

Detection of Significant Risk Factors of Heart Diseases by Educating Multimodal Features and Implementing Machine Learning Techniques

Kanyifechukwu Jane Oguine¹, Ozioma Collins Oguine², Chito Frances Ofodum³

¹Department of Computer Science, University of Abuja, Abuja, Nigeria

²Department of Computer Science, University of Abuja, Abuja, Nigeria

³College of Medicine, University of Lagos, Lagos, Nigeria

ABSTRACT

Article Info

Volume 7, Issue 4

Page Number: 621-630

Publication Issue :

July-August-2021

Article History

Accepted : 07 Aug 2021

Published : 14 Aug 2021

One of the major reasons for deaths worldwide is heart diseases and possible detection at an earlier stage will prevent these attacks. Medical practitioners generate data with a wealth of concealed information present, and it's not used effectively for predictions. For this reason, the research will convert the unused data into a dataset for shaping using different data mining techniques. People die having encountered symptoms that were not taken into considerations. The main objective of this paper is to analyze the most significant risk factors of Heart Diseases of patients by extracting multimodal features and predicting the occurrence of heart diseases using different classification techniques comparatively. This study will help improve the decision-making of medical professionals on the occurrence of heart diseases patients in a bid to enhance early detection by implementing comparatively several machine learning techniques resulting in an improved prediction accuracy using patient records (Multimodal Features).

Keywords : Heart Diseases Detection, Machine Learning, Multimodal Features, Logistic Regression, Naïve Bayes, Random Forest.

I. INTRODUCTION

The main reason for death worldwide is heart attack diseases and possible detection at an earlier stage will prevent these attacks. Medical practitioners generate data with a wealth of concealed information present, and it's not used effectively for predictions. For this reason, the research will convert the unused data into a dataset for shaping using different data mining techniques. People die having encountered symptoms

that were not taken into considerations. There is a requirement for medical practitioners to define heart diseases before they occur in their patients. The features that increase the chances of heart attacks are smoking, lack of physical exercise, high blood pressure, high cholesterol, unhealthy diet, detrimental use of alcohol, and high sugar levels. Cardio Vascular Diseases (CVD) constitute coronary heart, cerebra-vascular or Stroke, hypertensive heart diseases, congenital heart, peripheral artery,

rheumatic heart diseases, and inflammatory heart diseases. Data mining is a knowledge discovery technique to examine data and encapsulate it into useful information [1].

The current research intends to forecast the probability of being diagnosed with heart diseases given a dataset of patients in a Clinic by educating multimodal causative features. Prediction and descriptions are principal goals of data mining; in practice “prediction” in data mining involves attributes or variables in the data set to locate unknown or future state values of other attributes. Description emphasizes discovering patterns that describe the data to be interpreted by humans. Early-stage detection of these diseases and predicting the probability of a person being at risk of getting them can reduce the death rate. Medical data mining techniques are used in medical data to extract meaningful patterns and knowledge. Medical information has redundancy, multi-attribution, incompleteness, and a close relationship with time. The problem of using the massive volumes of data effectively becomes a major problem for the health sector. Data mining provides the methodology and technology to convert these data mounds into useful decision-making information. This prediction system for heart diseases would facilitate Cardiologists in taking quicker decisions so that more patients can receive treatments within a shorter period, resulting in saving millions of lives.

As indicated by the most recent WHO information distributed in 2017 Coronary Heart Diseases Deaths in Bangladesh achieved 112,791 or 14.31% of total death [2]. According to the American Heart Association (AHA), in the United States (U.S.) someone has IHA about every 40 seconds that is frightening [3]. Nowadays, heart disease is one of the biggest problems in the healthcare sector because 100,600 deaths were caused by ischemic heart diseases, 100,780 by strokes, and by hypertensive heart diseases

about 28,000, in 2013. The main reason for this condition is the food habits of people [4].

They love to eat too oily, rich foods and they don't have the habit of doing exercise though many of them are smokers [4].

II. Empirical Review Literature

Cardiovascular diseases are the primary cause of death worldwide over the past decade. According to the World Health Organization (WHO), it is estimated that over 17.9 million deaths occur each year because of cardiovascular diseases, and out of these deaths, 80% is attributed to coronary artery diseases and cerebral stroke. Many habitual factors such as personal and professional habits and genetic predisposition account for heart diseases. Various risk factors such as smoking, overuse of alcohol and caffeine, stress, and physical inactivity along with other physiological factors like obesity, hypertension, high blood cholesterol, and pre-existing heart conditions are often deciding factors for heart diseases. The efficient, accurate and early medical diagnosis of heart diseases plays a pivotal role in taking preventive measures to avoid the complications that arise due to such diseases [5].

The major challenge faced in the world of medical sciences today is the provision of quality service and efficient and accurate prediction. The latter problem can be solved by automation with the help of Data Mining and Machine Learning. Data mining is defined as a process used to extract usable data from a large set of raw data. It implies analyzing patterns in large batches of data by making use of various software. It also involves effective data collection and warehousing coupled with computer processing.

Machine Learning (ML) which is a subfield of data mining that deals with large scale well-formatted dataset efficiently. In the medical field, machine

learning can be used for the diagnosis, detection, and prediction of various diseases. Various Machine Learning algorithms such as Logistic Regression, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest, and the ensemble technique of XGBoost are compared to find the most accurate model [6].

[5] used Naive Bayes and decision tree data mining techniques for predicting different types of diseases. They mainly concentrated on the prediction of heart diseases, diabetics, and breast cancer. The results were derived from the confusion metrics.

[7] proposed a Naive Bayes classifier approach for the prediction of cardiovascular diseases. The authors have considered few important risk factors for deciding the heart diseases.

[8] studied various ML algorithms that can be used for the classification of heart diseases. The research was carried out to study Decision Tree, KNN, and K-Means algorithms that can be used for classification, and their accuracy was compared. This research concludes that accuracy obtained by Decision Tree was highest further it was inferred that it can be made efficient by the combination of different techniques and parameter tuning.

[9] has designed an ML model comparing five different algorithms. Rapid Miner tool was used which resulted in higher accuracy compared to MATLAB and Weka tool. In this research, the accuracy of Decision Tree, Logistic Regression, Random Forest, Naïve Bayes, and SVM classification algorithms were compared. The decision tree algorithm had the highest accuracy.

[10] executed a survey including different classification algorithms used for predicting heart diseases. The classification techniques used were Naive Bayes, KNN (K-Nearest Neighbor), Decision tree, Neural network, and accuracy of the classifiers was analyzed for the different number of attributes.

[11] created a highly accurate hybrid method for the diagnosis of coronary artery diseases. The proposed method can increase the performance of the neural

network by approximately 10% by enhancing its initial weights using a genetic algorithm.

[6] proposed the development of a framework based on associative classification techniques on heart datasets for the diagnosis of heart-based diseases. The implementation of work is done on the Cleveland heart diseases dataset from the UCI machine learning repository to test different data mining techniques. The various attributes related to the cause of heart diseases are gender, age, chest pain type, blood pressure, blood sugar, etc. that can predict early symptoms of heart diseases.

III. METHODS AND MATERIAL

This research examines the empirical relationship between the set of features such as age, sex, and blood pressure with the probability of being diagnosed with heart disease. Therefore, this study is a quantitative case study. The main objective of this paper is to analyze the most significant risk factors of Heart Diseases of patients by extracting multimodal features and predicting the occurrence of heart diseases using different classification techniques comparatively.

3.1 Data Collection and Description

The data is collected from the UCI machine learning repository. The dataset is named "Heart Diseases Dataset" and can be found in the UCI machine learning repository. The UCI machine learning repository contains a vast and a varied number of datasets which include datasets from various domains. These data are widely used by the machine learning community ranging from novices to experts to enhance empirical analysis. Various academic papers and researches have been conducted using this repository. There are 76 attributes in the dataset but only 14 attributes are used for this study as shown in table 3.1 below.

Table 3.1 : Dataset Attributes

Features	Description
Age	Age in years
Sex	Gender instance (0 = Female, 1 = Male)
Cp	Chest pain type (1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic)
Trestbps	Resting blood pressure in mmHg
Chol	Serum Cholesterol in mg/dl
Fbs	Fasting blood sugar > 120 mg/dl (0 = False, 1 = True)
Restecg	Resting ECG results (0: normal, 1: ST-T wave abnormality, 2: LV hypertrophy)
Thalach	Maximum heart rate achieved
Exang	Exercise-induced angina (0: No, 1: Yes)
Oldpeak	ST depression induced by exercise relative to rest
Slope	The slope of the peak exercise ST segment (1: up-sloping, 2: flat, 3: down-sloping)
Ca	Number of Major vessels colored by fluoroscopy (values 0 - 3)
Thal	Defect types: value 3: normal, 6: fixed defect, 7: irreversible defect
Num	Diagnosis of heart diseases (0: Healthy, 1: Unhealthy)

3.2 Data Preprocessing and Splitting

The data set has 6 missing values. Since the missing values are from categorical attributes it was replaced with the mode value of the respective columns. There are 43 outliers in the dataset but it is not a noise or wrong input. Data falling outside the 3 standard deviation are considered outliers. Patients who are diagnosed with heart diseases have these outliers, meaning their cholesterol and blood pressure data is above the normal scale. So, in a real clinical scenario, it is not an outlier, and therefore it is decided to keep these outliers as they are. Standardization was used to rescale the data. That helped the machine learning models to perform better.

For the dataset, Sequential Backward Elimination was used. Any features in the dataset whose p-value is less than 0.05 are significant and are considered essential in this dataset as shown in table 3.2.

For each experiment, the entire dataset was split into 70% training set and 30% test set. We used the training set for resampling, hyperparameter tuning, and training the model and I used the test set to test the performance of the trained model. While splitting the data, a random seed (any random number) was specified, which ensured the same data was split every time the analysis was run.

Table 3.2 : Rescaled dataset description

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	167	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

3.3 Feature Selection and Extraction

In this section we are going to distribute the target value from the feature variable, it is vital for choosing appropriate accuracy metrics and consequently properly assess different machine learning models. First of all, we will count the values of explained variables otherwise known as the determining variable which is going to give us the prediction of a patient being affected by heart diseases or not. Secondly, we will separate numeric features from categorical features. And lastly, present the relation (table 3.3) between the categorical features in various plots and try to figure out or rather observe the influence of those categorical features in the actual determining variable “diagnosis”.

Feature generating is the main stage of data processing, where data features are extracted and stored for the next stages. The output is a list of

extracted features with values, to make the prediction model balanced.

Table 3.3: Resulting dataset after feature selection

	age	sex	resttbps	chol	fbs	restecg	thalach	exang	oldpeak	ca	...	cp_1	cp_2	cp_3	thal_0	thal_1	thal_2	thal_3	slope_0	slope_1	slope_2
0	63	1	145	233	1	0	150	0	2.3	0	...	0	0	1	0	1	0	0	1	0	0
1	37	1	130	250	0	1	187	0	3.5	0	...	0	1	0	0	0	1	0	1	0	0
2	41	0	130	204	0	0	172	0	1.4	0	...	1	0	0	0	0	1	0	0	0	1
3	56	1	120	236	0	1	178	0	0.8	0	...	1	0	0	0	0	1	0	0	0	1
4	57	0	120	354	0	1	163	1	0.6	0	...	0	0	0	0	0	1	0	0	0	1

5 rows x 22 columns

3.4. Diagrammatic Modeling of Prediction Processes

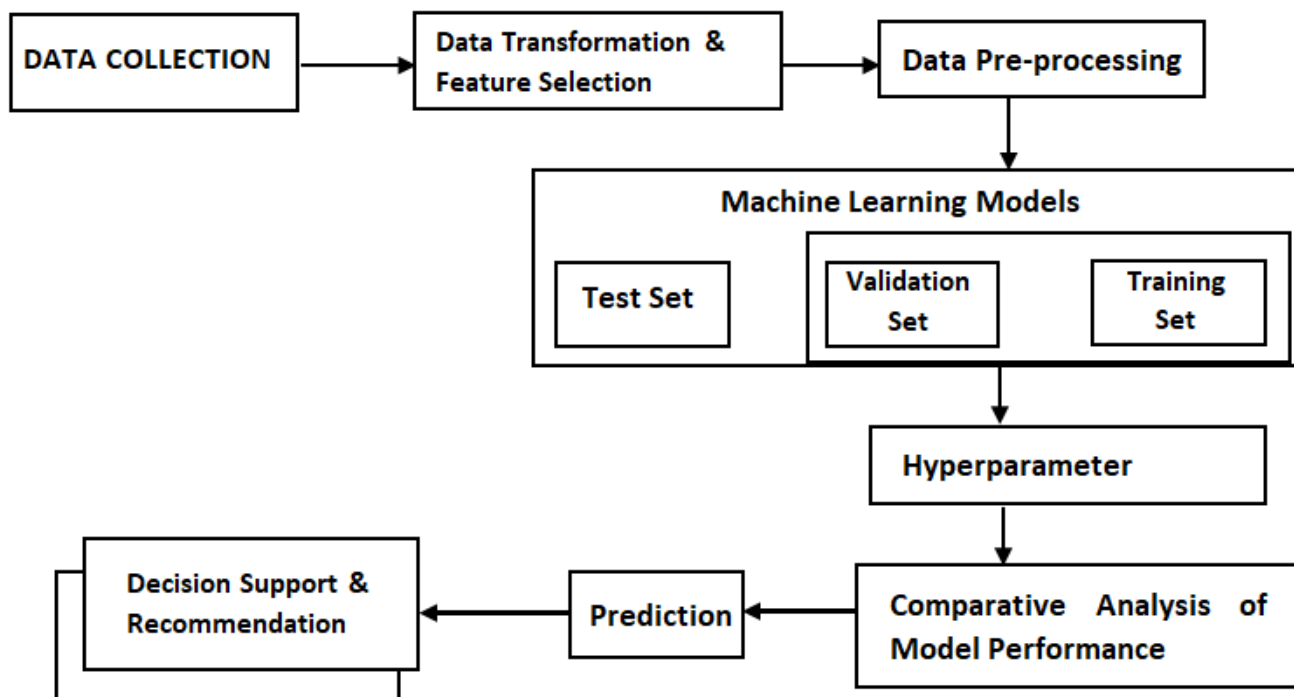


Fig 3.1 : Flow Diagram of the proposed Model for the comparative analysis of the Predictive Algorithms

In the fig. 3.1 above, the heart diseases prediction process is modeled in detail, firstly the dataset is obtained from a reliable source, after which the dataset is cleaned and missing values are dropped. For a machine-learning algorithm to learn from the dataset and to avoid overfitting, feature engineering is done on the dataset to know which features have a great effect on the target variable.

The dataset is then split into test, train, and validation test, the training set is tuned using hyperparameters, which are used to control the learning process of the model. Then the tuned dataset is fed to the model to learn from after which the performance of the model is tested using the Test set.

3.5 Data Mining Techniques/Algorithms

The data collected in this research was evaluated using three major statistical algorithms (Logistic

Regression, Naïve Bayes, and Random Forest), conclusions and recommendations were also drawn based on the comparative performance of the resulting models.

3.6 Choice of Programming Language

Python programming language and Jupyter Notebook were used for data mining predictive tasks. Python is a well-known general-purpose and dynamic programming language that is being used for different fields such as data mining [12], machine learning [13], [14], and the internet of things [15], [16]. Data mining algorithms are being implemented using python with the help of special-purpose libraries. The models were developed using 5-fold cross-validation.

IV. Experimental Result Presentation and Discussion

Three data mining algorithms were applied directly to the dataset using the python programming language. However, the model developed with Random Forest (RF) algorithm was found to be the most efficient with an f1 score of 84.46 % in contrast to the Naïve Bayes algorithm and the Logistic Regression with a relatively equal f1 score of 81.24% as shown in table 4.4 and fig 4.1.

4.1 Performance Evaluation of Models

For classification problems, the metrics used to evaluate an algorithm are accuracy, confusion matrix, precision, recall, and F1 values **Error! Reference source not found..** Since the project is a classification task, the comparative accuracy and the classification report of the two models were computed. Below is a little explanation of the evaluation techniques:

Accuracy - Accuracy is the percentage of correctly classified instances. It is one of the most widely used classification performance metrics.

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total Number of predictions}} \quad (1)$$

In the case of binary classification models. The accuracy can be defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Precision - Precision is the number of classified Positive or fraudulent instances that are positive instances.

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

Recall - Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made. Unlike precision that only comments on the correct positive predictions out of all positive predictions, recall indicates missed positive predictions. The recall is calculated as the number of true positives divided by the total number of true positives and false negatives.

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

F1 Score - The F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

$$F1\ Score = \frac{2 * (Recall*Precision)}{(Recall + Precision)} \quad (5)$$

Where:

TP = True Positive;

TN = True Negative;

FP = False Positive;

FN = False Negative

4.2 Comparative Analysis of models

Data mining models are evaluated using evaluation techniques to determine their accuracy. The techniques determine the quality and efficiency of the model using the data mining algorithm or machine

learning algorithms. These main performance evaluation techniques for the data mining model include specificity, sensitivity, and accuracy. However, in this study, the only accuracy is considered to evaluate the developed models.

Analysis with Logistic Regression

The Accuracy of the Logistic Regression model after applying principal component analysis was 86.89%. Along with Accuracy, other performance metrics; Precision, Recall, and F1 score raise after the introduction of PCA which can be seen from Table 4.1.

Table 4.1 : Performance metrics for the Logistic Regression Model

	precision	recall	f1-score	accuracy
Logistic Regression	0.867102	0.867102	0.867102	0.868852

Analysis with Naïve Bayes

The Accuracy of the Naïve Bayes model after applying principal component analysis was 86.89%. Along with Accuracy, other performance metrics; Precision, Recall, and F1 score raise after the introduction of PCA which can be seen from Table 4.2.

Table 4.2 : Performance metrics for the Multinomial Naïve Bayes Model

	precision	recall	f1-score	accuracy
Naïve Bayes	0.867102	0.867102	0.867102	0.868852

Analysis with Random Forest

The Accuracy of the Random Forest model after applying principal component analysis was 88.53%. Along with Accuracy, other performance metrics; Precision, Recall, and F1 score raise after the

introduction of PCA which can be seen from Table 4.3.

Table 4.3: Performance metrics for the Random Forest Model

	precision	recall	f1-score	accuracy
Random Forest	0.885165	0.881808	0.883238	0.885246

In table 4.4 below, we compared the three algorithms used in the analysis of the dataset. The Random Forest model performs better in all metrics while the other two (the Naïve Bayes and Logistic Regression Model) have relatively equal valuations. Below are visual representations of the comparative result in fig. 4.1.

Table 4.4 : Comparative Model Performance

Classifiers	Accuracy	Precision	Recall	F1 Score
Random Forest	0.867102	0.867102	0.867102	0.868852
Naïve Bayes	0.867102	0.867102	0.867102	0.868852
Logistic Regression	0.835165	0.836829	0.833897	0.834445

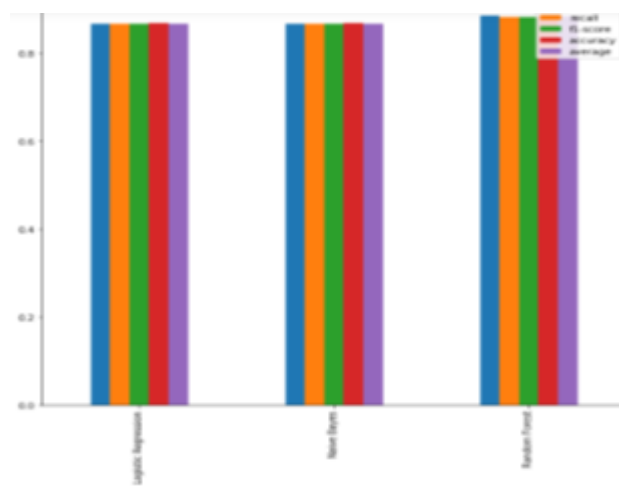


Fig. 4.1: Visual Representation of the Comparative performance of the 3 models

4.3 Prediction

The model predicted the most significant risk factors of heart diseases of patients by extracting multimodal features. Based on our analysis, two main significant risk factors served as modalities for making predictions. The modalities are age and gender of the

patients as shown in the bar visualization (fig 4.2 and fig. 4.3) below;

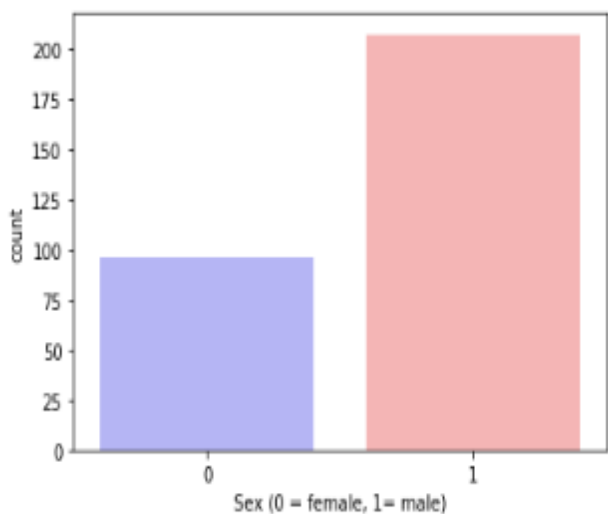


Fig 4.2: Heart Diseases Count by Gender

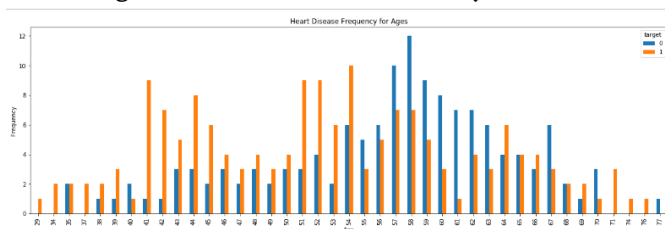


Fig. 4.3: Heart Diseases Frequency by Age

In the fig. 4.2 above, we can see that male patients are more susceptible to heart diseases than their female counterparts, however further analysis indicates that some underlying factors either due to habits or lifestyle might be the major reason for the deduction. It was also discovered from analysis as shown in fig. 4.3, that there is a high frequency of heart diseases amongst patients from the age range 40 – 60. This also means that critical measures and medical diagnosis are to either be taken to advise Male patients in their 40s or 50s of their susceptibility to heart diseases.

V. Summary

Heart diseases are one of the leading causes of death worldwide and the early prediction of heart diseases is important. The computer-aided heart diseases prediction model when implemented as a tool helps physicians and medical practitioners in heart diseases

diagnosis and early decision making. Some Heart Diseases classification systems were reviewed in this project. From the analysis, it is concluded that data mining plays a major role in heart diseases classification. Traditional machine learning with offline training is good for diseases prediction in the early stage and the good performance of the system can be obtained by preprocessed and normalized datasets. The classification accuracy can be improved by a reduction in features.

VI. Conclusion

This research compares different machine learning algorithms and predicts if a certain person, given various personal characteristics and symptoms, will get heart disease or not. The main motive of our report was to comparing the accuracy and analyzing the reasons behind the variation of different algorithms. We have used the UCI dataset for heart diseases which contains 303 instances and used 10-fold Cross-Validation to divide the data into two sections which are training and testing datasets. We have considered 14 attributes and implemented three different algorithms to analyze their comparative accuracy in prediction. By the end of the implementation part, we have found that the models used performed excellently, with the Random Forest Model [roving the most effective in predicting the likelihood of occurrence of heart diseases. In addition to that, we also observed that Increased collaboration in the development of the AI prediction models can enhance their applicability in the clinical practice and assist healthcare providers and developers in the fight against heart diseases.

VII. Recommendation

The dataset that is used in our project is very small and old. Moreover, no new dataset regarding heart diseases has been introduced so far. There is a need for a new dataset and we can collect that from various

hospitals. We can also evaluate the efficiency of each classifier and also such classifiers in combination, by employing the bagging, boosting, and stacking techniques.

Creating a database for echocardiography examination results and other examination results would be helpful for searching, retrieving, and minimizing memory space.

Currently, patient history and knowledge used for interpreting the echocardiography results are not stated in the dataset. The researcher believes that stating this hidden knowledge can have a positive impact on researches that will be conducted in the future. The echocardiography offers two-dimensional images during the examination. Unfortunately, the images generated from each examination are not stored by the hospital instead; they are discarded as soon as the examination is over. The hospital should find a way to store the image so that it can be used to extract relevant information related to the diseases using intelligent image recognition systems.

VIII. Abbreviations

WHO, World Health Organization; US, United States; RF, Random Forest; UCI, University of California, Irvine; AHA, American Heart Association; CVD, Cardio Vascular Diseases; ML, Machine Learning; KNN, K-Nearest Neighbor, SVM, Support Vector Machine.

Data Availability

The data are publicly available on the UCI machine learning repository.

Funding

The authors received no funding for this work.

Disclosure

The authors reported no conflicts of interest for this work.

Acknowledgments

The authors gratefully acknowledge the UCI Machine Learning Repository [18] for publicly releasing updated datasets utilized in this research.

IX. Author Contributions

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work.

X. REFERENCES

- [1]. R. Kasabe and G. Narang, "Cardio Vascular Ailments Prediction and Analysis Based On Deep Learning Techniques", *Int J Eng and Appl Phys*, vol. 1, no. 2, pp. 174–178, May 2021.
- [2]. M. Raihan, P. K. Mandal, M. Islam, T. Hossain, P. Ghosh, S. Shaj, A. Anik, M. Chowdhury, S. Mondal, and A. More, (2019) "Risk Prediction of Ischemic Heart Disease Using Artificial Neural Network", in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, pp. 1-5.
- [3]. P. Palma, (2015). "Heart diseases now a plague in Bangladesh", *The Daily Star*.
- [4]. M. Raihan, M. Islam, P. Ghosh, S. Shaj, M. Chowdhury, S. Mondal, and A. More, (2018) "A Comprehensive Analysis on Risk Prediction of Acute Coronary Syndrome Using Machine Learning Approaches", in 2018 21st International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 1-6.

- [5]. Gomathy, B. T. Nagamani, S. Logeswari, Heart Disease Prediction using Data Mining with Mapreduce Algorithm, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.
- [6]. Sayantan M., Shouvik B., Anamitra B. R., Nilanjan D.; Wavelet-Based QRS Complex Detection of ECG Signal' International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 3, May-Jun 2012, pp.2361-2365
- [7]. Miranda G., Shailendra N., Singh E., (2015): Predictions in heart diseases using techniques of data mining.
- [8]. Golande M., Chandrasegar T., Gautam S., (2019): Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7.
- [9]. Fahd S. A.; "Implementation of Machine Learning Model to Predict Heart Failure Disease", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.
- [10]. Theresa (2019): Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7.
- [11]. Amma, N.G.B "Cardio Vascular Disease Prediction System using Genetic Algorithm", IEEE International Conference on Computing, Communication, and Applications, 2012.
- [12]. Hasan MK, Islam MM, Hashem MMA (2016) Mathematical model development to detect breast cancer using multigene genetic programming. In: 2016 5th International Conference on Informatics, Electronics, and Vision (ICIEV). IEEE, pp 574–579
- [13]. Islam A.S., Milon I.M.; (2019) Diabetes Prediction: A Deep Learning Approach. Int J Inf Eng Electron Bus 11:21–27.
- [14]. Hasan M, Islam MM, Zarif MII, Hashem MMA (2019) Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet of Things 7:100059.
- [15]. Islam M, Neom N, Imtiaz M, Nooruddin S, Islam M, Islam M (2019) A Review on Fall Detection Systems Using Data from Smartphone Sensors. Ingénierie des systèmes d'Inf 24:569–576.
- [16]. Nooruddin S, Islam MM, Sharna FA (2020) An IoT-based device-type invariant fall detection system. Internet of Things 9:100130.
- [17]. Oguine O.C., Oguine K. J., Okoli C. I., Oguine M.B.: "Comparative Analysis and Forecasting on the death rate of Covid-19 patients in Nigeria using Random Forest and Multinomial Bayesian Epidemiological models".: Journal of Clinical Case Studies Reviews & Reports; 2021.
- [18]. UCI Machine Learning Repository <http://archive.ics.uci.edu/ml>. Irvine, CA: University of California, School of Information and Computer Science.

Cite this article as :

Kanyifeechukwu Jane Oguine, Ozioma Collins Oguine, Chito Frances Ofodum, "Detection of Significant Risk Factors of Heart Diseases by Educating Multimodal Features and Implementing Machine Learning Techniques", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 4, pp. 621-630, July-August 2021. Available at doi : <https://doi.org/10.32628/CSEIT2174131>
Journal URL : <https://ijsrcseit.com/CSEIT2174131>