Grenze International Journal of Engineering and Technology, June Issue

A Comprehensive Review of Recent Advances in Heart Disease Prediction using Machine Learning Algorithms with Optimization Techniques and Feature Selection

Sobia Mir[1] and Dr. Sunanda[2]
[1]MTech student computer science and engineering Shri Mata Vaishno Devi University, Katra, India
Email: sobiamir24@gmail.com
[2]Assistant Professor school of computer science and engineering Shri Mata Vaishno Devi University, Katra, India Email: Sunanda.gupta@smvdu.ac.in

***Abstract*—The biggest challenge with heart disease is its detection. There are various tools available that can predict the heart disease; but they are expensive when analysing the chances of heart disease in humans. Early prediction of heart disorders helps reduce the mortality rate and other problems. In this modern world, we have huge amount of data, therefore a variety of machine learning approaches can be used to detect heart disease accurately and efficiently as well. This review paper aims to provide an overview of the current state of Machine learning-based heart disease prediction, including the various Machine learning techniques and models that have been used, the data sources, optimization techniques, feature selection methods employed, and the performance of these models in terms of accuracy. Overall, this review paper will provide a comprehensive overview of the current state of ML-based heart disease prediction and help guide future research in this field.**

***Index Terms*— Heart disease, Machine Learning, Optimization techniques, Feature Selection.**

I. INTRODUCTION

Heart disease disorders are regarded as the most common among all fatal diseases. Medical professionals undertake many surveys on heart disorders to collect data on heart patients, their symptoms, and the course of the condition. More cases of humans with common diseases that exhibit typical symptoms are being documented. People in this fast-paced world want to live very opulent lives, so they work extremely hard to earn a lot of money and lead comfortable lives. As a result, they do not take care of themselves, which causes them to change their eating habits and lifestyles altogether, become more tense, develop blood pressure and sugar issues at a young age, and neglect to give themselves enough rest, eat what they are given and don't even care about the food's quality when they are ill, they take their own prescription, and all of these minor acts of carelessness add up to a significant threat: heart disease. According to World Health Organization study, due to heart disease, a huge number of individuals worldwide pass away every single year.
Machine learning is used to make it easier to make predictions and decisions from the wide range of data generated by the industry of healthcare. ML techniques can be of great assistance. Irrespective of the fact that

Grenze ID: 01.GIJET.9.2.8_3

heart disease can be present in a number of ways, perhaps vital risk factors are there that determine if someone would be at risk for certain conditions or not.
Among all the methods, there are numerous ways to diagnose heart disease. Most doctors employ angiography, a well-known technique, all around the world. There are several restrictions related to the angiography procedure, though. These drawbacks encourage the expansion of a non-invasive technology for heart disease prediction because the procedure is quite expensive and requires doctors to consider numerous aspects in order to identify a patient. These conventional procedures, which deal with patient health reports, are time consuming, they may produce inexact results since they are carried out by people, and they deal with patient medical reports. We need an automated method in order to prevent these errors and to get the best and fastest outcomes. Researchers had discovered over the past few years that ML techniques are particularly effective in examining the medical data. The datasets are provided to ML algos, and those algos operate in accordance with their nature and produce results. By altering the algos, machine learning mainly concentrates on acquiring and self-learning strategies. Machine learning is also mainly effective in managing a broad range of data, whether it is labelled or not. Large-scale optimization uses machine learning extensively for classifying medical data. Outlier and imbalanced data, however, may appear in the ground of machine learning and have an effect on the efficiency of the prediction model. Optimizations are also needed in Machine learning Algos to reduce the computation and identify the most significant features from the dataset. It may have an impact on how well machine learning algorithms perform overall. However, when developing machine learning algorithms for vast amounts of data, optimization approaches can be crucial. Current presumptions need to be examined because machine learning models are becoming more complicated. Due of this reason, machine learning optimization and feature selection is quite helpful.

II. LITERATURE SURVEY

A variety of experiments and research projects have been conducted in recent years, with the relevant major papers being published, with rising development in the fields of medical science and machine learning. [1] RAHMA ATALLAH et al. in 2019 used the CLEVELAND DATASET, a famous UCI repository, to create extra detailed and reliable model to detect the possibility of having a cardiac condition. Accuracy of 90% was achieved by using ensemble learning model which surpassed other classifiers.
[2] AMIN AL HAQ et al. in 2019 executed trials to gain the effectiveness of various algorithms. The accuracy of the models was increased by using ensemble learning approaches like bagging, boosting, and stacking. Additionally utilized for classification, the BPNN technique was contrasted with ML Algorithms. With the help of relief feature selection, the SVM was able to classify the features with an accuracy of 88%. The support vector machine classification accuracy was greatly enhanced by ensemble learning approaches, increasing it to 92.30%. In their tests, the BPNN had a 93% accuracy rate. In the end, the deep neural network method (BPNN) was more successful in detecting heart disease based on all experimental study and outcomes. [3] LIAQAT ALII et al. in 2019 proposed an automated diagnostic system for HD diagnosis. The projected classification employed deep neural network for classification and the 2 statistical model for feature refinement.
Using 6 distinct evaluation criteria, the projected diagnostic system’s strength was assessed. This proposed method increased the accuracy for HD detection by using chi square-DNN model 93.33%. Although the present study didn’t look at hybrid diagnostic system's predictable time complexity.
[4] NORMA LATIF FITRIYANI et al. in 2020 introduced an efficient heart disease detection model for heart disease detection. This model integrates SMOTE_ENN, DBSCAN, and XG Boost based machine learning algos. Outlier data were found and removed using DBSCAN, the uneven training dataset were balanced using SMOTE_ENN, and the prediction model were learned and created using XGBoost MLA. The experimental outcomes showed that the suggested model surpassed leading models in performance and earlier research findings, with accuracy levels of up to 98.40% and 95.90% for the Cleveland and Statlog datasets, respectively.
[5] JIAN PING LI et al. in 2020 proposed a diagnosis system based on ml for heart disease prediction. This system was designed using algorithms. The feature selection problem was solved using traditional feature selection algorithms. The Cleveland heart disease dataset was utilised. The effectiveness of the identifying system was further evaluated using performance evaluation indicators. In comparison to previously proposed approaches, the accuracy of support vector machine using this feature selection algorithm (FCMIM) was 92.3%, which was quite decent.
[6] JIKUO WANG et al. in 2020, proposed a stacking based 2-level model in which the two levels are base and metalevel. The base-level prediction classifiers were used as input for the meta-level classifiers. Classifier with lowest correlation was initially determined by means of the greatest information and Pearson-correlation

coefficients. The top classifiers to combine to achieve the best final result were then found using the enumeration approach. According to experimental findings, the suggested model successfully detected CHD by sensitivity, accuracy and specificity of 95.84%, 95.43%, and 94.44% respectively.

[7] In 2020, Bayu Adhi Tama et al. projected (CHD) prediction model using ensemble. Three separate ensemble learners were stacked to create the suggested methodology. Four independent publicly accessible datasets—namely, Z-Alizadeh Sani, Cleveland, Statlog and Hungarian were used to evaluate the proposed approach. According to the discoveries of the experiments, the model was able to exceed cutting edge CHD recognition approaches in relationship of accuracy, F1, AUC value. Accuracy for the datasets from Z-Alizadeh Sani, Cleveland, Statlog and Hungary was 98.13%, 93.55, 86.49, and 91.18%, respectively.

[8] ABDALLAH ABDELLATIF et al 2022 proposed a model that integrates Inf-FSs, IWRF, and BO. With the help of the three techniques, the most chief features are chosen, the imbalanced data classification problem that arises in medicinal datasets is addressed, and the weighting factor is adjusted. Two open datasets were used to evaluate the generated model and compare it to earlier research. According to the experimental findings, the planned model was more successful in producing better outcomes without altering the data distribution. Additionally, the suggested improved weighted random forest outperforms the standard RF in terms of CVD detection by 3.62% and 6.3%.

[9] ABDALLAH ABDELLATIF et al 2022 projected an effective machine learning model decision support system to detect the presence or absence of HD and to check the severity level. ET, SMOTE, and HB were all included in the suggested model. In terms various evaluation parameters, the results exposed that the model beats other by up to 99.2%, 99.33%, 99%, and 0.983 for the Cleveland dataset and 98.52%, 98.8%, 98.080%, and 0.960% for the Statlog dataset. For multiclass (severity level) classification, the model achieved 0.939% ,95.73%, 96.35%, 95.73%, 95.78%, and of MCC, Accuracy, Precision, recall, f1-score, respectively.

GHULAB NABI AHMAD et al in 2022 proposed a medical diagnosis of heart disease prediction using Machine Learning and GridSearchCV. They used many machine learning models like KNN, GBC, SVM, LR together with GridSearchCV. They also used a 5-fold cross-validation for the verification. They considered many datasets like Cleveland, Switzerland, Hungary, long BeachV. They found through their experiment that XGBoost classifier with GridSearchCV gave the highest accuracy of testing 100% and training 99% for all the datasets. It was also found that XGBoost also gave the best accuracy without GridSearchCV among all the other classifiers. Table 1. shows the comparative study of the literatures that we have discussed above in this section.

TABLE I. COMPARATIVE STUDY OF LITERATURE REVIEW

Authors	**Year**	**Dataset**	**Classifiers**	**Feature selection used**	**Accuracy**
Atallah et al. [1]	2019	Cleveland	Machine learning ensemble techniques: SGD, KNN, Random Forest, Logistic Regression, Hard Voting Ensemble Method	No	Accuracy of Hard Voting Ensemble Method: 90%
Haq et al [2]	2019	Cleveland	SVM, LR, KNN, NB, ANN, DT and Ensemble learning techniques such as bagging, boosting, and stacking, BPNN algorithm	Relief	DT-78.54%, NB- 86%, SVM- 92.30% BPNN-93%
Ali et al. [3]	2019	Cleveland	DNN	Chi square (for feature elimination)	prediction accuracy of 93.33%
Fitriyani et al. [4]	2020	Statlog and Cleveland	SMOTE-ENN, and XGBoostbased ML algorithm.	N/A	Accuracy: 95.90% for statlog dataset. 98.40% for Cleveland dataset.
Li et al. [5]	2020	Cleveland	KNN, LR, ANN, SVM, NB, and DT	Relief, MRMR, LASSO, LLBFS, FCMIM	Accuracy of SVM with FCMIM: 92.37%

Wang et al. [6]	2020	Z-Alizadeh Sani, Statlog, SPECTF, big dataset	ML: Extra tree, RF, ADB, SVC, multi-MLP, XGB, GPC, GNB, LR, GB	Chi square, RFECV, XGB	Accuracy: 95.43% on Z-Alizadeh Sani Accuracy: 90.7% on statlog 92.2% on SPECTF dataset 73.2% on big dataset
Tama et al. [7]	2020	Z-Alizadeh Sani, statlog, Cleveland, and Hungarian	RF, GB machine, extreme gradient boosting machine	Correlation-based FS + PSO	Accuracy: Two-tier ensemble PSO-based feature selection-98.13% (Z-Alizadeh Sani), 93.55(statlog), 86.49(Cleveland), 91.18% (Hungarian dataset).
Abdellatif et al. [8]	2022	Statlog and Cleveland	Improved Weighted Random Forest, and Bayesian optimization	INF-FSs	Accuracy: IWRF (with FS)-97.7%(statlog) 95.9%(Cleveland) 98.3% with optimization (statlog), 97.2% with optimization (Cleveland).
Abdellatif et al. [9]	2022	Statlog and Cleveland	SVC, KNN, LR, SGD, TREEBASED ENSEMBLE SMOTE, ET.	No	Accuracy: (SMOTE+ ET) 99.2%(Cleveland), 98.52%(Statlog), 95.73% (severity level).
Ahmad et al. [10]	2022	Cleveland, Hungary, Switzerland, longbeachCV	LR, SVM, XGBoost, RF. KNN GridSearchCV	No	Accuracy: XGBoost with GridSearchCV: - 100%

III. Generic Approach for Heart Disease Prediction

"Fig 1". shows the generic flow chart that outlines how heart disease prediction system works.

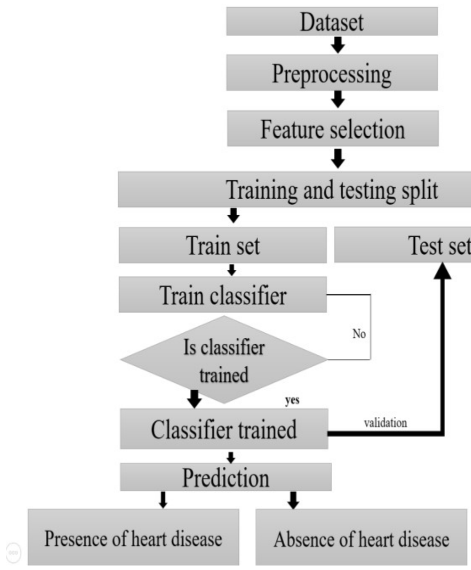


Figure 1. Generic Flowchart of heart disease prediction system

A. Collection of datasets

Heart Disease Datasets: These datasets are available for free. The dataset's descriptions are provided below, and a summary of their attributes is provided in Table 2.

a) Z-Alizadeh Sani: There are 303 patients in the sample, 216 of them have CHD. Each patient provides 55 input variables as well as a class label variable. The variables take into account a few aspects of the patient's profile, including demographics, symptomatology, physical examination, electrocardiography, and laboratory tests. The class label variable in the original dataset has four categories: normal, LAD, LCX, and RCA.

b) Statlog: The dataset's initial iteration contains 75 attributes. 13 features have been used for CHD identification by numerous researchers. There are no missing values, and 261 occurrences in which 114 patients had CHD were successfully gathered.
c) Cleveland: Three hundred three samples of healthy people and CHD patients make up the dataset. The original has 76 features however we take into account 13 features in line with previous studies. Because there are five integer values in the original one ranging from zero i.e.; no CHD to four i.e., severe CHD, class label features are normalised in two outputs i.e.; absence of CHD, No existence of CHD.
d) Hungarian: Overall 294 observations are present in this dataset and 13 input characteristics. Additionally, 106 patients have been diagnosed as CHD victims, while as the remaining patients are in good health that is CHD not found.

B. Data pre-processing

Data pre-processing involves a set of techniques to prepare the raw data for the model. It can include cleaning, normalizing, and transforming the data to make it suitable for the algorithm to work with. This can include tasks such as handling missing values, removing outliers, and scaling the data. It is an important step in the machine learning pipeline as the quality and structure of the data can greatly impact the performance of the model.

C. Feature selection

Feature selection is a method where in irrelevant, unimportant attributes/features are removed to reduce the complexity. Only a little portion of the dataset's variables is used to create a model for machine learning; the remaining are either unimportant or noisy. The performance gets affected with these noisy attributes in the dataset. To keep the performance and efficiency unaffected, it is important to identify and choose the most significant attributes from the dataset, which is achieved with the benefit of feature selection in machine learning. In [2] authors used Relief FS technique and the accuracy of SVM on features selected by relief was increased from 86% to 88%. In [5] authors used SVM with fast conditional mutual information (FCMIM) feature selection algorithm. The accuracy of SVM with the proposed FS was pretty good (92.37%) as compared to the previous proposed methods. In [8] authors have applied feature selection on different machine learning models (G-NB, LR, KNN, SVM, RF, XGBoost and IWRF) and the accuracy of each models improved using the same.
In “Fig 2”. We have shown different feature selection techniques available to us.

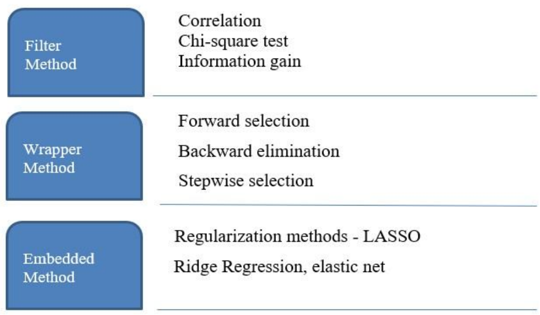


Figure 2. Different feature selection techniques

IV. Overview of Machine Learning Techniques for Classification

“Fig 3”. shows different machine learning techniques that are available to us.
There are several machine learning techniques that have been used for heart disease prediction including support vector machine, k nearest neighbour, decision tree, naïve bayes. Random forest, logistic regression, extreme gradient boosting. All these techniques have their own advantages and disadvantages and the best technique to use depend on the characteristics of the data and the specific problem. There are also many ensemble learning techniques such as bagging boosting, and stacking which are used for classification. Different researchers have applied different machine learning techniques for heart disease prediction and have achieved decent accuracies in their work. In [1] K was taken 7 and it provided the best results based on GridSearchCV. On unseen test data the accuracy of KNN was 82% and on running GridSearchCV the accuracy was found out to be 87% much better than the previous one. In [9] KNN is also used for severity level classification of heart disease and achieved a good accuracy. In [2] authors have used support vector machine for heart disease classification with full as well as selected features. The results declared that SVM with full features got low accuracy (86%) as compared to reduced feature with an improved accuracy of 88%. SVM is mostly used when there is a very pure

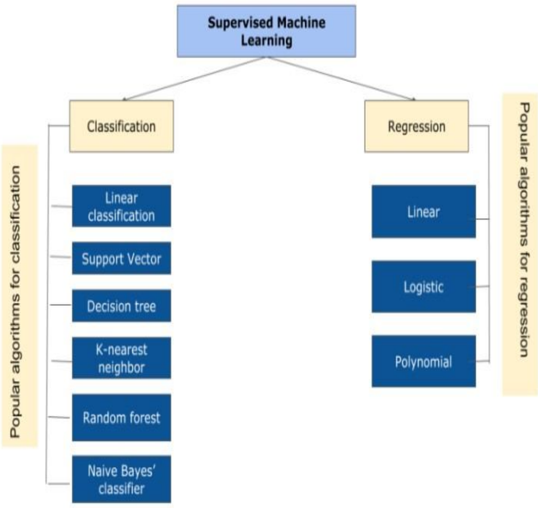


Figure 3. Different machine learning techniques

margin of separation between classes. Also, in [2] authors experimented that bagging method resulted in reducing the accuracy of SVM from 88% to 87%. The performance of SVM [5] was pretty good (92.37%) on selected features chosen by fast conditional mutual information (FCMIM) FS technique. As heart disease datasets are not much large and SVM is mostly used for heart disease prediction for the reason that SVM does not work better for large datasets. in [10] authors used LR, SVM, KNN, and GBC (gradient boosting classifier) with GridSearchCV and used a 5-fold cross validation system for the confirmation and in their experiment GBC with GridSearchCV gave the highest training and testing accuracy by 99.03% and 100% and GBC also resulted in highest accuracy without GridSearchCV among other classifiers concluding that GBC was the best model among other three models. GridSearchCV is a method in the scikit-learn library in Python that is used to tune the hyperparameters of a machine learning model. The method takes an estimator (such as a classifier or regressor), a parameter grid, and a scoring function as input. It then performs a search over the parameter grid by training and evaluating the estimator using different combinations of the parameters specified in the grid. The goal of GridSearchCV is to find the combination of parameters that results in the highest score according to the specified scoring function. In [8] the authors applied Improved Weighted Random Forest (IWRF) to Statlog and Cleveland dataset and resulted in improved prediction accuracy up to 95.5% and f- measure of 94% for Statlog dataset as compared to other existing models (G-NB, LR, KNN, SVM, RF, XGBoost). Improved weighted random forest is a variant of the random forest algorithm that assigns different weights to the individual trees in the forest based on their accuracy or performance. The idea behind this is that some trees in the forest may perform better than others, and by assigning higher weights to those trees, the overall accuracy of the forest can be improved. This can be done using techniques such as boosting or bagging. Additionally, it is possible to weight the individual observations in the dataset differently. This can be useful when there is a class imbalance in the data, as it allows the algorithm to give more importance to the under-represented class.

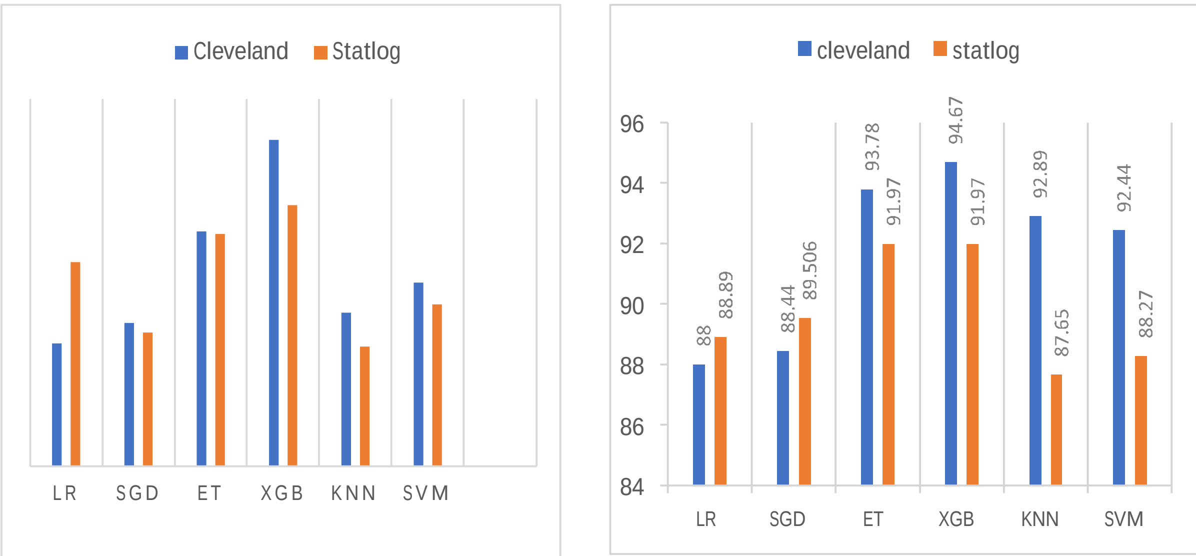
V. Overview of Optimization Techniques

In [7] authors have used PSO (particle swarm optimization-based feature selection) to choose the most significant feature set for each dataset and achieved an excellent accuracy of 98.13% on Z-Alizadeh Sani dataset. Inspired by swarm behaviour seen in nature like birds and fish schooling PSO is a powerful metaheuristic technique, mimics a simplified social structure. The goal of this technique is to graphically imitate a flock of birds doing an elegant but unpredictable composition. It deals with the complex and difficult problems in a move that create a connection between the environment and particular simple agents. PSO is relatively simple algorithm that is easy to implement and understand. It does not require the calculation of gradients. PSO is global optimization algorithm that means it can find the global optimum of a problem, even if it has multiple local optima. It also handles large number of parameters.

There are other bio-inspired optimization techniques which haven't been much used in recent researches like Grey Wolf Optimization, Genetic Algorithm etc. The meta-heuristic technique known as Grey Wolf Optimization (GWO) is used to categorize and predict data. GWO optimizer is now used in the majority of medical applications to predict outcomes with the highest degree of accuracy. A genetic algorithm (GA) is a computational method for optimization and search that is inspired by the mechanics of natural evolution. The basic idea behind a genetic algorithm is to model the process of natural selection, in which the fittest individuals in a population are selected to produce the next generation. These have many applications in the field of research like engineering, management and identifying various medical diseases. Selection, crossover and mutation are the three prime operators of Genetic Algorithm (GA). To elect the best individual in the population selection operation is the most important operation in GA. Genetic algorithms are best for multiclass classification. For detecting the severity level of heart disease this will be very helpful.

VI. Discussion

In this survey, the recent literature on heart disease classification utilizing machine learning algorithms, optimization methods and feature selection is revealed. At present there are massive data available universally. But then most of the datasets which are accessible are not structured correctly and not ordered in good manner for reuse of the algorithms. Even some of the datasets also have poorer quality to extract the vital attributes to achieve an exceptional precision. Many optimization and feature selection procedures are being used to extract the important features from the dataset. Data unbalancing is also a major issue and to overcome this problem SMOTE is used [9]. Synthetic Minority Oversampling Technique (SMOTE) is a method to handle the issue of imbalanced dataset. Authors [9] have applied SMOTE to various machine learning algorithms where Extreme Gradient Boosting (XGBoost) algorithm with SMOTE attained highest results than other models. The comparison of accuracies of various Machine Learning algorithms with and without using SMOTE is shown in “Fig. 4”. Authors [9] have also used Hyperband (HB) Optimization along with SMOTE to many machine learning algorithms where Extra Tree performed very well by achieving maximum accuracy of 99.2 %. Hyperband optimization is the method of electing a set of optimal hyperparameters for a machine learning algorithm. Accuracy and efficiency can be also improved using feature selection. Authors [8] have used these techniques Inf-FSs and IWRF where their model outperformed the other models by achieving an improvement of 2.4% and 4.6% in accuracy on Statlog and heart disease clinal records dataset respectively. Among the machine learning algorithms, it is seen that Extra tree used with SMOTE is resulting in highest accuracy as compared to other algorithms used with the technique SMOTE [9]. In [7] authors used a two-tier ensemble method and chose the best features by running different number of particles in Particle Swarm Optimization (PSO) and in their experiments it is shown that PSO with 20 particles is the better prediction performance on the Z-Alizadeh Sani dataset. From this study we can conclude that applying machine learning models along with optimization, feature selection (discussed in section 4) and balancing the dataset are very essential steps for predicting heart disease more accurately and efficiently.



(a) (b)

Figure 4. a) Accuracies of ML models before applying SMOTE. b) Accuracies of ML models after applying SMOTE [9]

VII. CONCLUSION AND FUTURE WORK

Heart disease is the deadliest disease among other diseases which causes people to die. Its investigation and diagnosis at the initial stages is the healthier method then cure. When heart disease is detected at right time, it will help the healthcare experts to prevent more problems. The use of machine learning techniques has shown a promising approach for identifying patients at risk of heart disease. By analysing a patient's medical history, symptoms, and test results, machine learning techniques can classify patients as having or not having heart disease with high accuracy. However, it is important to note that these models must be validated and tested on a large and diverse population to ensure their robustness and generalizability. Further research is also needed to improve the models and to develop more accurate and efficient methods for predicting heart disease. From this review we come to know that Feature selection methods and optimization techniques play a vital role in heart disease prediction by improving the accuracy and efficiency of machine learning models. Engineers and researchers can get support from this paper to analyse the best machine learning models for classification. Researchers in future can focus more on these techniques to provide a more generalised heart disease prediction system. In future researchers can also utilise the feature selection and optimization techniques for detecting the severity level (multi-class classification) of heart disease. The datasets mentioned in this review are public and freely available and accessible.

REFERENCES

[1] R. Atallah and A. Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method," 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923053.

[2] A. U. HAQ et al., "Identifying the Predictive Capability of Machine Learning Classifiers for Designing Heart Disease Detection System," 2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, 2019, pp. 130-138, doi: 10.1109/ICCWAMTIP47768.2019.9067519.

[3] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed and J. A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}}$ Statistical Model and Optimally Configured Deep Neural Network," in IEEE Access, vol. 7, pp. 34938-34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[4] N. L. Fitriyani, M. Syafrudin, G. Alfian and J. Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," in IEEE Access, vol. 8, pp. 133034-133050, 2020, doi: 10.1109/ACCESS.2020.3010511.

[5] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in IEEE Access, vol. 8, pp. 107562-107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[6] J. Wang et al., "A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease," in IEEE Access, vol. 8, pp. 37124-37133, 2020, doi: 10.1109/ACCESS.2020.2975377.

[7] Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble (hindawi.com)

[8] A. Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "Improving the Heart Disease Detection and Patients' Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest," in IEEE Access, vol. 10, pp. 67363-67372, 2022, doi: 10.1109/ACCESS.2022.3185129.

[9] Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods," in IEEE Access, vol. 10, pp. 79974-79985, 2022, doi: 10.1109/ACCESS.2022.3191669.

[10] G. N. Ahmad, H. Fatima, S. Ullah, A. Salah Saidi and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV," in IEEE Access, vol. 10, pp. 8015180173, 2022, doi: 10.1109/ACCESS.2022.3165792.