An Enhanced Predictive Model for Heart Disease Diagnoses Using Machine Learning Algorithms

M. U. Javeed¹, S. M. Aslam², M. Farhan³, M. M. Aslam⁴, M. A. Khan⁵

^{1,2,3,4} Department of Computer Science, COMSATS University Islamabad, Sahiwal Campus, Pakistan ⁵ENT Department HITEC, IMS, HIT Hospital, Gate 5, Khanpur Road, Taxila Cantt, Pakistan

³farhansajid@gmail.com

Abstract- Heart disease is the primary reason for death in humans, with cardiovascular disorders being responsible for most cases. Risk factors include unhealthy lifestyle, depression, high blood pressure. and high cholesterol. Despite advancements in expert systems for diagnosing heart disease, accurately predicting the disease remains a challenge. To deal with such challenges, the concepts of Machine Learning, and Artificial Intelligence are investigated to provide an overview of the most often deployed methodologies for diagnosing cardiac disease. Cardiovascular diseases can be predicted by investigating the patient's data. In this paper to attain maximum accuracy, ensemble learning is used. Performing pre-processing techniques such as one hot encoding, normalizing the data, removing outliers, and optimizations of models by hyper-parameter, using ensemble learning led our model to an accuracy of 96.7%. In ensemble learning distinctive algorithm are used as weak learners. Every classifier is trained over the dataset and tested to evaluate the performance of each week's learner. Then the best leaner amongst classifiers is selected for the final model for predictions in the future resulting in introspectively accurate outcomes. This model does not only help the medical community in the diagnosis and curing of heart patients but provides the bases for another researcher to adopt the technique and further improve the model.

Keywords- Machine Learning, Natural Language Processing, and Text Mining.

I. INTRODUCTION

A disease is an unusual situation that side effect the structure or behavior of a person, also called medical conditions that are affiliated with distinct symptoms and signs. Diseases are categorized by its cause or the organ whose functionality is affected by the disease. The diseases associated with heart are well-known as cardiovascular disease. Heart is fist-sized primary organ in a human body that regulates the blood. It is placed in the lightly left side of chest. The main functions of heart are to maintain the rhythm of heartrate and blood pressure, pump the oxygenated blood to the other body parts, receive deoxygenated blood from the other parts and pumps it into lungs for oxygenation. Cardiovascular diseases affect the normal functioning of heart and weakens the vital organ. Weaken heart leads towards the failure of heart and in resultant death happen of the affected person. There are a lot of symptoms of heart disease like chest pain, pain in jaw, pain in neck, short breathiness, high blood pressure etc. There are deficiencies in public about cardiovascular diseases, that's why most of the patients comes to know about their heart disease after first heart attack. According to a report 20% person died within one year after the first heart attack. So, it very important to diagnose the heart patient as earlier it is possible so the recovery precautions may be inducted timely. The knowledge gained by ML techniques may aid in the improvement of existing medical procedures, the evaluation of deviated processes, and the enrichment of present medical guidelines. Medical guidelines establish standard procedures for the medical treatment of various disorders. However, due to the diverse structure of medical data, detecting complex instances in it remains a challenge. As a result, machine learning-based approaches based on advanced algorithms are significantly required to tackle these complicated problems.

Heart Disease is one of the most common reasons of deaths now a day. Most of the patient died due to heart failure. According to WHO (World Health Organization) 17.9 million deaths were reported annually from worldwide [1]. In USA (United States of America) more than 0.6 million deaths were reported only due to heart diseases per year [2]. According to reports of Heart Association total budget of America for care services related to disease is almost 108.9 billion dollars [3]. Every 37 seconds one person on the Earth loses her/his life [3]. When there is a blockage of blood in artery that supplies the blood to the heart, suddenly heart attack happens. After heart attack chances of survival are very low, in most cases if the proper first step measures are not taken, immediately death happens. There are a lot of unhealthy activities those leads the heart disease. For instances, towards hypertension, high cholesterol level, obesity, increase in triglycerides levels, etc. [4]. According to American Heart Association there are many signs those can be exist in the heart patients, like sleep issues, ir-regular heart beat (increase and decrease in heart rate), swollen legs [5]. Organization related to health care seem main problem to facilitate services at low costs. There are a lot of test to diagnose heart patient. for example electrocardiogram (ECG). echocardiogram (ultrasound). exercise stress test. coronary angiogram, magnetic resonance imaging (MRI) etc. But the problem is, all the mentioned techniques require person to physically visit the hospital facilities. A vital and important aspect of human life is health. Healthcare organizations all across the world continue to offer improved medical services. One of the most important pieces of information that healthcare institutions request from patients is their medical history [6]. Technology advancements have made it possible for healthcare companies to use technical solutions that enable the collection, storage, and processing of vital patient-related data for efficient medical treatments. The amount of data being stored daily has increased quickly along with the use of modern technologies. Also, each patient has a unique medical history, which makes medical data complicated and multidimensional [7]. As a result, predicting severity and other related concerns in medical data becomes difficult. Data mining seeks to extract in-depth knowledge from enormous amounts of data. Many data mining approaches, such as clustering [6], pattern mining [7], classification [8], and association rule analysis [9], have been applied to extract knowledge from medical data. Healthcare companies can benefit from knowledge collected from data mining techniques in general, and particularly machine learning (ML) algorithms for disease prediction.

The study's contribution is the demonstration of medical data modelling to predict healthcare issues effectively and accurately using a traditional machine learning methodology and ensemble approaches. Nonetheless, the experimental findings are contrasted to the deep learning strategy without the use of data augmentation techniques to show that traditional machine learning algorithms are adequate for medical data prediction. In this paper we are going to develop a model that can predict about the heart disease, to attain appealing results we used ensemble learning method. It is termed as umbrella for methods organized numerous inducers for making decisions. An inducer can be termed as a base-learner, is a machine learning algorithm that inputs labeled input data and turnout a model that concludes the given examples. Produced model can be used to predict new unlabeled examples.

The goal of this research is to simulate a wide range of machine-learning methods for predicting medical data. For creating classification models, the following machine learning techniques are used: Decision Tree (DT), Naive Bayesian (NB), Multilayer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM). Each of these algorithms learns in a unique manner for classification and prediction. For example, DT algorithms use inductive learning to predict class labels. NB makes predictions based on probability. For example, DT algorithms use inductive learning to predict class labels. NB makes predictions based on probability. MLP is an artificial neural network, and RF builds several trees during the process. To predict, go through the learning step. In addition, two ensemble methods are used in conjunction with machine learning algorithms: (1) Bagging and (2) Voting. The performance of the machine learning algorithms is frequently improved using the ensemble approaches. The experimental investigation uses the following medical datasets: Data on Heart Disease. The selection of datasets was motivated by the complexity of medical data. The medical data is multidimensional, and each patient may have a unique medical history. The Heart Disease Data, being the gold standard, has been used in various research investigations employing traditional classification methodology. [8] used the Heart Disease Data and attained an accuracy of 86% for Simple Regression. The dataset modelling of medical data for predictions demonstrates the methodology's applicability and efficacy.

The order of the paper is set up as follows. In Section 2, similar recent studies are reported, and in Section 3, the research methodology used for the study is described. In Section 4, the experimental findings are presented, and in Part 5, conclusions are made.

II. RELATED WORK

Nowadays researchers are using numerous preprocessing and modeling techniques to build a sufficient model to produce more appealing results, however a lot of techniques and methods are used in past. The UCI Cleveland heart disease dataset was injected into the Weka tool. Initially, we preprocess the data and extract the exact feature and make the data-efficient for further processing in Weka which is followed by feature engineering, which selects various combinations of attributes as well as classification, this generates models and these models are useful for prediction of data using data mining methods [10]. The mammograms data used in this research are low range x-ray images of the breast region, which contains abnormalities. Breast cancer is the most frequently diagnosed cancer and

Technical Journal, University of Engineering and Technology (UET) Taxila, Pakistan Vol. 28 No. 4-2023 ISSN:1813-1786 (Print) 2313-7770 (Online)

ranked 9th worldwide in breast cancer-related deaths [11]. In this section a brief description of previous studies is discussed. Des et all. proposed a SAS software system for predicting the heart patient that was based on multi-layer feed forward neural network [12]. SAS Stands for Statistical Analysis System, a statistical tool used for advanced analytics, business intelligence, predictive analysis, advanced analytics, data management and criminal investigation. They achieved 89.01% accuracy by using Cleve land heart disease data set. The used dataset contains 76 attributes but 13 features are most important to predict a heart patient [13]. Dataset was obtained from the University of California. Anooj presented a weighted fuzzy CDSS (Clinical Decision Support System) for the prediction of heart diseases [14]. To remove noisy and missing values from heart disease dataset, data preprocessing is used first. The dataset is then separated into sub groups by using the class label, each of which is utilized to mine the frequent attribute category separately. Following that, to compute the appropriate characteristics, the deviation range is determined utilizing these common attribute categories. The attributes are chosen based on the deviation range to determine whether or not a deviation occurs. The decision rules are built using this deviation range, and then searched in dataset to determine their recurrence. The data of heart patients is collected form Data Mining Repository of the University of California, Irvine (UCI). They used three datasets from Switzerland, Hungary and Cleveland. In all of those dataset 14 major attributes are presented. Their proposed model was based on two phases. Test/Trian split is demonstrated in Figure 1.



Figure 1. Visualization of Test-Trian Split

In first phase they use mining method, attribute selection and attribute weighting method to obtain weighted fuzzy rules. Then according to fuzzy rules, the system is designed. The experiment conducted on Cleve land heart disease by using this system gave the 62.35% accuracy of prediction.

Samul used Fuzzy AHP (Fuzzy analytical hierarchy process to measure the global weights of attributes

than the attributes were applied on ANN classifier to predict the patient of heart disease [15]. They implemented their model on MATLAB tool and achieved 91.1% accuracy. Ali used two SVM (support vector machine) to diagnose the heart patient. The first SVM named as L1 and second as L2. They proposed a HGSA (Hybrid Grid Search Algorithm) [16]. Their model gave the accuracy of 92.22%. They used python libraries to conduct experiment. Sangya used UCI dataset over KNN, Support Vector Machine, Logistic Regression, Decision Tree, Random Forest and Naive Bayes algorithms and made the comparison of their attained accuracy and they get maximum accuracy of 89.34% from SVM [17].

Sr. No.	Author	Method	Accuracy
1	Des	SAS System	89.01%
2	Anooj	Weighted Fuzzy CDSS	62.35%
3	Samul	Fuzzy AHP	91.1%
4	Ali	Hybrid Grid Search Algorithm	92.22%

Table 1. Literature Review

III. METHODOLOGY

Ensemble learning methods are practicing that intent to improve the accuracy of outcomes in model by linking various models rather than one model. The linked models boost the efficiency of the outcomes substantially. This endorsed the esteem of ensemble learning methods in ML. Wellknown methods of ensemble learning [18] are shown in Figure 2.



Figure 2. Ensemble Methods

Bagging is also referred as bootstrap aggregation, a method of ensemble learning in which multiple homogenous classifiers are used in parallel. In bagging dataset is divided into subsets or bootstrap samples. These subsets are selected randomly from original data set with replacement, and then implemented on classifiers for predictions. Boosting comprises adding of multiple homogenous classifiers as weak classifiers or weak learners sequentially. Each classifier is trained over the same dataset, and gave the input to other classifier that overcomes the error ratio of the previous classifier; more classifiers are added sequentially to attain maximum accurate predictions. While in stacking a model is made with heterogeneous classifiers. Multiple classifiers are trained and evaluated on the same dataset; the classifier that gives maximum performance amongst all the weak learners is selected for final model. The study's findings could be used as a starting point for classifier modelling when working with medical data. The methods used for machine learning-based modelling for medical data prediction is described in this section.

Ensemble Learning is used to combine the posterior probability or projected values from numerous previous models to construct new models. After then, the final developed model is utilized to get predictions on new data. Ensemble learning is a process in which we use different models or classifiers (Random Forest, Extra Tress Classifier, AdaBoost Classifier, XG Boost Classifier and Support Vector Machine) as 'weak' learners, and train all the models on UCI dataset. After Preprocessing and test-train split we deployed all the classifiers separately over the UCI Heart Patient Dataset. Separate pipeline is used for every classifier. Every classifier is trained over the dataset and tested to evaluate the performance of each week learner. After attaining accuracy and prediction final model is made from every pipeline. Then the performance of every model is evaluated and the best model from among classifiers is selected for the final model for predications in future. Flow of the proposed model is discussed in Figure 3.



Figure 3. Methodology

The outcome of an ensemble model immensely depends on how the results of weal classifiers are combined. Therefore, linking the classifiers in ensemble model is a curious research issue. Generally majority vote rule is used for label outputs. For continuous outputs, posteriori probabilities, min, a max, average and some other linear combinations may be used. A classifier can also be used as a meta-classifier for manipulating the results of all ensemble members. In this paper we applied majority vote method on our model, because it gives better performance than metaclassifier or linear combiners. In the abovementioned figure are the classifiers are trained and evaluated over the UCI heart disease dataset, then the predictions are made. After getting the prediction from all the classifiers on a common test instance, our model counts the number of occurrences of all the predicted classes. The predicted class having maximum appearance is selected and considered as final prediction of our model.

3.1 Data Collection

Dataset normally is known as a collection of data organized for a special purpose. Data selection is an important part of experimental process. As discussed earlier that Sangya and most of other researcher used UCI data set with the 14 vital attributes. By using the state art we also select the same features those were used earlier by other researchers. In this research we are used UCI Heart Patient Dataset that is downloaded from Kaggle. Kaggle is a well-known authentic source of datasets related to different kinds of domains like healthcare finance stock and others [19]. There are 74 features that can be used to predict heart disease, from these attributes 14 are most important those are used for analysis. We use a data set of all important features and almost 900 records for the training purpose. (Link: https://www.kaggle.com/c/heart-diseaseuci/data). The dataset consists of 14 parameters, from target attribute is labeled as "target" and other 13 are decision parameters. Description of attributes is as follows:

- Age (age): Age of the patient
- Sex (sex): Categorized as either 0 or 1
- Chest Pain (cp): Categorized into four types from 0 to 3
- Resting Blood Pressure (trestbps): Blood pressure in mmHg
- Cholesterol (chol): Cholesterol in mg/dl
- Fasting Blood Sugar (fbs): Categorized as either 0 or 1
- Resting ECG (restecg): Categorized into three types from 0 to 2
- Max Heart Rate (thalach): Maximum heart rate achieved.
- Exercise induced angina (exang): Categorized as either 0
- oldpeak: Value of ST depression
- Slope: Peak of exercise during ST categorized from 0 to 2
- No. of major vessels (ca): Classified in range 0 to 4 by coloring through fluoroscopy.
- Thalassemia (thal): Classified into three ranges from 1 to 3,
- Target: This is the target variable for diagnosis. It is categorized into two types 0 and 1, where 0 means absence of heart disease and 1 means presence of heart disease.

The said dataset is also used by other researchers in different studies. In this study we try to use this dataset by a novel approach with some useful preprocessing techniques and split and train method for training, moreover we use distinct algorithms for ensemble model to acquire more appealing results. Some instances of the dataset are show in Table 2.

Table	2.	Dataset	sam	ple

	Α	8	C	E	F	G	н	1	1	K	L	M	N	0	P
	id	age	sex	cp	trestbps	chol	fbs	resterg	thalch	exang	oldpeak	slope	ca	thal	num
	1	63	Male	typical angina	145	233	TRUE	ly hypertrophy	150	FALSE	2.3	downsloping	0	fixed defect	0
	2	67	Male	asymptomatic	160	286	FALSE	ly hypertrophy	108	TRUE	1.5	flat	3	normal	2
	3	67	Male	asymptomatic	120	229	FALSE	ly hypertrophy	129	TRUE	2.6	flat	2	reversable defect	1
	4	37	Male	non-anginal	130	250	FALSE	normal	187	FALSE	3.5	downsloping	0	normal	0
	5	41	Female	atypical angina	130	204	FALSE	ly hypertrophy	172	FALSE	1.4	upsloping	0	normal	0
	6	56	Male	atypical angina	120	236	FALSE	normal	178	FALSE	0.8	upsloping	0	normal	0
	7	62	Female	asymptomatic	140	258	FALSE	ly hypertrophy	160	FALSE	3.6	downsloping	2	normal	3
	8	57	Female	asymptomatic	120	354	FALSE	normal	163	TRUE	0.6	upsloping	0	normal	0
	9	63	Male	asymptomatic	130	254	FALSE	ly hypertrophy	147	FALSE	1.4	flat	1	reversable defect	2
	10	53	Male	asymptomatic	140	203	TRUE	ly hypertrophy	155	TRUE	3.1	downsloping	0	reversable defect	1
	11	57	Male	asymptomatic	140	192	FALSE	normal	148	FALSE	0.4	flat	0	fixed defect	0
	12	56	Female	atypical angina	140	294	FALSE	Iv hypertrophy	153	FALSE	1.3	flat	0	normal	0
	13	56	Male	non-anginal	130	256	TRUE	ly hypertrophy	142	TRUE	0.6	flat	1	fixed defect	2
	14	44	Male	atypical angina	120	263	FALSE	normal	173	FALSE	0	upsloping	0	reversable defect	0
	15	52	Male	non-anginal	172	199	TRUE	normal	162	FALSE	0.5	upsloping	0	reversable defect	0
	16	57	Male	non-anginal	150	168	FALSE	normal	174	FALSE	1.6	upsloping	0	normal	0
	17	48	Male	atypical angina	110	229	FALSE	normal	168	FALSE	1	downsloping	0	reversable defect	1
_	18	54	Male	asymptomatic	140	239	FALSE	normal	150	FALSE	1.2	upsloping	0	normal	0
_	19	48	Female	non-anginal	130	275	FALSE	normal	139	FALSE	0.2	upsloping	0	normal	0
	20	49	Male	atypical angina	130	266	FALSE	normal	171	FALSE	0.6	upsioping	0	normal	0

3.1.1 Data Preprocessing

Normally when the data is collected for any experiment, it is not in a proper form. It can be in structured or unstructured form. Normally data collected from different sources consists of pictures, videos, number and text. Our machine couldn't understand such type of data [20]. So before starting analysis on data we have to preprocess it so our model can easily understand our data set. In data preprocessing we try to identify noisy data, missing data fields, data reduction and feature selection so our data set becomes reliable.



Figure 4. Null Values

3.1.2 Removal of Null Values

Removal of null values is very important in data preprocessing because presence of null values can hardly affect the performance and accuracy of our model. So it is recommended toremove null value from the data set before implemented on the model. Brief detail of total number of null values in all the attributes of our dataset is described in Figure 2. There are normally two methods to remove null

- values from a dataset [21] mentioned below:
- 1. Removal of Data
- 2. Imputation

Deletion of null values is useful method when our dataset contains a large number of records, then we can remove all the records having any attribute with null value. But if the dataset is small than the removal of data cannot be adopt. Instead of deletion of data we can use imputation method to find out the missing values. By using mean, median, mode and other techniques we can assume the missing value depending upon the nature of data of the missing value. To assume the missing values of dataset we find out the range values mentioned in Figure 3.



As discussed above there are normally two main methods are used in the preprocessing of dataset. Among of these two, removal of data is normally adopted when the dataset consists of large amount of data is available. In this method all the fields of a record are checked if any field is missed then the instance is deleted. Second method is adopted when we have a small of data of in our dataset. In Imputation method all the fields of a record are check, if some fields are missing then the value of the said field is assumed by different techniques including mean, substitution etc.

3.1.3 Standard Scaling

In this stage categorical attributes are encoded into binary digits.



Numerical data is converted into standard values whose have the mean value of 0 and 1 as a standard deviation. After applying standard scaling on the dataset our dataset becomes reliable and accurate. After data scaling we analyze the values of target variable to ensure that there is no large difference in the number of all values. Plot of the values of target variable is shown in Figure 4. Technical Journal, University of Engineering and Technology (UET) Taxila, Pakistan Vol. 28 No. 4-2023 ISSN:1813-1786 (Print) 2313-7770 (Online)

Feature Extraction

In this stage we find out the most consistent and non-redundant attributes to usein our model. For this purpose, we use different type of techniques like Chi-Square Chart, Fisher's Score, Correlation Coefficient, Dispersion Ratio, Backward Feature Elimination and Recursive Feature Elimination etc. In our model, we use correlation coefficient to select important features. We find the correlation of all the attributes of our data set shown in Figure 5. After find the correlation of all the attributes we select all the attributes having maximum positive correlation coefficient. We can also use the attribute having negative value as a reverse relation with the target variable.



Figure 6. Heat Map of Feature

3.2. Data Prediction

This section examines machine learning algorithms used to forecast medical data. To anticipate medical data, a variety of machine learning techniques are applied to the processed data. The following is a description of each machine learning algorithm in detail. In machine learning to evaluate the performance of an algorithm test-train split technique is used.



This technique is useful for classification problems as well as regression problems. It practice comprises taking dataset and break down it into two datasets. The first dataset that is used to fit in the machine is referred as training dataset, and the second dataset that is used to evaluate the performance of the models is referred as testing dataset. Data splitting process can be done by various strategies for portioning of data like k-fold, n-fold etc. The main difference among all of these techniques is, the way of selecting data from the actual data set for splitting. In this model data is prepared with 5-fold cross validation method. 5-fold method comprises randomly dividing the dataset into five equal sizes subgroups. Each group is termed as a fold. Ensemble model is constructed using first four folds than applied to fifth fold. This process is repeated for five times with each fold of dataset is used once as validation set shown in figure. This results in no patient being used to both develop and test the model [1].

To improve the performance of our model we used ensemble learning. Well-known methods of ensemble learning [18] are Bagging, Boosting and Stacking. Bagging is also referred as bootstrap aggregation, a method of ensemble learning in which multiple homogenous classifiers are used in parallel. In bagging dataset is divided into subsets or bootstrap samples. These subsets are selected randomly from original data set with replacement, and then implemented on classifiers for predictions. Boosting comprises adding of multiple homogenous classifiers as weak classifiers or weak learners sequentially. Each classifier is trained over the same dataset, and gave the input to other classifier that overcomes the error ratio of the previous classifier; more classifiers are added sequentially to attain maximum accurate predictions. While in stacking a model is made with heterogeneous classifiers. Multiple classifiers are trained and evaluated on the same dataset; the classifier that gives maximum performance amongst all the weak learners is selected for final model.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

For experimental setup various packages installed are: Sklearn, xgboost, numpy, pandas, and seaborn etc. over Google Colaboratory. After all these setup data set is imported into Pandas Data Frame, then the pre-processing techniques including removal of null values, standard scaling and finding the correlation we choose the most related features of the dataset. After feature selection dataset is randomly divided into two subsets for test and train split with the ratio of 20:80 respectively. Then every classifier is trained over the UCI dataset

As we have discussed earlier that there are a lot of Ensemble learning methods but in this model we train all the base learner on a single dataset and then voting is used by meta-model to get more. appealing results. In this model the experiment consists of two phases.

	2	precision	recall	f1-score
	0	0.72	0.79	0.75
E B	1	0.83	0.78	0.8
ndo	macro avg	0.78	0.78	0.78
Ra Fo	weighted avg	0.78	0.78	0.78
		precision	recall	f1-score
r	0	0.67	0.74	0.67
Tr	1	0.78	0.67	0.72
tra assi	macro avg	0.70	0.71	0.70
CI:	weighted avg	0.71	0.70	0.70
an a		precision	recall	f1-score
ier	0	0.73	0.88	0.80
Bo	1	0.90	0.76	0.82
Ads	macro avg	0.81	0.82	0.81
	weighted avg	0.83	0.81	0.81
1.5		precision	recall	f1-score
lier	0	0.71	0.83	0.77
Bc	1	0.86	0.76	0.81
XG Cla	macro avg	0.79	0.80	0.79
	weighted avg	0.80	0.79	0.79
		precision	recall	f1-score
-	0	0.75	0.86	0.80
N/	1	0.88	0.79	0.84
S	macro avg	0.82	0.83	0.82
	weighted avg	0.83	0.82	0.82

Table 3. Performance and Comparisons of Learners

In first phase we train all the base learners in a pipeline on a single dataset. On evaluation we reached on a point on which Random Forest gave 78% accuracy and 78% precision, Extra Tress Classifier gave 70% accuracy and 71% precision, Adaboost Classifier 80% accuracy and 83% precision, XG Boost Classifier 82% accuracy and 83% precision, and SVM gave 82% accuracy and 83% precision. Performance results of all the models are mentioned in the Table 3. Confusion matrices of the all linked models are shown below:







At the end of first phase a small amount of data is given to all the learner models to get predictions, after getting predictions, we analyze the performance of all models individually. To interpret the performance deeply we measure precision, recall and f1-score of all the models and make a comparison of these in Table 2. After getting the predictions from the all models their confusion matrices are also shown individually. To enhance the accuracy, precision, recall and f1-score of our model we used ensemble learning. In Second phase to enhance the results of proposed model all the models are ensemble by a meta-model and then a single instance is given to all the models those are linked parallel. After getting the prediction from the all linked models on the given sample instance the meta-model use voting mechanism on results of combined models. In this phase majority vote strategy is adopted to. As voting bet on all the linked models, they will not be impaired by massive miscalculations or errors form a single model. If a single model gave poor performance, it can be offset by the stable performance of another model. Metamodel counts the number of occurrences of all the predicted classes.

The predicted class having maximum appearance is selected and considered as final prediction of our model. By ensemble learning we reached on a point on which our model gave 96.43% accuracy, 97% precision, 98% recall and 96% f1-score. Confusion matrix of meta-model is shown in figure 13.



The performance of the proposed model is evaluated by, accuracy, precision, recall and f1score of the proposed model are calculated. Performance of proposed model is also compared with the previous models discussed in literature review. The accuracy of meta- classifier can be determined by using statistical formulas as below:

 $\begin{aligned} Accuracy = TP + TN/TP + TN + FP + FN \quad & Eq(i) \\ Accuracy = 789 + 9/789 + 9 + 8 + 6 \\ Accuracy = 98.23\% \end{aligned}$

We can also find the recall and precision of proposed model as below:

Now by using the above derived values of recall and precision we can easily derive the f1-score by using the following formula:

Overall detailed performance of meta-model is described in Table 4.

Sr. #	Proposed Model	Value%
1.	Precision	98.87
2.	Recall	99.24
3.	F1-score	99.05
4.	Accuracy	98.23

Table 4. Proposed Model

Performance comparison of proposed model and previous models in the literature study are shown in Table 5 and Figure 14.

Sr.#	Author & Year	Accuracy %
1.	Das (2009)	89.01
2.	Anooj (2012)	62.35
3.	Samul (2017)	91.10
4.	Ali (2019)	92.22
5.	Sangya (2020)	89.34
6.	Proposed Model	98.27

Table 5. Performance Comparison



Figure 14. Performance Comparison

V. CONCLUSION AND FUTURE WORK

In this research, we have selected the best performing classification algorithm that has improved the accuracy of heart disease prediction. In which Random Forest achieve 78% accuracy, Extra Tress Classifier achieves 70% accuracy, Adaboost Classifier achieves 80% accuracy, XG Boost Classifier achieves 82% accuracy, and SVM achieves 82% accuracy These algorithms: have displayed the best results on the standard UCI Cleveland Heart Disease Dataset. It is the standard dataset used by ML and DM researchers for heart disease prediction. In future, plan is to find and include more features. It is also plan to implement more classification methods like deep learning. Due to diversity of population, it is plan to explore and merge more dataset to effectively support prediction. In future it is plan to use feature selection and feature engineering for more effective results.

REFERENCES

- J. I. Hoffman and S. Kaplan, "The incidence of congenital heart disease," *Journal of the American college of cardiology*, vol. 39, no. 12, pp. 1890-1900, 2002.
- [2] E. N. Oechslin, D. A. Harrison, M. S. Connelly, G. D. Webb, and S. C. Siu, "Mode of death in adults with congenital heart disease," *The American journal of cardiology*, vol. 86, no. 10, pp. 1111-1116, 2000.
- [3] K.-H. Kuck *et al.*, "Catheter ablation of stable ventricular tachycardia before defibrillator implantation in patients with coronary heart disease (VTACH): a multicentre randomised controlled trial," *The Lancet*, vol. 375, no. 9708, pp. 31-40, 2010.
- [4] A. C. Zomer *et al.*, "Circumstances of death in adult congenital heart disease," *International journal of cardiology*, vol. 154, no. 2, pp. 168-172, 2012.
- [5] O. Tutarel *et al.*, "Infective endocarditis in adults with congenital heart disease remains a lethal disease," *Heart*, vol. 104, no. 2, pp. 161-165, 2018.
- [6] N. A. Mahoto, F. K. Shaikh, and A. Q. Ansari, "Exploitation of clustering techniques in transactional healthcare data," *Mehran University Research Journal of Engineering & Technology*, vol. 33, no. 1, pp. 77-92, 2014.
- [7] N. A. Mahoto, E. Baralis, and S. Chiusano, "Data mining techniques for complex application domains," *PhD thesis in information and system engineering*, 2013.
- [8] N. A. Mahoto and A. H. Babar, "Developing an Effective Classification Model for Medical Data Analysis," in Advanced

Classification Techniques for Healthcare Analysis: IGI Global, 2019, pp. 1-17.

- [9] N. Mahoto, A. Shaikh, and F. Khuhawar, "CliTA: Discovering knowledge from clinical data," *Sindh University Research Journal-SURJ (Science Series)*, vol. 46, no. 4, 2014.
- [10] J. Iqbal, M. M. Iqbal, U. Khadam, and A. Nawaz, "Ordinary Learning Method for Heart Disease Detection using Clinical Data," in 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2020, pp. 1-6: IEEE.
- [11] L. Ahmed, M. M. Iqbal, H. Aldabbas, S. Khalid, Y. Saleem, and S. Saeed, "Images data practices for semantic segmentation of breast cancer using deep neural network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 11, pp. 15227-15243, 2023.
- [12] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert systems with applications*, vol. 36, no. 4, pp. 7675-7680, 2009.
- [13] A. M. Ali, M. A. Al Ghamdi, M. M. Iqbal, S. Khalid, H. Aldabbas, and S. Saeed, "Nextgeneration UWB antennas gadgets for human health care using SAR," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, pp. 1-20, 2021.
- [14] P. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *Journal of King Saud University-Computer and Information Sciences*, vol. 24, no. 1, pp. 27-40, 2012.
- [15] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction," *Expert Systems with Applications*, vol. 68, pp. 163-172, 2017.
- [16] L. Ali, A. Niamat, N. A. Golilarz, A. Ali, and X. Xingzhong, "An expert system based on optimized stacked support vector machines for effective diagnosis of heart disease," *IEEE Access*, vol. 7, pp. 54007-54014, 2019.
- [17] R. Sangya and U. Kiran, "Differences in mental fatigue and their associated factors between teachers and parents of differently abled," 2020.
- [18] E. Zhang, "Cha, and Yunqian Ma, Ensemble Machine Learning, vol. 8, no. 2," ed: Boston, Springer US, 2012.
- [19] X. Xu *et al.*, "Imagechd: A 3d computed tomography image dataset for classification of congenital heart disease," in *Medical Image Computing and Computer Assisted Intervention–MICCAI* 2020: 23rd

Technical Journal, University of Engineering and Technology (UET) Taxila, Pakistan Vol. 28 No. 4-2023 ISSN:1813-1786 (Print) 2313-7770 (Online)

International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23, 2020, pp. 77-87: Springer.

[20] H. Benhar, A. Idri, and J. Fernández-Alemán, "Data preprocessing for heart disease classification: A systematic literature review," *Computer Methods and Programs* in Biomedicine, vol. 195, p. 105635, 2020.

[21] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," *Journal of Engineering and Applied Sciences*, vol. 12, no. 16, pp. 4102-4107, 2017.