute of Technology, Coimbatore, TN, India

ABSTRACT: Wearable devices and the various applications of Wearable devices are increasing greatly which encourages Researchers to focus much on Internet of Medical Things (IoMT). As we know, Machine Learning Models and the Internet of Medical Things jointly enabled methodologies for supporting healthcare services particularly for Cardiovascular Disease Pattern Prediction, classification and accurate Diagnosis. It was noticed that there were two IoMT based Models proposed recently. They are Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT) and Hybrid Random Forest-Linear Model(HRFLM). The above mentioned two Models were Studied thoroughly. Both the Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT) and Hybrid Random Forest-Linear Model(HRFLM) were implemented and carefully analyzed its performances during Training and Testing Processes in terms of Prediction Accuracy, Precision, Sensitivity, Specificity, FScore and Average Processing Time(ms). From the experimental results, it was noticed that the Hybrid Random Forest-Linear Model(HRFLM) is performing well during Training and Testing Processes in terms of Prediction Accuracy, Precision, Sensitivity, Specificity, FScore and Average Processing Time(ms) as compared with Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT).

KEYWORDS: Cardiovascular Diseases; Patterns Prediction; Internet of Medical Things; Random Forest; Machine Learning

I. INTRODUCTION

Health care industries consider that Cardiovascular Diseases are the primary reasons to heart attack and human death in the world. That is the reason why most of the Researchers are focusing Technologies and proposing techniques through ML / DL to predict Cardiovascular Diseases well in advance. This better prediction is mandatory as for as health care data analysis are concerned [2,3,8,10].

From the literature survey, it was noticed that the Cardiovascular Diseases is considered as the prime reasons which leads cause of human death.It is also noticed that around 6,10,000 human beings are losing their life due to Cardiovascular Diseases in US. That is 25% of deaths due to Heart Diseases. As far as Indian statistical reports concerned, there were predicted that in our population, there were 27%of deaths due to Heart Diseases as compared with Global Average 23.5%. [1,2,9,14]. This is the common disease for both Male and Female. Thus, it is the major issue to be addressed. But however, it seems it is really challenging prediction to predict Cardiovascular Diseases in advance as it was observed that there were numerous factors are involving for this disease. Pulse Rate, LDL, HDL, BP and Diabetes are considered as the risk factors.

Artificial Intelligence, Machine Learning and Deep Learning were introduced better prediction by removing noise and extracting useful information from datasets. These techniques are sued to remove unwanted information from the datasets by applying dimensionality reduction techniques. This will help Medical Practioners to take wise decision with better disease prevention treatment. From the literature survey, it was noticed that a few techniques proposed for diagnosing cardiac diseases. [1,2,8,13]. It is also noticed that though numerous methods and techniques proposed for this purpose, we need still better classifiers for the best prediction with highest prediction accuracy.

From the Literature Survey, it was noticed that there were still various intelligent techniques proposed namely pattern recognition, classification and predictive methods for better cardiovascular problems prediction. It is also noted that ML/DL based classifiers like kNN, DT, SVM, RF, ANN and linear and Nonlinear Regression techniques to predict Cardiovascular diseases.As Machine learning is considered as the best approach for better prediction, more researchers

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104077 |

studying and analyzing through the above mentioned techniques [1,2,9,11,13,14]. The approach and the concept and architecture of Machine Learning is shown in the Fig. 1a and Fig. 1b.

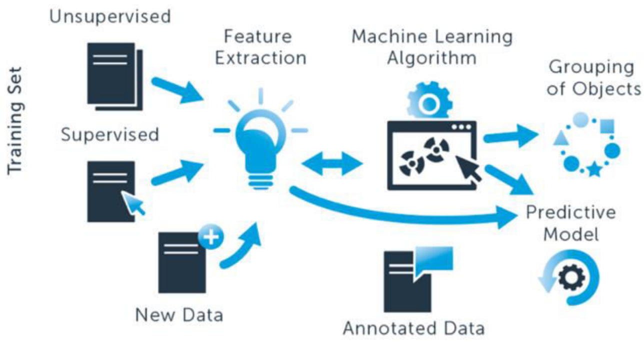


Fig. 1a. Concept of Machine Learning [15]

II. RELATED WORK

Padmaja et al. [11] was proposed an efficient Machine Learning Technique, which is proposed for better cardiovascular diseases patterns. It considered as useful method for Medical Practitioners to understand and predict the actual level of heart diseases.

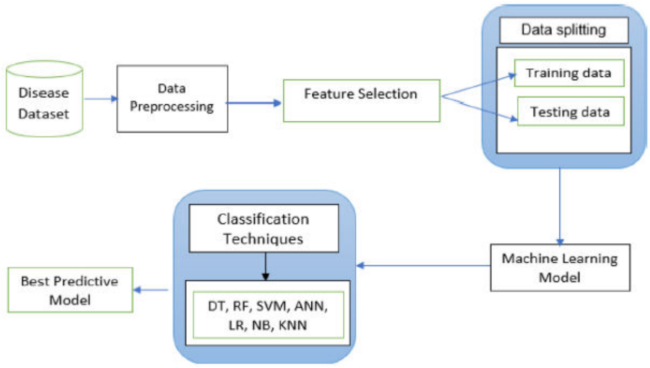


Fig. 1b. Architecture of CVD Pattern Classification / Prediction [15]

Geetha [10] was providing an Machine Learning Technique in association with Artificial Neural Networks (ANN) for predicting Cardiovascular diseases patterns prediction. The author used 13 features and predict the patterns. The author implemented and shown this methods performance and noted the efficiency of the classification accuracy 70%.

Gao et al. [7] has introduced an ML based Classifier which is the hybrid model implemented and tried to achieve better classification accuracy. The Kaggle dataset was considered for implementation and analysis. This Model ensemble both bagging and boosting approaches.

As both PCA and LR Models, it selected required information from datasets for better classification. From the results, it was noticed that the classification accuracy was 98.6%.

Xiaoming Yuan and et. al.[14] designed a Fuzzy GBDT technique which combines Fuzzy and Boosting DT for better classification prediction. It is much suitable to avoid overfitting. It was noticed from the experimental results that it got better classification accuracy.

Mohan et al. [9] has introduced an effective Random Forest Model called Random Forest Hybrid with a Linear Model (HRFLM). This is the integrated model with Linear Model and Random Forest Model. The accuracy was predicted and measured as 88.7%. This work considered the dataset named Hungarian from the database UCI.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104077 |

From the literature survey, it was noticed that the above mentioned IoMT based Models were relatively better. ie we would like to implement and analysis both the Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT) [14] and Hybrid Random Forest-Linear Model(HRFLM) [9] as it was predicted as much suitable of IoMT.

III. Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT)

Bootstrap Aggregation proposed by [14] Xiaoming Yuan and et. al. is used for reducing the DT Variance. For training this technique, the features were taken in random pattern. All the predictions were found average and it is considered as the best robust model.

The learning approach [1,4,14] is obtaining better area under the curve ie AUC and this greatly reduced the variance by bagging technique.

Xiaoming Yuan and et.al. [14] utilized the great ML technique GBDT. The GBDT diagram is as shown in the Fig. 2. From the experimental report, it is clearly noticed that it reduced loss function and achieved better accuracy which is demonstrated in the equation 1.

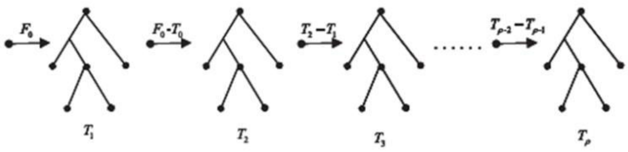


Fig. 2 Schematic diagram of the GBDT Technique14]

The author [14] has implemented this classifier in parallel which is able to predict patterns in parallel. Thus the complexity is reduced in the datasets. It addresses the overfitting issues as well.

Fuzzy-GBDT: Fuzzy Logic Integrates GBDT Algorithm

Data Fuzzification: As we know, if we needed better classification accuracy, that will lead to more process and data complexity. When executing diagnosis processes, we can notice that the prediction report might be the same for different patients whereas all the attributes are common.

This might be lead to complexity. This issue was solved in this model as Fuzzy Logic was employed. The datasets used under hierarchical model and narrates degree for each and every attributes.

As this model employs membership function, its complexity is much reduced [2,7,14]. The Membership function was shown in the Fig. 3.

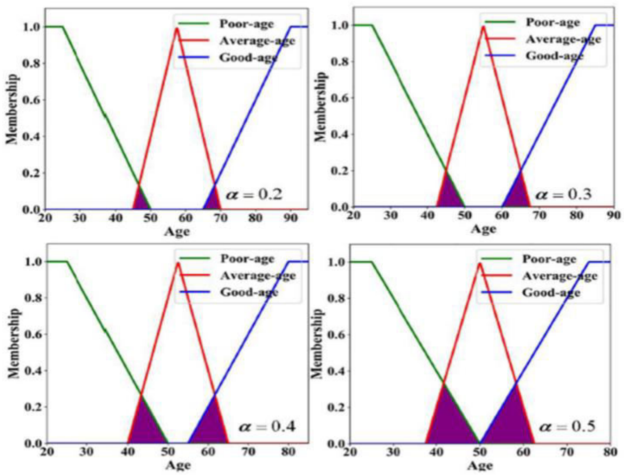


Fig. 3 Effect of different values α in membership [14]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104077 |

The fuzzification with different slope is shown in the Fig. 3. In the data samples, the age distribution is from 0 to 90. The slope can be correlated with age by adjusting α and the data set has 14 attributes.

IV. Hybrid Random Forest-Linear Model(HRFLM)

Mohan et al. [9] has introduced an effective Random Forest Model called Random Forest Hybrid with a Linear Model (HRFLM). This is the integrated model with Linear Model and Random forest Model. These Features took from UCI Dataset is feeding to Linear Method in association with Random Forest is feed again to HRFLM. This will predict the pattern based on the input and will be classified. As mentioned, the classification accuracy was 88.7%.

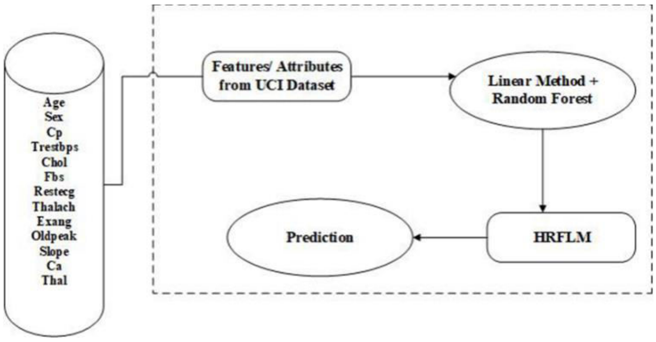


Fig. 4 Heart Disease Prediction with HRFLM[9]

HRFLM[9] approach was used for combining the characteristics of Random Forest and Linear Method (Logistic Regression). HRFLM proved to be quite accurate in the prediction of heart disease. There are three steps in performing the hybrid technique

- Finding out the output probabilities of each model. To implement this Mohan et al using the pred_proba function which gives the target probability in the sequesnce list form. As we know, for each category, the unique probability level are fixed with target variable.
- With the help of log loss function, finding the optimized weight that perfectly combines the two models which has low classification error rate. Log Loss function is a metric which considers prediction uncertainty based on input features and labels.
- Using the optimized weight from the above step combining the two models with the help of weighted average and then performing the prediction. From the results, it was noticed that it achieved better classification accuracy as compared with other classifiers.

V. RESULTS AND DISCUSSION

The dataset from repository of Machine Learning UCI was downloaded[1]. The Heart Disease Dataset namely Switzerland, Hungary, Cleveland were used for performance analysis. The downloaded dataset has 303 records and there were 76 attributes. We considered 14 attributes for Cardiovascular Patterns Prediction.

As mentioned earlier, this project Work has developed with the VC++ Professional 2022 and MSVC Tool that is developed for extracting and validating various patterns of Heart Diseases. The standardized and Cleaned Dataset was feeding to BioWeka to analyze the identified Models. Simulations are carriedout to evaluete the performances and classification and prediction abilities of the above listed our proposed classifiers.

This work considered 10 different Data Sets grouped together for predicting cardiovascular disease patterns, and each group has 50,000 samples out of 5,00,000 samples used for prediction analysis of the existing model. The experiments were executed again and again for measuring the efficiencies of the classifier.

The performances of the identified two classifiers called Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT) and Hybrid Random Forest-Linear Model(HRFLM) were implemented and carefully analyzed its performances during Training and Testing Processes in terms of Prediction Accuracy, Precision, Sensitivity, Specificity, FScore and Average Processing Time(ms).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104077 |

This Research Work is implemented the following two Classifiers to classify and predict various Cardiovascular Disease Patterns. They are

- Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT)
- Hybrid Random Forest-Linear Model(HRFLM)

The above mentioned classifiers are implemented and conducted Simulations for studying the efficiencies of the Classifiers.

As mentioned earlier, the dataset contains 303 records with 14 attributes were considered for performance analysis. From the Fig. 5 and Fig 7, it was noticed that during training and testing processes, the Hybrid Random Forest-Linear Model(HRFLM) is achieving better classification and prediction in terms of True Positive and True Negative Rate as compared with Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT).
It is also observed that False Positive and False Negative Rate of Hybrid Random Forest-Linear Model(HRFLM) is less than that of Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT) during Training and Testing Processes.

From the Fig. 6 and Fig.87, it was observed that during training and Testing processes, the Hybrid Random Forest-Linear Model(HRFLM) is achieving better classification and prediction in terms of Accuracy, Precision, Sensitivity, Specificity, FScore as compared with Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT).
It is also observed that the Average Processing Time(ms)of Hybrid Random Forest-Linear Model(HRFLM) is less than that of Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT) during Training and Testing Processes.

Parameter: Sensitivity (%)

DATA	FGBDT	HRFLM
7	21.57143	22.64957
14	42.34992	43.71267
21	54.10237	56.38179
28	62.94178	65.15769
35	69.3433	71.67413
42	74.59764	77.45564
49	78.9429	82.13671
56	83.00283	86.24155
63	86.31213	90.18695
70	89.91597	93.47826

Parameter: Specificity (%)

DATA	FGBDT	HRFLM
7	21.57143	22.49284
14	42.19059	43.73253
21	54.16948	56.32118
28	63.14408	65.01912
35	69.83502	71.31226
42	75.66714	77.26436
49	80.28235	81.59809
56	84.21694	85.38998
63	87.8331	89.09069
70	91.54519	92.25353

Parameter: Accuracy (%)

DATA	FGBDT	HRFLM
7	21.57143	22.57143
14	42.27107	43.72262
21	54.13565	56.35133
28	63.04215	65.08809
35	69.58608	71.49168
42	75.12102	77.35967
49	79.59747	81.86512
56	83.59892	85.8107
63	87.05702	89.63124
70	90.71429	92.85714

Parameter: Precision (%)

DATA	FGBDT	HRFLM
7	21.57143	22.71429
14	42.78639	43.64353
21	54.54134	56.11277
28	63.42992	64.85849
35	70.21362	71.07076
42	76.18488	77.18488
49	80.72829	81.44257
56	84.50201	85.2163
63	88.08269	88.93984
70	91.71429	92.14286

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104077 |

Parameter: FScore (%)

DATA	FGBDT	HRFLM
7	21.57143	22.68188
14	42.56703	43.67807
21	54.32096	56.24696
28	63.18491	65.00774
35	69.77573	71.37117
42	75.3829	77.32002
49	79.82561	81.78817
56	83.74571	85.72585
63	87.18843	89.55905
70	90.80622	92.80576

Parameter: Average Processing Time (ms)

DATA	FGBDT	HRFLM
7	1711	1146
14	1692	1020
21	1823	1094
28	1831	1082
35	1803	1086
42	1824	1073
49	1814	1111
56	1799	1022
63	1762	1058
70	1735	1046

Fig. 5. Comparative Analysis -Sensitivity, Specificity, Accuracy, Precision, FScore and Average Processing Time(ms) of FGBDT and HRFLM during Testing

As mentioned earlier, this project Work has developed the VC++ Interfacing Tool for extracting and validating various patterns of Heart Diseases. The standardized and Cleaned Dataset was feeding to BioWeka to analyze the identified Models.

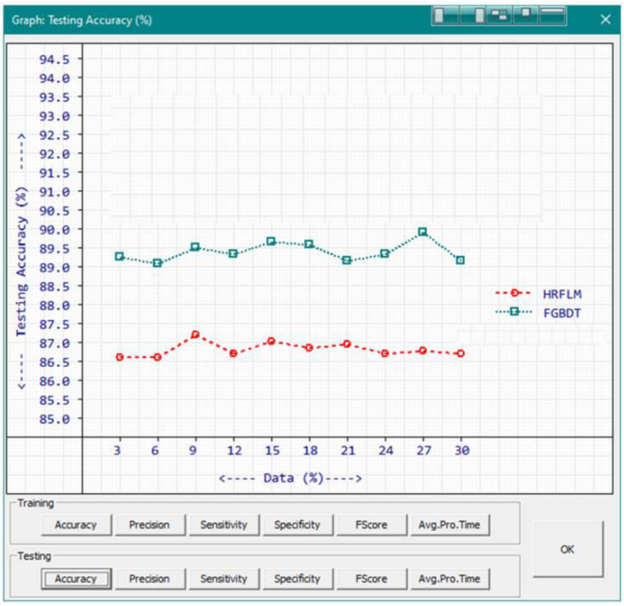


Fig. 6. Comparative Analysis –Prediction Accuracy of FGBDT and HRFLM during Testing

VI. CONCLUSION AND FUTURE WORK

The recently proposed Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT) and Hybrid Random Forest-Linear Model(HRFLM) Classifiers were Studied thoroughly and implemented carefully to analyze its performances during Training and Testing Processes in terms of Prediction Accuracy, Precision, Sensitivity, Specificity, FScore and Average Processing Time(ms). From the experimental results, it was noticed that the Hybrid Random Forest-Linear Model(HRFLM) is performing well during Training and Testing Processes in terms of Prediction Accuracy, Precision, Sensitivity, Specificity, FScore and Average Processing Time(ms) as compared with Bagging-Fuzzy-Gradient Boosting Decision Tree (FGBDT). However, it is predicted that the Hybrid Random Forest-Linear Model(HRFLM) is not suitable for better prediction when Datasets have more different patterns with very less dissimilarities. To address this identified issues, this research work is planned to propose an Intelligent Classifier for better prediction.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104077 |

REFERENCES

1. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, “Classi cation models for heart disease prediction using feature selection and PCA,” Informatics in Medicine Unlocked, Vol. 19, 2020.
2. Lakshmanarao, Y. Swathi, and P. S. S. Sundareswar, “Machine learning techniques for heart disease prediction,” International Journal of Scientific & Technology Research, Vol. 8, No. 11, 2019.
3. Abdallah Abdellatif , Hamdan Abdellatef, Jeevan Kanesan, Chee-Onn Chow, Joon Huang Chuah, and Hassan Muwafaq Gheni, “An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods,” IEEE Access, 2022.
4. Agrahara A. “Heart Disease Prediction Using Machine Learning Algorithms,” International Journal of Scientific Research in Computer Science, Engineering and Information Technology Vol. 6(4), pp. 137–49, 2020
5. Alotaibi FS. “Implementation of machine learning model to predict heart failure disease,” International Journal of Advanced Computer Science and Applications, Vol. 10(6), pp.261–268, 2020.
6. G. Eason, B. Noble, and I. N. and Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, 2021.
7. Gao X-Y, Amin Ali A, Shaban Hassan H, and Anwar EM. “Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method,” Complexity 2021.
8. Ghulab Nabi Ahmad, Hira Fatima, Shafiullah, Abdelaziz Salah Saidi and Imdadullah, “Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV,” IEEE Access, pp. 80151 – 80173, 2022
9. Mohan S, Thirumalai C, and Srivastava G. “Effective heart disease prediction using hybrid machine learning techniques,” IEEE Access, Vol. 7:Pp. 81542–81554, 2019.
10. P. D. C. Geetha S, Kalaivani V, Haritha CJ, and Preetha G. “Prediction Techniques of Heart Disease and Diabetes Disease using Machine Learning,” Turkish Journal of Computer and Mathematics Education Vol. 12(10):3316–25, 2022.
11. Padmaja B, Srinidhi C, Sindhu K, Vanaja K, Deepika N, and Patro EKR. “Early and Accurate Prediction of Heart Disease Using Machine Learning Model. Turkish Journal of Computer and Mathematics Education (TURCOMAT) Vol. 12(6), Pp. 4516–28, 2022.
12. S. Goel, A. Deep, S. Srivastava, and A. Tripathi, ``Comparative analysis of various techniques for heart disease prediction,'' in Proc. 4th Int. Conf. Inf. Syst. Comput. Netw. (ISCON), Mathura, India, pp. 88-94,2019
13. V. A. S. Hernndez et al., “A practical tutorial for decision tree induction: Evaluationmeasures for candidate splits and opportunities,” ACM Computational Survey, Vol. 54, No. 1, pp. 1–38, 2021.
14. Xiaoming Yuan , Jiahui Chen, Kuan Zhang , Yuan Wu and Tingting Yang, “A Stable AI-Based Binary and Multiple Class Heart Disease Prediction Model for IoMT,” IEEE Transactions on Industrial Informatics, Vol. 18, No. 3, 2022.
15. https://alphaxsalt.medium.com/machine-learning-basics-e1d1be0eff1
16. https://github.com/VenkateshBH99/Hybrid-Random-Forest-Linear-Model#readme

BIOGRAPHY

Subhasini SVis a Scholar pursuing M.E. Computer Science and Engineering, KIT- Kalaignar Karunanidhi Institute of Technology, Coimbatore, TN, India.