

Analysis of classification based predicted disease using machine learning and medical things model

¹Hemanta Kumar Bhuyan, ²T.Arun sai, ³M.Charan, ⁴K.Vignesh Chowdary, ⁵Biswajit Brahma
^{1,2,3,4} Vignan's Foundation for Science, Technology & Research (Deemed University), Guntur, AP, INDIA
⁵McKesson Corporation, San Francisco, California, USA
hmb.bhuyan@gmail.com, thulluruarunsai4@gmail.com, mallelacharan987@gmail.com
Vigneshchowdarykorrapati@gmail.com, Biswajit.Brahma@gmail.com

Abstract: Health diseases have been issued seriously harmful in human life due to different dehydrated food and disturbance of working environment in the organization. Precise prediction and diagnosis of disease become a more serious and challenging task for primary deterrence, recognition, and treatment. Thus, based on the above challenges, we proposed the Medical Things (MT) and machine learning models to solve the healthcare problems with appropriate services in disease supervising, forecast, and diagnosis. We developed a prediction framework with machine learning approaches to get different categories of classification for predicted disease. The framework is designed by the fuzzy model with a decision tree to lessen the data complexity. We considered heart disease for experiments and experimental evaluation determined the prediction for categories of classification. The number of decision trees (M) with samples (MS), leaf node (ML), and learning rate (I) is determined as MS=20, ML=3, I=0.1, then mean test score(m) is 20.

Index Terms—Fuzzy set, gradient decision tree, Internet of Medical Things (IoMT), machine learning.

I. INTRODUCTION

Although, different diseases affected the human body as their environment, those diseases couldn't be existed in long-term periods or certain periods. Heart disease is one of the them which is a very harmful for the human body. This is out of expectation of live as its high death and illness rates [1]. This disease is a very fear ness disease because it comes with uncertain times. It can be predicted or detected through the AI model [2]. Thus, these disease affecters need to consult with a physician at an early stage of treatment, otherwise, there is uncertain life due to this type of disease. The electronics-healthcare systems can benefit from AI learning algorithms that can effectively process large amounts of Internet of Medical Things (IoMT) data. IoMT helps to find out the heart disease through various outcomes. The system also minimizes the costs and takes responsibilities of various treatment processing of patient disease data. Moreover, Problems that must be overcome include ensuring that machine learning algorithms and models are accurate, generalize well, and remain stable.

Different disease prediction models have received a lot of attention from scientists working to improve their accuracy. Various approaches such as Logistic regression, K-nearest neighbors, etc. are machine learning classifiers used in the

classification system design [5,6] As a method for classifying cardiac illness, the gradient boosting decision tree (GBDT) has proven to be more effective and relevant [7,8]. Continuous and discrete values are among the various data that it can handle [9].

A binary classification problem is solved by most existing prediction models or algorithms; however, they do not take into account the risk level of heart disease. Five categories, from zero (no presence) to four (severe heart disease), are defined by angiographic results. Each individual's risk was projected by Thomas et al., [10]. However, the accuracy of the original multiclassification system for categorizing various levels of cardiovascular disease risk can still be improved. The reduction of variance and deviation is a common strategy for enhancing algorithm accuracy in machine learning. To improve AUC and reduce variance, the bootstrap aggregating (bagging) approach is added to the learning model [11]. The accuracy of different kind of disease prediction is rarely disclosed in prior studies, which have great data complexity.

Thus, we proposed the framework for classifications of heart disease with various methods with steady and high-accuracy prediction approach. We use lessen variables for methodology with fuzzy logic to improve the performance while decreasing model deviation. We incorporate fuzzy logic into the GBDT algorithm. We also use the bagging algorithm to reduce the model's variance via multiple random sampling and also to improve its stability. Furthermore, our method can forecast the severity of cardiac disease more accurately than previous algorithms, which can only predict whether patients are unwell. The following is a list of our major contributions.

- (a) Fuzzy-GBDT-based approach is proposed to detect heart disease with lessening the complexity.
- (b) Enhancing GBDT's capability to learn from new data sets.
- (c) Fuzzy-GBDT and bootstrap aggregation are combined to avoid overfitting.
- (d) This new version of the Fuzzy-GBDT reduces the prediction model's variance and deviation, resulting in better accuracy and stability.
- (e) To improve the diagnosis and treatment of different illness kinds, we developed the Bagging-Fuzzy-GBDT multiple classifications, prediction models.

- (f) During the experiments, the model was found to be very accurate and stable in its ability to predict cardiac disease.

The rest of this article is arranged in the following manner. The background of this paper is presented in Section II. In section III, we developed the architecture using medical things. for the system. Bagging-fuzzy-GBDT prediction has been implemented into the proposed algorithm in Section IV. The simulation findings are discussed and explained in Section V. The paper is concluded in Section VI.

II. BACKGROUND

Over the past few decades, heart disease prediction has received a lot of attention. Heart disease diagnosis can be accurately predicted using a variety of methods [3–8]. Information on cardiac illness is extremely complicated. Limitations exist in the use of data mining techniques to forecast cardiac disease. Reducing the number of attributes would simplify and improve the system. Heart disease diagnosis was complex because of the use of four feature selection methods and six classification algorithms [6]. Jabbar et al. [12] provided a genetic algorithm-based associative classification. Using a genetic algorithm for disease prediction may yield the most accurate set of attributes based on the quantity of the actual data. Using a decision tree to train classification algorithms is a well-known method. All of which were developed for specific purposes. Decision trees, neural networks, and fuzzy based feature selection by Bhuyan et al. [13]. The decision tree was found to be the most accurate in predicting heart disease. Using a decision tree to process continuous data is tricky. Using hybrid learning procedures, cardiac disease is diagnosed [14]. Heart disease was predicted using seven different data mining techniques [14], including SVM, NB, as well as the multilayer perceptron and deep neural network.

It was shown that SVM utilizing the improvement strategy was the best identification method than others. Data complexity and accuracy can be improved by using the fuzzy logic method [15]. Fuzzy ANN was used to estimate the likelihood of cardiovascular illness [16,21]. However, the problem of overfitting in disease prediction should be taken into consideration. To acquire the results of heart disease prediction using binary classification, it needs to have the illness. Sometimes, Doctors are unable to aid for treatment of heart disease because the risk level is not provided. The prediction of multiple cardiac disease classifications is also important. Additional risk variables for coronary artery disease, such as coronary calcium score, were investigated by Polonsky et al. [17]. Few categorized in terms of future cardiovascular events are risk. The results of the model's evaluation have improved the risk classification. From above analysis of related work and challenging task, we have proposed the framework for monitoring system using medical things as section III.

III. FRAMEWORK OF MONITORING SYSTEM THROUGH MEDICAL THINGS

We proposed the framework for processing of disease data through various components using medical things with wireless connectivity as shown in fig.1. Different components are helped

to detect disease as per prior data processing. Doctors and other authorized staff could view the health monitoring data at any time and from any location. We use machine learning and other AI technologies to train health data for detecting the disease and make treatment accordingly. It's possible to greatly increase the accuracy and stability of the prediction model. The diagnosis framework is built based on the well-known GBDT technique, which is based on machine learning. We considered Gradient boosting is part of GBDT for the decision tree which is a weak classifier due to sequential training. We have created new decision tree to reduce residuals. The loss function's negative gradient value is used to approximate the residual in the GBDT technique. The eq. 2 is considered for creating GBDT algorithm where F_0 represents the starting value and T represents the decision tree that has been traversed. GBDT generates a prediction model consisting of a collection of mediocre predictions.

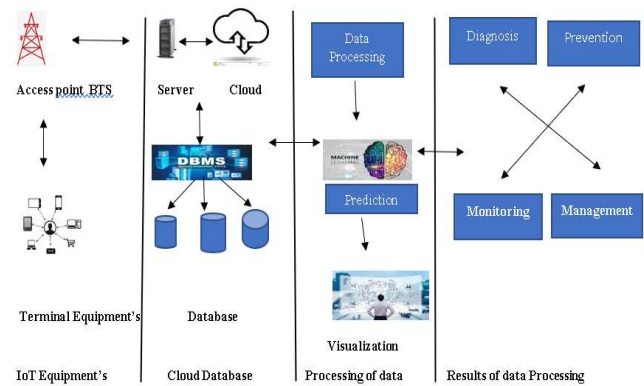


Fig. 1: Medical Thing based architecture

The regression problem is simulated using GBDT and the present model's negative gradient loss function. GBDT uses this value. It is a measure of the model's trustworthiness that the loss function is used to find out the error during evaluation of data. The less error creates more accuracy in experiments. The equ. (1) is defined with weight coefficients which are denoted by a, b, c, \dots , in the expression (1). To conclude, it combines the findings of all of the preceding trees. By including the sigmoid function, the GBDT algorithm find the solution for both regression and classification issues.

$$F(x) = aF_0(x) + bF_1(x) + cF_2(x) + \dots + \rho F_p(x) \quad (1)$$

IV. DIFFERENT CLASSIFICATION METHODS BASED DISEASE PREDICTION

We considered the finest algorithms for fitting in classical approaches of machine learning algorithms as GBDT. In addition to predicting cardiac disease, GBDT can process a wide range of data. Two categories of classification such as Binary and multiple classification techniques for detecting the heart disease are implemented in this section. To begin, we simplify the data on heart illness by applying the fuzzy logic method. Prediction of cardiac disease using the Fuzzy-GBDT algorithm is proposed. The Bagging-Fuzzy-GBDT algorithm is also coupled with a bagging algorithm to avoid overfitting and produce multiple classification predictions.

A. Fuzzy Logic based GBDT Algorithm

1) Fuzzy data set: The intricacy of the data frequently affects the accuracy of the predictions and diagnoses. Patients' diagnostic outcomes are often the same as per wide variations in data value when grouped under the same criterion. It creates the difficulty of analyzing disease data. Lessening difficulty and improve data processing, we considered fuzzy logic to design GBDT algorithm for both the precision and generalizability of the prediction model. Our fuzzy sets are considered as hierarchical data objects that define how closely a data object fits an attribute. We have taken membership functions for considering minimize the complexity of a fuzzy set as [15]. When it comes to data fuzzification, the most important consideration is the role played by the membership level. Membership is developed by mathematical approaches with appropriate function which has a degree range of 0 to 1.

Vital clinical records are examined in this paper for diagnosis and severity assessment of harshness disease. The membership function is defined with the help of the triangle function. The optimum values of the data interval depend on the membership degree. Here, we considered μ which is converted into three membership degrees (μ_1, μ_2, μ_3), we define $\sigma = \frac{b-a}{3-2\alpha}$, ($0 < \alpha < 1$), where α is the degree of subintervals. Let $\theta = a + (1 - \alpha)\sigma$, Different equations are formulated as per above data as follows:

$$\mu_1 = \begin{cases} 0, & x > a + \alpha \\ \frac{a+\sigma-x}{\sigma}, & a \leq x \leq a + \sigma \\ 1, & x < a \end{cases} \quad (2)$$

$$\mu_2 = \begin{cases} 0, & x < \theta \text{ or } x > a + (2 - \alpha)\sigma \\ \frac{x-\theta}{\sigma/2}, & \theta \leq x \leq \theta + \sigma/2 \\ \frac{a+(2-\alpha)\sigma-x}{\sigma/2}, & \theta + \frac{\sigma}{2} < x < \theta + \sigma \end{cases} \quad (3)$$

$$\mu_3 = \begin{cases} 0, & x < a + (2 - 2\alpha)\sigma \\ \frac{x-\sigma}{\sigma}, & a + (2 - 2\alpha)\sigma \leq x \leq b \\ 1, & x > b \end{cases} \quad (4)$$

To see how different values affect data fuzzification outcomes. The age of an attribute is fuzzed based on the regularity of the data samples, with a range of 0 to 90 for this attribute. Accuracy in age subinterval cohesion correlates positively with the number. Data about heart disease is broken down into 14 different features. In Section V, Readers can find further information on the dissemination of data. We apply the algorithm of [18] for our proposed work as follows.

Algorithm 1: Fuzzy decision tree for Classification.

Input: Data set D, T - iterations, loss function L, fuzzy feature A_i , indicator function I

Output: Strong learner $f_T(x)$

1: Find max of (μ_1, μ_2, μ_3) $\rightarrow (\alpha_i, \beta_i, \gamma_i)$ and assign to B_i

2: Now B_i assigns to A_i

3: Start with $f_0(x)$

4: for $t \in [1, T]$ do

5: for $i \in [1, N]$ do // determine negative gradient y_i

6: $y_i = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]$ $f(x) = f_{(t-1)}(x)$

7: end for

8: Make a regression tree for target c_{ij} with terminal regions R_{ij} , $j = 1, 2, \dots, J$

9: for each $j \in [1, J]$ do (Determine the best fit c_{ij} for the leaf node j)

10: $c_{tj} = \arg \min_c \sum_{x_i \in R_{tj}} \log(1 + \exp(-y_i(f_{t-1}(x_i) + c)))$

11: end for

12: Update $f_t(x) = f_{t-1}(x) + \sum_{j=1}^J c_{tj} I(x \in R_{tj})$

13: end for

2) Fuzzy decision tree for Classification Algorithm

Description: The algorithm 1 maintain fuzzy parameters α_i, β_i , and γ_i with the three intervals of i th feature A_i . We developed the membership functions as (μ_1, μ_2, μ_3), respectively.

Let the training set $D = \{(x_i, y_i)\} i = 1 \dots n$, and start with weak

Fuzzy-GBDT $f_0(x)$ in step 3 as $f_0(x) = 0.5 * \log\left(\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n 1 - y_i}\right)$. In

step 6, we used the cross-entropy loss function as L, where $p_i = \frac{1}{1 + e^{-f(x_i)}}$, so

$$L(y_i, f_i(x)) = -\{y_i \ln p_i + (1 - y_i) \ln(1 - p_i)\} \quad (5)$$

$$y_i = -\left[\frac{L(y_i, f(x_i))}{f(x_i)}\right] \quad (6)$$

where J is leaf nodes and the indicator function $I \in [0, 1]$. In last, the strong learner is formulated as equ (7) as follows.

$$f_T(x) = f_0(x) + \sum_{t=1}^T \sum_{j=1}^J c_{tj} I(x \in R_{tj}). \quad (7)$$

B. Explanation of Bagging-Fuzzy-GBDT Algorithm for Binary Classification

The generalizability of the Fuzzy-GBDT algorithm may be enhanced. GBDT, on the other hand, is highly sensitive and cannot efficiently analyze large amounts of medical data. Parallel integrated learning is a assembly learning technique known as "bagging". Bagging removes the interference of a single sensitive spot by conducting several self-samplings. When it comes to classifier outputs, it can successfully evade overfitting lessening the variation.

(1) **Bagging-Fuzzy-GBDT Algorithm:** Data complexity is reduced and the result is standard data. Many different week classes are created using the bagging method. It is possible to achieve excellent adaptability to the integrated Bagging-Fuzzy-GBDT algorithm. Prediction and diagnosis of cardiac disease can be done with complete confidence to the high level of accuracy and consistency in place. The following are the detailed steps that must be followed to complete the task.

(1) Assume that T is a training set with a size of N.

(2) If you have a large enough training set, we can substitute the subsets T_i , $i = (1, 2, \dots, m)$ with it.

(3) Train the classifiers with the sampled subset T_i once it has generated using the Fuzzy-GBDT classifier.

(4) Fuzzy-GBDT classifiers can be generated by repeating steps 3 and 4.

(5) Classifier m weak learners vote for the best Bagging-Fuzzy-GBDT classifier.

(2) Complexity of Algorithms: There are N leaves in the tree, a vector of leaf scores is generated as per nodes. A data dimension d that determines the definition of the tree $f(x)$. The GBDT algorithm's time complexity is dictated by the regularization term $\Omega(f_t)$. The following steps could lead to it:

$$\Omega(f_t) = \gamma N + \frac{1}{2} \lambda \sum_{j=1}^N \bar{w}_j^2 \quad (8)$$

As a general rule, it is impossible to list every potential tree structure. As a result, we have switched to a greedy algorithm. It begins at a single leaf and grows the tree from there iteratively. The time complexity of sorting of a single tree has an $O(n \log n)$. The depth of the tree determines the difficulty of GBDT. Due to the triangle membership function used by fuzzy logic, data complexity is reduced, but GBDT's technique is unaffected. There is no difference in complexity between Fuzzy-GBDT and GBDT. In binary classification, the strongest Fuzzy-GBDT classifier is voted on by numerous weak Fuzzy-GBDT classifiers concurrently with time complexity of $O(dn \log n)$.

V. EXPERIMENTAL EVALUATION

Our Bagging-Fuzzy-GBDT model is put up against a variety of other classification models in this section. The results of the study show that the algorithm provided is valid for detecting heart disease.

A. Dataset.

The open-source dataset on cardiovascular disease from the University of California, Irvine (UCI) is used in this paper [19]. A total of 836 records are included in this database, which includes 14 critical features. Other complete databases' average values are used to fill up the gaps. For this study, the dataset is separated into two parts: a training set and a test set.

B. Evaluation Metrics

The suggested model's performance is assessed using five metrics, including the ROC curve, AUC, accuracy, precision, and F1. A receiver operating characteristic (ROC) curve has horizontal axes for false-positive rates and vertical axes for genuine positive rates [20]. According to this, the classifier's accuracy is compromised by the amount of false-positives it incorrectly classifies. ROC curve area and the circle around the axis of coordinates are referred to as AUC. Accuracy is primarily measured as the proportion of correctly identified using the following formula:

$$\text{Accuracy} = \frac{TS+}{TS+TN+FS+FN} \quad (9)$$

There are types of people in this equation: those with heart disease and those who are not; those who are predicted to have the disease but are not; these are known as true positives and

negatives, respectively. Precision and recall can be quantified using F1's thorough description of the accuracy and recall. Following is a list of all the equations:

$$\text{Precision} = \frac{TS}{TS+FS} \quad (10)$$

$$\text{Recall} = \frac{TS}{TS+F} \quad (11)$$

$$F1 = 2 * \frac{\text{Precision} + \text{Re}}{\text{Precision} * \text{Recall}} \quad (12)$$

C. The Bagging-Fuzzy-GBDT Algorithm Parameter Determination.

Six parameters [9] are to be determined for the Bagging-Fuzzy-GBDT algorithm. The proposed prediction model's stability and accuracy are directly influenced by the values of these parameters. It is therefore critical to figure out how to get the best results.

- (i) A decision tree has the number M of branches. Each iteration of the Bagging-Fuzzy-GBDT growth process produces a decision tree. As the number of decision trees in a training set grows, so does the accuracy of the predictions.
- (ii) Maximum depth (MD) is the optimum nodes in a decision tree. To put it another way, MD's value cannot be too high or too low. It takes longer to train each tree if the MD is too large, hence the process takes longer.
- (iii) To split an internal node, three samples are needed at the very least. MS's value is influenced by two factors such as int and float. Each node has a minimum sample size of (MS MD).
- (iv) ML samples must be present at a leaf node to do the analysis. ML training samples must be left in both the left and right branches of any splitting point to be evaluated. This improves the fit of the model, particularly in regressions.
- (v) There are m samples to be bagged. Appropriate m values can assist improve prediction accuracy.
- (vi) To prevent the model from overfitting, the regularization method is known as Learning rate $I(0 < I < 1)$ is employed to limit the model's reliance on a single decision tree i.e., Bagging-Fuzzy-GBDT.

Additionally, we used a grid search for parameter selection in the Bagging-Fuzzy-GBDT method which perform in a significant increase in instability. To find the best possible parameter combination based on cross-validation scores, grid search makes use of the cross-validation approach. This is the highest score that the testing set achieved during the optimization process. Because of this, we utilize the average test score to determine which parameters are best. Before performing the grid search to discover the best M , we left the other parameters alone to estimate the number of decision trees that should be used. We considered 35 decision trees for experiments, and optimum at mean test score. Second, grid search is used to optimize MD, which has a fixed value of 35 and the rest is the default value. It's demonstrated that the average test score for the various maximum depths of each decision tree.

The mean test score reaches its greatest point at MD=6. Measuring the mean test score lowers consistently when the MD is bigger than 6. As a result, the number 6 is the most suitable. The following are the outcomes: All of these numbers are in various parameter values such as MS = 20, ML=3, I=0.1, and m = 20.

Table I (a): Values before grid search

Experiment No.	Accuracy	Precision	Recall	F1
1	0.87	0.85	0.89	0.94
2	0.87	0.85	0.89	0.94
3	0.88	0.86	0.90	0.94
4	0.87	0.85	0.89	0.94
5	0.875	0.86	0.893	0.94
6	0.873	0.858	0.891	0.94
7	0.87	0.84	0.89	0.94
8	0.877	0.86	0.895	0.94
9	0.87	0.85	0.89	0.94
10	0.87	0.857	0.897	0.94

Table I (b): Values after grid search

Experiment No.	Accuracy	Precision	Recall	F1
1	0.902	0.90	0.92	0.94
2	0.902	0.90	0.92	0.94
3	0.902	0.90	0.92	0.94
4	0.902	0.90	0.92	0.94
5	0.902	0.90	0.92	0.94
6	0.902	0.90	0.92	0.94
7	0.902	0.90	0.92	0.94
8	0.902	0.90	0.92	0.94
9	0.902	0.90	0.92	0.94
10	0.902	0.90	0.92	0.94

It's possible to see how evaluation markers have changed over time in Table I(a). According to the graph, the precision and accuracy are always fluctuating, with only the recall rate being constant. The model's stability has increased significantly since the optimization as shown table II (b).

D. The Algorithm's Performance Test

Binary Classification's performance evaluation: We have used five prediction models to test the proposed models' rationale. Comparisons are made between Bagging-Fuzzy-GBDT and other models as well as a decision tree, GBDT, and Bagging-Fuzzy-GBDT. Using grid search, we found that the Bagging-Fuzzy-GBDT model is best at predicting heart disease, and we've listed the metric values of five other models in fig2.

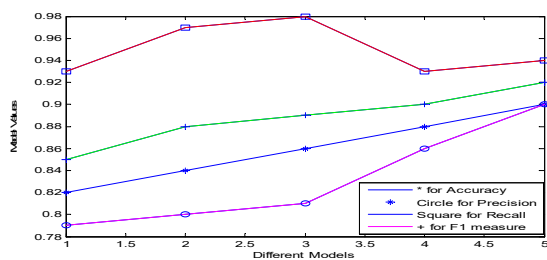


Fig 2: The Indicator Values Of Different Models

With the Bagging-Fuzzy-GBDT technique, accuracy and precision are greatly improved. In the first place, GBDT outperforms the typical decision tree prediction model when it

comes to predicting heart disease. As far as accuracy and precision go, the Bagging-FuzzyGBDT technique is the most similar to the other four approaches. The number of incorrect predictions is what determines the recall value. To put it another way, recall is affected by precision. It isn't always beneficial to have a high value for recall. To ensure that the prediction model is evaluated fairly and objectively, we introduce F1. The prediction model's accuracy improves with increasing F1 value. Bagging-Fuzzy-GBDT has the greatest F1 value in fig. 2., demonstrating that it is the most successful and stable method. A comparison is made between the Bagging-Fuzzy-GBDT approach and other current academic approaches. According fig 3, a proposed Bagging-Fuzzy-GBDT outperforms the literature.

Table II: AUC of different models

S.No.	Models	AUC values
1	Decision Tree	0.8
2	CBDT	0.82
3	Bagging CBDT	0.84
4	Fuzzy CBDT	0.87
5	B F GBDT	0.9

Table II shows the ROC and AUC evaluation results for various modes (b). The Bagging-Fuzzy-GBDT method outperforms the decision tree and GBDT in terms of speed and efficiency. In terms of ROC, their AUC values are 0.8 and 0.82. If you look at Table II, it can see that our proposed Bagging-Fuzzy-GBDT technique has a 0.9 AUC. This further demonstrates the model's superior overall performance. The AUC value tends to rise with the addition of the bagging method and fuzzy logic, respectively. Using these two strategies, we have improved the GBDT algorithm's accuracy. A great degree of stability and accuracy can be found in both theory and experiment using Bagging-Fuzzy-GBDT.

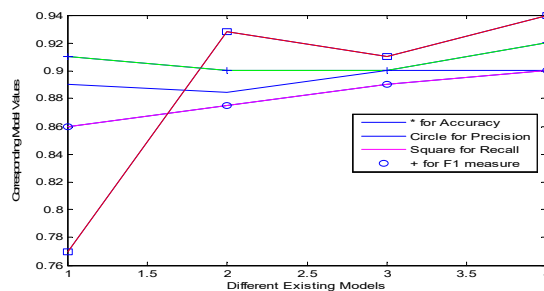


Fig. 3: Performance Comparison Results of Several State-Of-The-Art Works

A multiclassification prediction of the disease's type 1, type 2, and type 3 and 4 based on angiography is made to further identify a given disease type. Each prediction type's average accuracy is calculated over ten studies. It depicts the details, with C-P denoting the number of right forecasts and I-P denoting the number of incorrect guesses. The vast majority of predictions are accurate for each category. To conduct the tests, 135 sets will be available, of which 55 will be type 1 and 41 will be type 2. As a result, type 4 data has a higher threshold for error. Each form of categorization prediction accuracy is displayed in table III for comparison (b). According to the experiment, the Bagging-Fuzzy-GBDT algorithm performs exceptionally well on multi-classification implementation, with an accuracy rate ranging from 80 to 95 percent. For the most

accurate forecast, type 2 has a 93% success rate on average. Type 1 and type 3 both have 85% accuracy in predicting the future. As a result, the diagnosis may be made more quickly and correctly, and patients are given alternative treatment options depending on the kind. However, comparing the accuracy of individual classifications for multiclassification is a waste of time and resources. The overall accuracy of the multiclassification model should be examined, which makes more sense when evaluating the model. Out of the 135 testing data, 118 are predicted correctly. Overall, Bagging-Fuzzy-multiclassification GBDT's prediction is 87.4% accurate, according to our results.

Table III: Accuracy of different types of heart disease. (a) GBDT. (b) Bagging-Fuzzy-GBDT. (c) LDBN. (d) Comparisons.

S.No.	Types of Disease	GBDT Accuracy	B-F-GBDT Accuracy	LDBN Accuracy	Comparisons. Accuracy
1	Type 1	0.85	0.85	0.75	0.65
2	Type 2	0.9	0.93	0.83	0.71
3	Type 3	0.7	0.85	0.72	0.74
4	Type 4	0.8	0.8	0.68	0.83
5	Total	0.83	0.87	0.74	0.87

Type 1 has a precision of 0.9, whilst type 3 has a precision of only 0.7. LDBNs are also evaluated for their ability to predict several types of heart disease categorization. Heart disease prediction is also valued by similar prediction technologies like deep belief networks and backpropagation neural networks. As shown in Table III, different types of heart disease performed better predictions. Above this illness prediction and determining heart disease severity could be transformed into binary and multiclass classification problems, respectively. In addition, the Bagging-fuzzy-GBDT algorithm could achieve excellent accuracy and steadiness in binary and multiclassification.

VI. CONCLUSIONS

As per our proposed framework, the Bagging-Fuzzy-GBDT technique is used for the prediction and detection of cardiac disease in IoMT that is both steady and accurate. The proposed Bagging-Fuzzy-GBDT method was able to predict heart disease in both binary and multiple classifications. For better performance, we used fuzzy logic and bagging algorithms. After the grid search parameters were found, the model's stability was greatly improved. The performance of the evaluation recognized the proposed model outperforms other traditional algorithms in terms of accuracy with other evaluation items and other indicators. Predicting disease with great accuracy is only part of what the Bagging-Fuzzy-GBDT algorithm can do. The e-healthcare can make use of this technology to improve patient diagnosis and care. To improve the proposed model, we will collaborate with regional hospitals to implement real time healthcare data in future.

REFERENCES

- [1] J. Henriques *et al.*, "Prediction of heart failure decompensation events by trend analysis of telemonitoring data," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 5, pp. 1757–1769, Sep. 2014.
- [2] G. Yang *et al.*, "Homecare robotic systems for Healthcare 4.0: Visions and enabling technologies," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2535–2549, Sep. 2020.
- [3] Y. Cheng, H. Zhu, J. Wu, and X. Shao, "Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 987–997, Aug. 2018.
- [4] Bhuyan H. K., Ravi V. K., Analysis of Sub-feature for Classification in Data Mining, IEEE Transaction on Engineering Management, 2021 (Early Published).
- [5] R. Tao *et al.*, "Magnetocardiography based ischemic heart disease detection and localization using machine learning methods," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1658–1667, Jun. 2019.
- [6] Bhuyan, H. K. and Reddy C. V. Madhusudan: Sub-feature selection for novel classification, International Conference on Inventive Communication and Computational Technologies (ICICCT), IEEE Explore, 20-21 April, 2018.
- [7] Z. Wen *et al.*, "Exploiting GPUs for efficient gradient boosting decision tree training," *IEEE Trans. Parallel Distrib. Syst.*, vol. 31, no. 12, pp. 2706–2717, Dec. 2019.
- [8] Bhuyan H. K., Chakraborty C, Shelke Y, Pani S. K., COVID-19 diagnosis system by deep learning approaches, Expert Systems, Volume 39, Issue 3, pp. 1-18, July 2021.
- [9] B. Zhang *et al.*, "Health data driven on continuous blood pressure prediction based on gradient boosting decision tree algorithm," *IEEE Access*, vol. 7, pp. 32423–32433, 2019.
- [10] J. Thomas and R. T. Princy, "Human heart disease prediction system using data mining techniques," in *Proc. Int. Conf. Circuit, Power, Comput. Technol.*, Nagercoil, India, 2016, pp. 1–5.
- [11] R. Khosla, S. Fahmy, Y. C. Hu, and J. Neville, "Predicting prefix availability in the internet," in *Proc. IEEE Int. Conf. Comput. Commun.*, San Diego, CA, USA, 2010, pp. 1–5.
- [12] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Heart disease prediction system using associative classification and genetic algorithm," *Comput. Sci.*, vol. 1, pp. 183–192, Dec. 2012.
- [13] Bhuyan, H.K., Mohanty M., Das S. R.: "Privacy preserving for feature selection in data mining using centralized network". International Journal Computer Science and Issues (IJCSI), vol(9), issue(3), pp. 434 - 440, April.2012,
- [14] S. Pouriyeh *et al.*, "A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease," in *Proc. IEEE Symp. Comput. Commun.*, Heraklion, Greece, 2017, pp. 204–207.
- [15] P. Ziemba, "NEAT F-PROMETHEE—A new fuzzy multiple criteria decision making method based on the adjustment of mapping trapezoidal fuzzy numbers," *Expert Syst. Appl.*, vol. 110, pp. 363–380, Nov. 2018.
- [16] Bhuyan, H. K., Kamila, N. K., and Dash, S. K.: "An Approach for Privacy preservation Of Distributed Data In Peer-To-Peer Network Using Multiparty Computation", International Journal Computer Science and Issues (IJCSI), vol(3), issue(8), pp.424-429
- [17] T. S. Polonsky *et al.*, "Coronary artery calcium score and risk classification for coronary heart disease prediction," *JAMA*, vol. 303, no. 16, pp. 1610–1616, Apr. 2010.
- [18] 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, March 2022.
- [19] *UCI Machine Learning Repository-Heart Disease Data Set*. Accessed: Feb. 10, 2021. [Online]. Available: [http://archive.ics.uci.edu/ml/datasets/Heart Disease](http://archive.ics.uci.edu/ml/datasets/Heart+Disease)
- [20] X. Yuan *et al.*, "A high accuracy integrated bagging-fuzzy-GBDT prediction algorithm for heart disease diagnosis," in *Proc. IEEE/CIC Int. Conf. Commun. China*, Changchun, China, 2019, pp. 467–471.
- [21] Bhuyan, H. K., Pani, S. K., Impact of world pandemic "COVID-19" and an assessment of world health management and economics, The Internet of Medical Things: Enabling Technologies and Emerging Applications, IET, pp-55, 2022.