2021 6th International Conference for Convergence in Technology (I2CT)
Pune, India. Apr 02-04, 2021

Heart Disease Prognosis Using Machine Learning Classification Techniques

Mohammed Nowshad Ruhani Chowdhury[1, a], Ezaz Ahmed[2, a], Md. Abu Dayan Siddik[3, a, b], Akhlak Uz Zaman[4, a]
[a]*Dept. of Computer Science and Engineering, Metropolitan University, Sylhet, Bangladesh*
[b]*Inst. of Information and Communication Technology, Shahjalal University of Science and Technology, Sylhet, Bangladesh*
[1]nowshad.cse@gmail.com, [2]ahmedezaz.cse@gmail.com, [3]dayansiddik94@gmail.com, [4]akhlak.ashik@gmail.com

2021 6th International Conference for Convergence in Technology (I2CT) | 978-1-7281-8876-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/I2CT51068.2021.9418181

***Abstract*— Heart Disease is one of the fatal causes of worldwide death in also third world countries like Bangladesh. Though heart disease prediction with a satisfactory accuracy level is a very demanding and challenging topic, it is achievable using advanced machine learning (ML) techniques. Manufacturing a proper ML system can not only predict cardiovascular disease with high accuracy but also can reduce human intervention, the requirement of extra medical tests. Quick prediction can deduct the death rate and severity of the disease. This paper describes our proposed methodology of predicting heart disease that targets a goal of finding important features by applying ML algorithms resulting in improving the accuracy of the prediction. Instead of collecting the dataset from the online repository, we collected data from the Sylhet region of Bangladesh by physically going to door-to-door hospitals and healthcare industries to make an appropriate questionnaire and to produce the most valuable dataset related to heart disease prediction. Our dataset consists of 564 instances and 18 attributes. We trained our model using classification algorithms like Decision Tree, Logistic Regression, K-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machine (SVM), etc. Though accuracy for different algorithms varies for a different number of instances in the dataset, SVM yielded the best performance with an accuracy level of 91% for the threshold instances of the dataset in our proposed system.**

Keywords*— heart disease prediction; cardiovascular disease prediction; heart disease prognosis; machine learning; classification algorithms; decision tree; logistic regression; k-nearest neighbors; Naïve Bayes; support vector machine.

I. Introduction

Worldwide mortality caused by heart disease is breathtaking if we investigate the present time statistics. According to the World Health Organization (WHO) fact sheets, Of the 56.9 million passing worldwide in 2016, the prominent part (54%) were because of the top 10 causes. Ischaemic (or ischemic) heart disease, known as coronary artery disease (CAD) or coronary heart disease, and stroke are the world's prime executioners, representing a combined 15.2 million passing in 2016 [1]. These illnesses have remained the prominent sources of death around the world over the most recent 15 years [1]. Bangladesh is also very vulnerable to heart disease. According to the World Health Organization - Noncommunicable Diseases (NCD), Bangladesh Profiles of 2018, NCDs account for 67% of all estimated deaths, and cardiovascular disease (CVD) alone accounts for 30% of estimated deaths in Bangladesh among all the NCDs [2].

Bangladesh likely has the most elevated paces of CVD but is the least studied among all South Asian nations. Bangladesh is a nation that is 'missing in action' in the worldwide battle against CVD [3]. As worldwide death and sickness due to CVD is increasing day by day, cardiovascular disease prediction becomes one of the most crucial subjects in the section of clinical data analysis. The data produced regarding heart disease is humongous in the healthcare industry. Machine Learning (ML) is compelling in helping with settling on choices and predictions with a satisfactory level of accuracy from the vast amount of information delivered by the medical services industry.

This paper describes our proposed methodology of predicting heart disease in the Sylhet region of Bangladesh that targets a goal of finding important features by applying ML algorithms resulting in improving the accuracy of the prediction. Setting an appropriate questionnaire is the most crucial and the first most step to build a satisfactory heart disease prediction model based on machine learning. We prepared our questionnaire according to the guideline and most valuable suggestions of the Head of the Department of Cardiology of Sylhet MAG Osmani Medical College as well as the Clinical & Interventional Cardiologist of Mount Adora Hospital, Akhalia, Sylhet honorable Dr. Muhammad Shahabuddin and the Assistant Professor of the Department of Cardiology of Sylhet MAG Osmani Medical College as well as the Cardiology Specialist of Mount Adora Hospital, Nayasarak, Sylhet honorable Dr. Md. Anamur Rahman Anam [4]. Dr. Md. Anamur Rahman Anam suggested 36 attributes for our questionnaire, and Dr. Muhammad Shahabuddin reduced those attributes into 18 and identified eight among them as the most vital components. A total of 564 people contributed to data collection based on the questionnaire by filling up a hard copy of the questionnaire. Among these 564 people, heart patients were 313 in total, and healthy people were 251 in total. We collected data of heart patients from six different medical institutions in Sylhet. They are the following: 1. Jalalabad Ragib-Rabeya Medical College and Hospital, Sylhet, 2. Mount Adora Hospital, Sylhet, 3. North East Medical College and Hospital, Sylhet, 4. Sylhet Women's Medical College and Hospital, 5. National Heart Foundation Hospital, Sylhet, 6. Sylhet MAG Osmani Medical College Hospital. But, we collected healthy people's data from educational institutions, relatives, and other people of the Sylhet region.

After collecting data manually, we input the manual data by filling up the survey form on the website built by us to store them into MySQL format [5]. After that, we checked the noise data and eliminated them from our dataset. Then we converted this dataset from MySQL to CSV file format. After that, we split the dataset into two separate sets: training and testing. Then we used these split datasets for machine learning (ML) implementation using the python programming language. We applied five classification ML algorithms to train the proposed model, cross-validate the train data set using the test set, find useful insights, visualize the results, and predict heart disease by generating the best possible accuracy. The algorithms are the following: 1. Decision Tree, 2. Logistic Regression, 3. K-Nearest-Neighbor (KNN), 4. Naive Bayes, 5. Support Vector Machine (SVM). We found that in our proposed model, SVM yields the best performance with an accuracy level of 91% for the threshold instances of the dataset in our proposed system.

The rest of the paper describes the whole research work elaborately immediately after mentioning some related works done earlier on the same topic.

II. RELATED WORK

There is enormous related work done in anticipating coronary illness utilizing machine learning or data mining procedures. Artificial Neural Network (ANN) has been familiar with producing the most raised prediction with exactness in the clinical area [6]. R. Das, I. Turkoglu, and A. Sengur [7, Sec. 2.2] analyzed a dataset taken from the Cleveland heart disease database. They utilized the backpropagation multilayer perceptron (MLP) of ANN to foresee heart disease. They contrasted the results with the results of existing models within the same sector, noticed improvement by their proposed method, and got 89.01% accuracy from the analyses. S. Mohan, C. Thirumalai, and G. Srivastava [8] assembled an AI model to anticipate coronary illness consolidating the qualities of Random Forest (RF) and Linear Method (LM). They got an accuracy level of 88.7% through the prediction model. To reduce the number of medical tests needed to be taken by a patient, Lessening redundant features and attributes is necessary. M. A. Jabbar, P. Chandra, and B. Deekshatulu [9, Sec. 4] used hybrid feature subset selection measures such as Symmetric Uncertainty (SU), Information Gain (IG), and Genetic search to lessen immaterial and repetitive properties and attributes.

Despite many research works that ensued in predicting heart disease, Bangladesh is probably the least studied among all South Asian nations [3, Sec. 1]. But some researchers did some beneficial research works. Many of these research activities are worth mentioning. M. Sultana, A. Haider, and M. S. Uddin [10, Sec. 3] worked with two data sets. One is the standard data set of 13 attributes with 270 records downloaded from the UCI Machine Learning Repository. Other is the dataset of 6 attributes with mere 100 records collected from Enam Medical Diagnosis Centre, Savar, Dhaka, Bangladesh. After collecting the data sets, the researchers examined their heart disease prediction procedure through Weka software using KStar, J48, SMO, Bayes Net, and Multilayer Perceptron algorithms. After analyzing the results, they found that SMO and Bayes Net showed optimum performances than the performances of KStar, Multilayer Perceptron, and J48 algorithms based on performance factors in their proposed system. S. M. M. Hasan, M. A. Mamun, M. P. Uddin, and M. A. Hossain [11, Sec. 4] worked with the Cleveland Heart Disease dataset of 14 attributes and 303 records downloaded from the UCI machine learning repository. They utilized information gain as a feature selection technique to omit unnecessary features and different classification algorithms like KNN, Decision Tree (ID3), Gaussian Naive Bayes, Logistic Regression, and Random Forest to predict coronary illness. After analyzing the results, they found that Logistic Regression yielded maximum accuracy with 92.76% than all the other algorithms in the research outcomes. K. C. Howlader, M. S. Satu, and A. Mazumder [12, Sec. 3] worked with the data set containing 34 attributes and mere 116 records collected from three hospitals in Dhaka, Bangladesh. They used seven classification algorithms to predict coronary illness. Those are Naive Bayes, Bayes Net, SMO, IBk, KStar, Heoffding Tree, and J48. Naive Bayes yielded the best accuracy (mere 70.83%) in their proposed model [12, Sec. 6]. But the authors mentioned the reason for not getting so many satisfactory outcomes is missing attributes in the data set that are replaced later with relevant values [12, Sec. 7].

We can conclude that Researchers' efforts related to predicting heart disease in Bangladesh are appreciable, but a few pieces of research occurred concerning this developing country. Moreover, most of the research activities were executed either by downloading data set from an online repository or by a few data collecting from Bangladeshi hospitals or medical institutions. This paper describes our efforts concerning this matter.

III. QUESTIONNAIRE AND DATA COLLECTION

The Head of the Department of Cardiology of Sylhet MAG Osmani Medical College, Dr. Muhammad Shahabuddin, helped us to finalize the questionnaire. He suggested eighteen attributes and pointed out the eight attributes of them as the main attributes given in Fig. I. They are the following: 1. Age, 2. Gender, 3. Smoking Habit, 4. Diabetes, 5. (Cholesterol) / Dyslipidemia H/O, 6. Chest Pain, 7. Hypertension (140/90), 8. Family History (any family member had heart disease or not).

TABLE I. QUESTIONNAIRE

Attributes	**Options**
1. Age	❑ **> 40** ❑ **< 40**
2. Gender	❑ **Male** ❑ **Female** ❑ **Other**
3. Smoking Habit	❑ **Yes** ❑ **No**
Attributes	**Options**
If the answer to question 3 is yes, then answer the question no. 4	
4. Smoking Condition	❑ Present Smoker ❑ Past Smoker
5. Regular Pulse	❑ Yes ❑ No
6. Physical Activity	❑ Gym

	❑ Walk ❑ Outdoor game ❑ No Activity
7. Diabetes	❑ **Yes** ❑ **No**
8. (Cholesterol) / Dyslipidemia H/O	❑ **Yes** ❑ **No** ❑ **Unknown**
9. Chest Pain	❑ **Typical angina** ❑ **Atypical angina** ❑ **Non-Cardiac chest pain** ❑ **No chest pain**
10. Hypertension (140/90)	❑ **Yes** ❑ **No**
11. Skip heart failure medicines when feeling better	❑ Yes ❑ No
12. Junk food eating habit	❑ Everyday ❑ 2-3 days a week ❑ 4-6 days a week ❑ No
13. Rice eating habit	❑ One time a day ❑ Two times a day ❑ Three times a day
14. Family History (any family member had heart disease)	❑ **Yes** ❑ **No**
If the answer to question 14 is yes, then answer questions no. 15-18	
15. Relation with the person	❑ Father ❑ Mother ❑ Sister ❑ Brother
16. Age of that family member	❑ Age < 65 ❑ Age > 65
17. Gender of that family member	❑ Male ❑ Female ❑ Other
18. Disease Type of that family member	❑ Heart Attack ❑ Heart Block

After finalizing the questionnaire, we collected data of 564 people based on the questionnaire by filling up a hard copy of the questionnaire. Among these 564 people, heart patients were 313 in total, and healthy people were 251 in total. We collected data of heart patients from six different medical institutions in Sylhet. They are the following: 1. Jalalabad Ragib-Rabeya Medical College and Hospital, Sylhet, 2. Mount Adora Hospital, Sylhet, 3. North East Medical College and Hospital, Sylhet, 4. Sylhet Women's Medical College and Hospital, 5. National Heart Foundation Hospital, Sylhet, 6. Sylhet MAG Osmani Medical College Hospital. But, we collected healthy people's data from educational institutions, relatives, and other people of the Sylhet region.

After collecting data manually, we input the manual data by filling up the survey form on the website built by us to store them into MySQL format [5]. We built the website using 1. HTML 5, 2. CSS 3, 3. Bootstrap 4, 4. JavaScript, 5. jQuery, 6. PHP 5.6.12, and 7. CodeIgniter 3.1.9. We build the site to store manual data into MySQL format and to use MySQL data to convert into CSV data format so that we can use the CSV data file in our python programming. After storing all the manual data into our website database, we checked the noisy data and eliminated them from our dataset. For example, if we found data related to Non-Cardiac Chest Pain (NCCP), which means chest pain that is not involved with heart disease, then we eliminated these types of data. Then we cleared the unnecessary portion of the dataset that did not need prediction, such as the name and phone number of the participants. We found 13 noisy data. After removing them from a total of 564 data, we got 551 data in the final dataset.

IV. Machine Learning Techniques Implementation

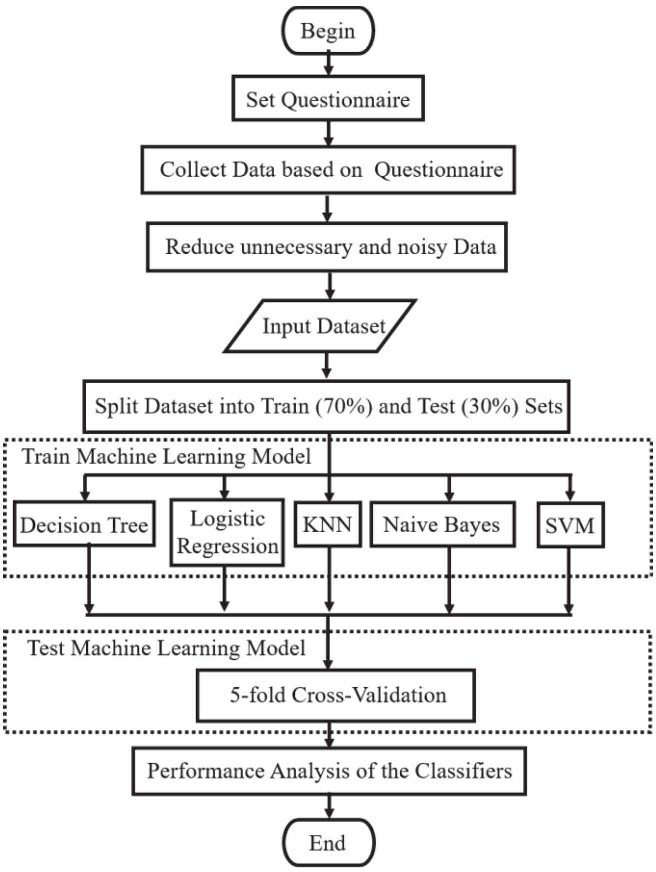

Fig. 1. Flowchart of the proposed research work activities.

After finalizing the dataset, we split the dataset into two separate sets: training and testing. 70% of data remained for training the machine learning model. The other 30% of data remained for testing the final model after the training of the model is done. Then we used these split datasets for machine learning (ML) implementation using the python programming language. Our outcomes of the experiment are not continuous numerical values; instead, they are values like true or false; because of that, we used classification algorithms of supervised machine learning techniques instead of regression. We applied five classification ML algorithms to train the proposed model. The algorithms are the following: 1. Decision Tree, 2. Logistic Regression, 3. K-Nearest-Neighbor (KNN), 4. Naive Bayes, 5. Support Vector Machine (SVM). We used Anaconda Jupyter Notebook as the Integrated Development Environment (IDE). The python libraries we used are 1. Pandas, 2. NumPy, 3. Scikit-learn, 4. Matplotlib, and 5. Seaborn. We used the Pandas library for reading our data file, NumPy for mathematical purposes, Scikit-learn library for splitting the dataset into the test and train sets, and Matplotlib and Seaborn for the visual representation of our model's results.

The five machine learning algorithms implementation for predicting heart disease are summarized below:

A. Decision Tree:

We executed the decision tree classification algorithm since it is truly outstanding and most utilized regulated learning strategies. Building a decision tree that is steady into a given data set is simple. The difficulty lies in building proper decision trees, which implies concise decision trees. We used ID3 ((Iterative Dichotomiser 3), which is the most widely used type among all the decision tree algorithms' types. To execute the algorithm, we required a few boundaries, and they are:

Criterion: We utilized the Gini Index first to see the results naturally. Besides, we used Entropy for Information Gain.

Splitter: We utilized the default Splitter, and that is "best."

Maximum Depth: Our Maximum Depth is six since this has the most extreme precision.

The decision tree yielded by our research is given below:

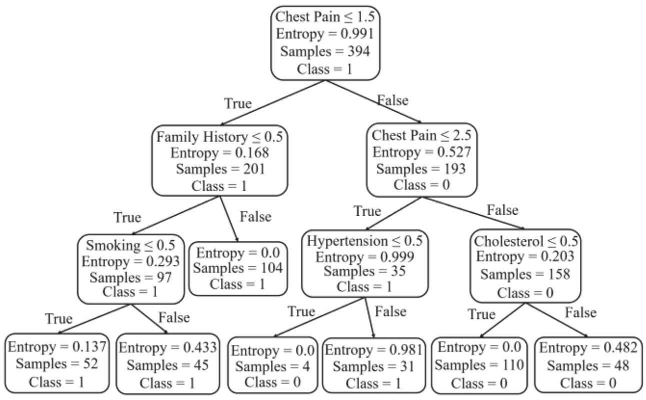

Fig. 2. The Decision Tree of the proposed model.

B. Logistic Regression:

Logistic Regression stands on the concept of probability that is a predictive analysis algorithm. We can consider a Logistic Regression as a Linear Regression model, but the Logistic Regression utilizes a more intricate cost function. This cost function is known as the 'Sigmoid function,' likewise as the 'Logistic Function.' The cost function's range is between 0 and 1 in the hypothesis of logistic regression. Therefore linear functions cannot enact it as linear functions can have a value less than 0 or greater than 1. The Logistic Regression Hypothesis Expectation is:

$$0 \leq h_{\theta}(x) \leq 1 \qquad (1)$$

In equation (1), $h_{\theta}(x)$ represents the hypothesis expectation of any instance or data point x. We used the sigmoid function to map the calculated values of probabilities. The sigmoid function maps any real value into another value between 0 and 1. The predictive model that we used for research is supervised and also looking for yes or no answer. That is the reason for using this algorithm for predicting heart disease.

C. K-Nearest Neighbor (KNN):

K-Nearest Neighbors (KNN) is probably one of the simplest algorithms utilized in machine learning for both regression and classification techniques. KNN calculations use the data and characterize new data points dependent on resemblance measures (e.g., distance function). In basic, we can say The KNN calculation accepts that comparative things are close to one another. In KNN, Classification occurs by considering the majority vote to its neighbors. The data point goes to the class that has the most intimate neighbors. As we increment the number of nearest neighbors, the estimation of k and accuracy may increment. The predictor variables anticipate the target variable or the dependent variable. For our situation, our target variable is the chance of coronary illness, and by utilizing the KNN algorithm, our model will predict the result.

D. Naïve Bayes:

Naive Bayes classifier is a machine learning approach that we also utilize for classifying and predicting the probability of an instance. Every Naive Bayes classifier assumes that the value of a particular feature is distinct from any other features' values given in the class variables. For example, a machine can consider any particular fruit as an apple if it is red, circular, and its diameter is approximately 10 cm. Naive Bayes classifier looks on every one of these highlights to look on distinctly to the likelihood that the fruit is an apple, paying little heed to any interrelation between the shading, roundness, and measurement features. The classifier has emerged from the Bayes theorem, which is the following:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (2)$$

Here above in eq. (2), utilizing the Bayes' theorem, we can discover the happening probability of an instance A, given that B has happened already. Here, we can consider B as the proof and A as the hypothesis. The supposition made here is that the predictors/features are independent. That is the presence of one specific characteristic that doesn't influence the other. Thus the Bayes theorem is called Naive.

E. Support Vector Machine (SVM):

Support Vector Machine (SVM) is a regulated ML calculation that we utilize for both regression and classification tasks. Notwithstanding, we generally use it in classification issues. In our model to implement SVM, we plot every single data item as a point in n-dimensional space (where n is the number of features you have) with the estimation of each component being the estimation of a particular coordinate. At that point, we perform classification by finding the hyper-plane that separates the two classes well overall.

V. RESULTS

Before finalizing the results of all classifiers, we needed to identify the number of correctly and incorrectly classified instances. Confusion matrix - a performance indicator gives information about the number of correctly and incorrectly classified instances for each classifier. We used a total of 551 collected data to find the confusion matrix for each algorithm. Among them, 300 are data of heart patients, and 251 are data of healthy people. Each classifier generated a 2x2 confusion matrix for two probable results (one is Negative, and another is Positive).

- True Positive (TP): TP is the number of total positive predictive instances that are categorized without error, and they are factually positive.
- True Negative (TN): TN is the number of total negative predictive instances that are categorized without error, and they are factually negative.
- False Positive (FP): FP is the number of total positive predictive instances that are categorized with error, and they are not factually positive.
- False Negative (FP): FN is the number of total negative predictive instances that are categorized with error, and they are not factually negative.
- Precision: precision discloses the proportion of the data points a model says is relevant and was relevant actually. That means classification models return only relevant instances in precision.
- Recall: recall discloses the ability to find all relevant instances in a dataset. That means classification models identify all relevant instances in a recall.

We took into account the following performance parameters to explore the confusion matrix:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Accuracy\ (ACC) = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

The results of the confusion matrix for all classifiers (for all 18 and main 8 attributes) are given in the following table:

TABLE II. CONFUSION MATRIX OF CLASSIFIERS

Classifications	**Confusion Matrix for all 18 attributes**		**Confusion Matrix for 8 main attributes**	
1. Decision Tree	TN (65) 38.46%	FP (12) 7.10%	TN (63) 37.28%	FP (14) 8.28%
	FN (11) 6.51%	TP (81) 47.93%	FN (9) 5.33%	TP (83) 49.11%
2. Logistic Regression	TN (58) 34.32%	FP (17) 10.06%	TN (58) 34.32%	FP (18) 10.65%
	FN (9) 5.33%	TP (85) 50.30%	FN (5) 2.96%	TP (88) 52.07%
3. K-Nearest-Neighbor (KNN)	TN (65) 38.46%	FP (19) 11.24%	TN (67) 39.64%	FP (10) 5.92%
	FN (8) 4.73%	TP (77) 45.56%	FN (13) 7.69%	TP (79) 46.75%
4. Naive Bayes	TN (57) 33.73%	FP (18) 10.65%	TN (67) 39.64%	FP (10) 5.92%
	FN (11) 6.51%	TP (83) 49.11%	FN (13) 7.69%	TP (79) 46.75%
5. Support Vector Machine (SVM)	TN (69) 40.83%	FP (13) 7.69%	TN (72) 42.60%	FP (8) 4.73%
	FN (8) 4.73%	TP (79) 46.75%	FN (7) 4.14%	TP (82) 48.52%

The accuracy of our proposed model is shown below by a figure:

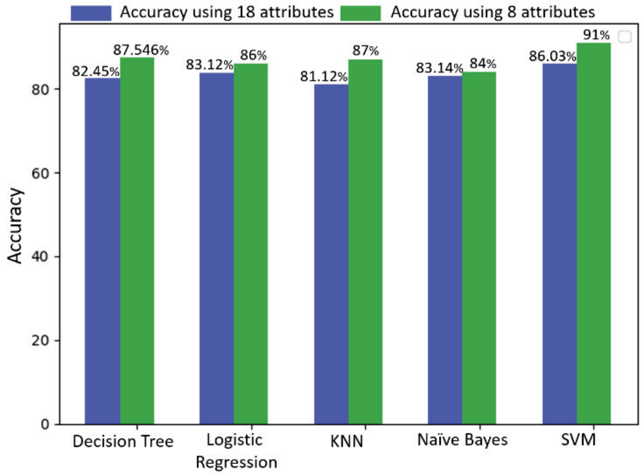

Fig. 3. Accuracy of all classification algorithms in the proposed model.

VI. CONCLUSION

In this era of technology and mass population, quick and accurate prediction of any disease is very necessary. The preprocessing of raw healthcare data into a well-organized machine learning model can resolve the issue. An appropriate machine learning model can not only predict a disease quickly but also can predict it with a satisfactory level of accuracy, improve treatment, reduce human intervention, and medical lab tests. As mortality for heart disease is increasing beyond imagination, building proper machine learning models for heart disease prediction is one of the obligatory requirements for the present time. This paper describes our effort to do so by dint of the real-word data collected from hospitals and healthcare industries of Sylhet, Bangladesh. Bangladesh likely has the most elevated paces of heart disease but is the least studied among all South Asian nations. Considering the issue, we want to collect heart disease-related data from healthcare industries of all-around Bangladesh as far as possible to investigate and to strengthen our proposed model. We also have a plan to include artificial neural networks (ANN), as ANN is well known to yield the highest accuracy prediction in the medical sector. We can also use regression analysis of supervised machine learning to determine the vulnerability level of a person towards heart disease in percentage instead of just only classifying either someone is a heart patient or not using classification.

ACKNOWLEDGMENT

We like to show our humble gratitude and thanks to the Head of the Department of Cardiology of Sylhet MAG Osmani Medical College, Dr. Muhammad Shahabuddin, and the Assistant Professor of the Department of Cardiology of Sylhet MAG Osmani Medical College, Dr. Md. Anamur Rahman Anam for giving their valuable time to guide us in preparing the questionnaire. We also like to show our humble gratitude and thanks to our respective Universities for their support.

REFERENCES

[1] "The top 10 causes of death," *World Health Organization*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. (accessed Oct. 24, 2020).

[2] "World Health Organization - Noncommunicable Diseases (NCD) Country Profiles (Bangladesh), 2018," *World Health Organization*. [Online]. Available: https://www.who.int/nmh/countries/bgd_en.pdf (accessed Oct. 24, 2020).

[3] A.K.M. Monwarul Islam, and A. A. S. Majumder, "Coronary artery disease in Bangladesh: A review," *Indian heart journal*, vol. 65, no. 4, pp. 424-435, Jul.-Aug. 2013, doi: 10.1016/j.ihj.2013.06.004.

[4] "Cardiologist Sylhet Heart Specialist," *Specialist Doctors Bangladesh*, 21-Sep-2020. [Online].

Available: https://specialistdoctorsbd.com/cardiologist-sylhet-heart-specialist/. (accessed Oct. 26, 2020).

[5] "Welcome To Heart Disease Prediction," *HDP*. [Online]. Available: https://syl-hdp.globentrust.com/. (accessed Oct. 27, 2020).

[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] 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.

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

[9] M. A. Jabbar, P. Chandra, and B. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 628–634, Nov. 27-29, 2012, doi: 10.1109/ISDA.2012.6416610.

[10] M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction," *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, Sept. 22-24, 2016, doi: 10.1109/CEEICT.2016.7873142.

[11] S. M. M. Hasan, M. A. Mamun, M. P. Uddin, and M. A. Hossain, "Comparative Analysis of Classification Approaches for Heart Disease Prediction," *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, Feb. 8-9, 2018, doi: 10.1109/IC4ME2.2018.8465594.

[12] K. C. Howlader, M. S. Satu, and A. Mazumder, "Performance Analysis of Different Classification Algorithms that Predict Heart Disease Severity in Bangladesh," *2017 International Journal of Computer Science and Information Security (IJCSIS)*, vol. 15, no. 5, pp. 332-340, May, 2017, doi: 10.1109/CEEICT.2016.7873142.