Early Prediction of Heart Disease using Modified Optuna-CNN Model: Performance Analysis

Jeen Retna Kumar R*, Shajeena J, Saumiya S, Bini Palas P, Anushia R M, Berakhah F Stanley

ABSTRACT - Heart disease is currently a major cause of death in the world. Primitive detection of heart disease is important for improving patient results and reducing the burden of cardiovascular disease on society. In recent years, deep CNN (convolutional neural networks) have shown promising results for detecting heart disease using patient parameters such as age, sex, blood pressure, cholesterol levels, BMI, and family history of heart disease. Also, choosing an appropriate hyperparameter is crucial for the success of model performance and obtaining optimum results. Optuna is a hyperparameter optimization algorithm that helps in finding the optimal set of hyperparameters. This paper presents a modified optuna-CNN model for early detection of heart disease by patient parameters. The procedure includes gathering information, preparing the data, choosing a suitable model, training the model, assessing its performance, and implementing it in a healthcare environment. The paper also discusses the potential benefits of using deep CNN for early detection of heart disease, including improved accuracy, sensitivity, and specificity, which can aid in early diagnosis and timely intervention. Overall, the application of deep convolutional neural networks (CNNs) to identify heart disease in its early stages based on patient characteristics is a hopeful field of study that could enhance patient results and minimize the consequences of heart disease on people and the community.

Index Terms – Heart disease Prediction, Cardiovascular disease, Convolutional neural network, Optuma optimization.

I. Introduction

One of the leading causes of death across the globe is heart disease, which accounts for around 33% of all deaths. Timely identification and treatment are critical to enhance patient outcomes and alleviate the impact of cardiovascular disease on society. In the past several years, deep convolutional neural networks (CNNs) have surfaced as a heart disease [1]. The use of deep CNN for early detection of heart disease offers several potential benefits, including hopeful and effective technique for identifying heart disease by considering patient factors, including age, gender, blood pressure, cholesterol levels, BMI, and familial history of improved accuracy, sensitivity, and specificity. By imposing patient parameters, deep CNN can identify patients at higher risk of developing heart disease, allowing for early diagnosis and timely intervention [2]. This article presents a procedure that involves utilizing deep CNNs to detect heart disease at an early stage based on patient characteristics. It covers various stages, including collecting data, pre-processing it, selecting an appropriate model, training the model, evaluating its performance, and implementing it in a clinical environment. Overall, the use of deep CNN for early detection of heart disease by patient parameters shows great potential for advancement in research and has the capability to enhance patient outcomes and reduce the impact of heart disease on individuals and society [3].

Artificial intelligence (AI) has been a revolutionary breakthrough for humanity, opening up a world of possibilities. AI has made significant advancements in all domains of life; it has extended its reach from rudimentary chat bots to self-driving cars and robots [4]. It has reinforced decision-making procedures and facilitated computer-based education by incorporating diverse artificial intelligence, which has made significant advancements in various fields, including biology, computer science, engineering, linguistics, mathematics, psychology, recognizing speech and faces, processing natural language, identifying images, and developing smart robots. [5]. Machine learning is a significant contributor to this advancement, enabling computers to learn from past experiences and data without explicit programming, thus making remarkable progress possible. With the increasing amount of data available, machine learning has become a valuable tool for efficiently handling and extracting useful information from raw data. In recent times, machine learning has made considerable progress in areas such as learning algorithms and pre-processing techniques, and its demand is on the rise due to the accuracy, consistency, and informative information it can provide from big data. Overall, the aim of machine learning is to facilitate machines acquiring knowledge and skills without requiring explicit programming, leading to remarkable advancements in various fields [6]. Deep learning, which was formally introduced as a subfield of machine learning in 2006, is a significant development that has significantly improved AI's capabilities. This is influenced by the neural networks found in the human brain and is distinguished by its use of numerous layers for data processing. Each layer takes input, applies weighted changes,

Manuscript received August 27, 2023; revised November 24, 2024.

Jeen Retna Kumar R is an Assistant Professor in the Department of ECE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India. (corresponding author e-mail: jejinrsrch@gmail.com).

Shajeena J is an Assistant Professor in the Department of CSE, SRM Institute of Science and Technology, Tiruchirapalli, India. (e-mail: shajeenaelmo@gmail.com)

Saumiya S is an Assistant Professor in the Department of ECE, Chennai Institute of Technology, Chennai, India. (e-mail: saumi2424@gmail.com).

Bini Palas P is an Assistant Professor in the Department of ECE, SRM Easwari Engineering College, Chennai, India. (e-mail: binipalas16@gmail.com).

Anushia R M is an Assistant Professor in the Department of CSE, DMI College of Engineering, Chennai, India. (e-mail: rmanushia85@gmail.com).

Berakhah F Stanley is an Assistant Professor in the Department of ECE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India. (e-mail: berakhahfstanley@gmail.com).

often as nonlinear functions, and then transfers it to the next layer. Deep learning has helped to overcome long-standing limitations in AI by excelling at uncovering complex patterns in large datasets. The wide-ranging applicability of deep learning has made it a valuable tool in numerous fields, including science, business, and government. [7].

Deep-learning models are based on a single architecture, whereas hybrid deep-learning (HDL) models are created by merging two or more deep-learning architectures or by combining deep-learning models with machine-learning models for classification purposes. The merging of individual deep-learning architectures is called hybridization. Despite the notable progress of deep learning in computer vision, there is a shortage of professionally curated, labeled datasets that are essential for training, testing, and validating algorithms. This remains a significant challenge not only in computer vision but also in the healthcare industry, where there is a scarcity of experts in clinical diagnostics [8]. The healthcare industry has been utilizing AI for the past three decades to enhance medical practice. The field has witnessed enhanced prospects owing to the recent advancements in machine learning and deep learning. AI-based medical technologies are rapidly growing and being implemented in clinical settings, spanning across a variety of medical services from diagnosis to surgery and remote patient care. These technologies are assisting healthcare professionals in identifying life-threatening diseases such as cancer, resulting in early detection, simplified treatment, and shorter hospital stays for patients. [9].

II. Related works

The development of AI-enabled robots is a remarkable achievement of AI technology, and it is finding increased use in several fields, including medicine. The integration of AI and robotics in healthcare has the potential to bring about a significant transformation in the provision of medical services. It offers an insight into how AI and robotics can transform healthcare. Scientists use AI to process data swiftly and accurately, leading to better treatment results for severe illnesses examples of severe illnesses that can benefit from AI-assisted treatment include Cancer and cardiovascular disease. [10].

Heart diseases, also known as cardiovascular diseases responsible for highest number in deaths globally, with heart failure and stroke accounting for 85% of those deaths, according to a 2019 report by the World Health Organization (WHO). The report indicated that heart diseases caused around 18 million deaths, which amounts to 32% of all deaths. Identifying heart diseases in their early stages is critical as 25% of individuals pass away suddenly without prior symptoms. As a result, it is essential to create a system that can identify heart diseases in their early stages. Out of the different types of heart diseases, coronary artery disease (CAD) is the most prevalent, and it is associated with individual behaviors like smoking, diabetes, excessive alcohol consumption, obesity, stress, and hypertension [11]. Diagnosis of Cardiovascular diseases typically requires significant time, effort, and resources. By incorporating image classification techniques, deep learning can aid healthcare providers in gathering useful insights about heart patients and enhancing diagnostic precision. [12]. Researchers who work on image classification with deep neural networks (DNNs) face a challenge of noisy image data. The denoising techniques are used to reduce or eliminate the noise in the images, which enables the classification algorithms to accurately identify the content of the images. By applying these techniques to these datasets, researchers can improve the accuracy and reliability of image classification systems even in the presence of noise.

Deep learning (DL) is type of machine learning (ML) that involves learning from data representation rather than task-based methods. Deep learning employs various learning techniques including supervised, unsupervised, or semi-supervised learning. DL models are influenced by the biological nervous system, but they have different structural and functional characteristics than the human brain. [13,14]. DL models, which include CNNs, deep neural networks, recurrent neural networks, and deep belief networks, have found diverse applications in research areas such as speech recognition, computer vision, audio recognition, natural language processing, machine translation, social media filtering, drug design, bioinformatics, medical image processing, board game programs, and material examination. These advanced ML models have produced outcomes that are similar to, or even surpass, human performance in some situations [15,16]. It is essential to have precise diagnosis and treatment for heart failure patients to achieve an accurate health evaluation. Therefore, it is necessary to investigate techniques that can enhance heart failure assessment. The field of computer-aided diagnosis (CAD) merges computer science with clinical medical knowledge to achieve this objective. [17]. Computer-aided diagnosis (CAD) is an interdisciplinary field that applies computational techniques to analyze complex medical data and generate user-friendly results for tracking and assessing various diseases. CAD is a useful tool for medical practitioners as it provides accurate reference results and alleviates their workload [18]. By reducing the chances of misdiagnosis and missed diagnosis, CAD can prevent medical errors. Additionally, CAD helps in developing personalized medical interventions, predicting potential risks, enhancing clinical indicators, and improving medical research, making it crucial to establish an effective model for diagnosing and classifying heart failure [19]. A less complex and fully automated model for predicting cardiovascular disease using CNN and fast fourier transform is proposed by Wafaa et al. [20]. Heart sound detection using CNN can be proposed by Shumpei et al. [21] with the help of utilizing heart sound signals. The performance of different machine learning algorithms and its comparative analysis in cardiovascular disease is given by Asif et al. [22,23].

III. Materials and methods

This methodology outlines the necessary steps for building a CNN-based heart disease prediction model. However, it's essential to note that the success of the model depends on the quality of the dataset, the CNN architecture, and the hyper parameter tuning [24,25]. Therefore, it's crucial to have expert knowledge and experience in building and training CNNs for accurate cardiovascular disease prediction. Our goal in conducting this study is to devise a blend of deep learning algorithms that can accurately identify the presence of heart ailments at an early stage [26]. Our proposed method has three primary contributions: First, it uses a deep learning network for early detection. Second, we compare the effectiveness of our approach. Third, we implement our methodology in real-time application. Our proposed approach will be evaluated based on several metrics, including accuracy, precision, and recall, to measure its effectiveness in detecting heart diseases at an early stage. The architecture of the proposed methodology is illustrated in Figure 1.

A. *Data Preprocessing*

The paper suggests a plan to detect heart disease early by utilizing a CNN. The aim of this research is to employ neural networks to calculate the likelihood of a patient having heart disease by considering different factors like blood pressure, heart rate, and cholesterol levels [27]. The dataset used for this project are Cleveland dataset [31], Corona Heart Disease (CHD) dataset and HD dataset. Though 76 attributes are in the Cleveland dataset 14 attributes are utilized in the

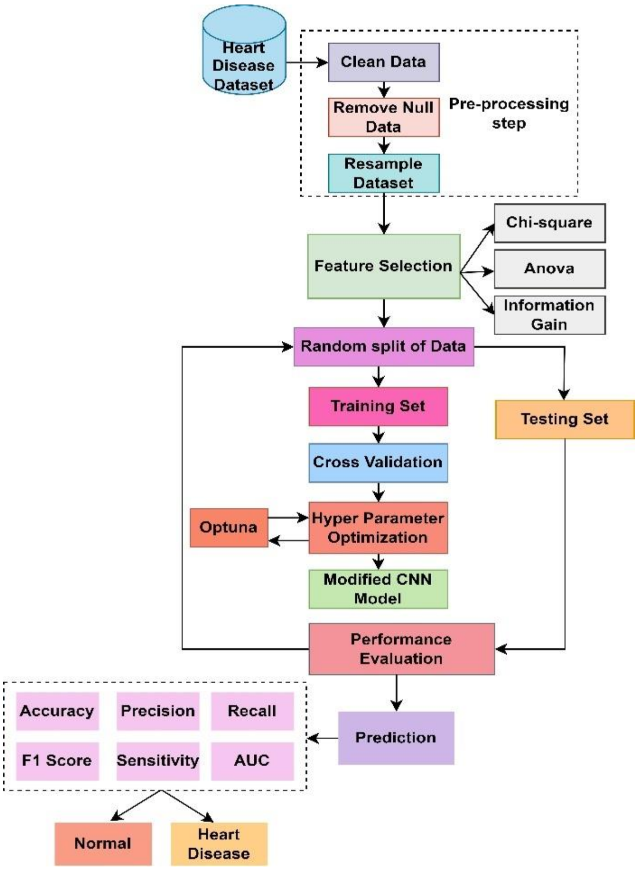

Fig. 1 Proposed methodology for heart-disease prediction.

proposed work. The CHD dataset contains information on 16 different attributes including age, sex, heart rate, etc. The dataset was comprised with a collection of data parameters of 4240 patients from Framingham Heart Institute out of which 644 were heart patients and 3597 were normal. The HD dataset consist of 11 attributes of 1190 patient records. The structure of the dataset description is depicted in Table I, II and III. The features in the dataset effectively offer more prominent information related to the prediction of cardiovascular disease. Age, Sex, Type of chest pain, blood pressure level and cholesterol levels are some of the primitive features which helps predominant in the prediction of heart disease.

TABLE I
DATASET DESCRIPTION OF CLEVELAND DATASET

Attribute	Description
Sex	Male=1; Female=0
Age	29 to 79
Chest Pain	1-4
Resting blood pressure	Continuous
Cholestrol	Continuous
Fasting blood sugar	Yes=1, No=0
ECG	Yes=1, No=0
Max Heart Rate	Continuous
Exercise induced angina	Yes=1, No=0
Old peak	0-4
Slope of peak exercise	1-3
Ca	0-3
Thal	3,6,7
CHD	Yes=1, No=0

TABLE II
DATASET DESCRIPTION OF CHD DATASET

Attribute	Description
Sex	Male=1; Female=0
Age	29 to 79
Education	0-4
CurrentSmoker	Yes=1, No=0
CigsPerDay	1 to 25
BPMeds	Takes, Yes=1, No=0
PrevalentStroke	Yes=1, No=0
PrevalentHyp	Yes=1, No=0
Diabetes	Yes=1, No=0
TotChol	Continuous
SysBP	Continuous
DiaBP	Continuous
BMI	Continuous
HeartRate	Continuous
Glucose	Continuous
TenYearCHD	Yes=1, No=0

TABLE III
DATASET DESCRIPTION OF HD DATASET

Attribute	Description
Sex	Male=1; Female=0
Age	29 to 79
Chest Pain	1-4
Resting blood pressure	Continuous
Cholestrol	Continuous
Fasting blood sugar	Yes=1, No=0
ECG	Yes=1, No=0
Max Heart Rate	Continuous
Exercise induced angina	Yes=1, No=0
Old peak	0-4
Slope of peak exercise	1-3
CHD	Yes=1, No=0

B. Data Preprocessing and Splitting

Preprocessing, which entails cleaning and preparing data before it is fed into a model for training, is an essential step in ensuring optimal performance and accuracy in prediction system. For better preprocessing the missing values should be handled effectively. Outliers are data points that differ significantly from the rest of the data and are often the consequence of measurement errors, data collection errors, or human error. In order to preserve the integrity of the data, any outliers found in this study are dealt with by being removed [28]. Data type conversion is an essential preprocessing procedure that involves transforming data from one data type to another. This stage is required since algorithms might not always be able to process the input data in its original format. This procedure entails scaling numerical data to normalize each feature, translating numerical data to categorical data, or converting text data to numerical data [29]. To process input data effectively and reliably proper data type conversion is essential for accurate predictions.

Typically, to develop a model, the dataset is typically split into two groups: the training set and the testing set. The training set is used to teach the model how to recognize patterns and features within the data. The testing set is then employed to determine the model's effectiveness in accurately predicting new data that it hasn't been exposed to during the training process. The splitting of the dataset is important to avoid overfitting, which occurs when the model becomes too specialized in fitting the training data, and cannot generalize well to new, unseen data. To obtain a more accurate estimation of the ability of a model to accurately predict outcomes on data it has not encountered before is referred to as its performance on new data. This is typically evaluated using a testing dataset that is separate from the training data used to develop the model, it is customary evaluate a model on separate dataset. The training dataset is usually larger than the testing dataset, with a common split being 80% for training and 20% for testing. The count in the splitting of data in the three datasets for the proposed work is shown in Table IV. However, the actual ratio of the split can depend on the size and complexity of the dataset and the specific needs of the problem being addressed.

TABLE IV
DATASET PARTITIONING

Dataset	Training	Testing
Cleveland	242	61
CHD	3392	848
HD	952	238

C. Feature Selection

In order to increase model accuracy and interpretability, feature selection in heart disease prediction entails choosing the most pertinent variables from a pool of potential predictors. This procedure is essential for creating prediction models that are both successful and economical to compute [30]. Demographic data, medical history, lifestyle factors, and physiological measurements like blood pressure, cholesterol, and ECG readings are features that are frequently taken into account when predicting heart disease. Statistical techniques like as ANOVA and chi-square, machine learning algorithms like LASSO regression and recursive feature elimination (RFE), and domain knowledge-driven methods can all be used for feature selection. Models can better capture the underlying patterns and relationships in the data by choosing the most useful features, which improves the accuracy of heart disease risk forecasts. In the proposed method, feature selection based on chi-square, ANOVA and Information gain is implementation for attaining better prediction results.

Feature selection, which makes use of statistical techniques like chi-square, ANOVA (Analysis of Variance), and information gain, is essential to the precise prediction of cardiac disease. By assessing the degree of independence between categorical factors and the target variable, the chi-square test can be used to pinpoint characteristics that are strongly linked to heart disease. ANOVA, on the other hand, determines which of the target's categories have significant differences by evaluating the variation in numerical properties among them. Information gain, on the other hand, highlights the most instructive characteristics for heart disease prediction by measuring the decrease in uncertainty about the target variable by including a certain trait. By combining these techniques, it is possible to minimize computing complexity and dimensionality while optimizing the predictive performance of heart disease models by selecting the most pertinent features.

D. Modified CNN Architecture

The architecture of the modified CNN algorithm used in the proposed method is given in figure 2. A modified version of the standard CNN architecture is referred to as a modified convolutional neural network (CNN). Due to their prowess at accurately capturing spatial relationships in data, CNNs are frequently employed for tasks like image categorization, object detection, and image recognition. The model used for the proposed method is comprised with four convolution layer, four fully connected layer and a softmax layer. The network depth is increased and more complex features are captured by adding more convolutional layers. The kernel size of convolutional layer can have an impact on the network's receptive field. While larger kernels can concentrate on more general aspects, smaller kernels are better at capturing minute details.

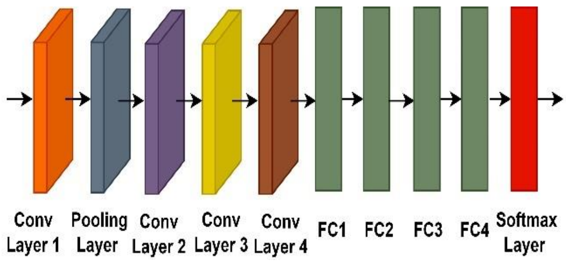

Fig. 2 Modified CNN Algorithm

The one-dimensional input data is first passed to convl layer1 with 64 filters with a kernel size of 3 x 1. The convolution layer is followed by a pooling layer of window size 2 with stride 2. Then the output is passed to conv layer2, conv layer3 and conv layer4 with filter size of 128, 512 and 1024 respectively. Rectified Linear Unit (ReLU) is applied for linear activation function. The features maps after passing through the convolutional layer, it is directed to four fully connected layers with 512, 256, 128 and 64 neurons respectively. Finally, the softmax layer gives the classification output.

E. Optimization using optuna Algorithm

The performance of CNN algorithm can be increased by learning model that can be employed with the hyperparameter optimisation framework. Optuna is one among the important optimization algorithm that aids in automating the process of selecting the ideal collection of hyperparameters for a CNN model's training. It uses methods like Bayesian optimization to find the ideal hyperparameters in an effective, automated manner. With capabilities like pruning, early stopping, and multi-objective optimization, Optuna can handle both straightforward and intricate optimization issues. The framework's scalability and user-friendly interface make it a popular option for effectively optimizing machine learning models. The schematic diagram for hyper optimization modelling is shown in Figure 3. The typical process for using Optuna to improve a CNN model is by deciding which hyperparameters is needed to optimise by specifying the search spaces for those parameters. The rate of learning, the number of layers, the number of filters per layer, the dropout rate, etc. are a few examples of things you would want to optimise. With Optuna, we can define search spaces in a variety of ways, including by selecting discrete values, setting ranges, or employing distributions [31].

Optuna's objective function is a crucial element. The hyperparameters are entered, used to build and train the CNN model, and then a scalar value representing the model's performance is returned. The performance of the model might be assessed by the objective function using cross-validation or train-test splits. Set up the optimisation research as follows: Make an Optuna study object and set the desired optimisation parameters for it. In doing so, it is necessary to indicate the optimisation algorithm (such as Bayesian optimisation or random search), the number of trials or time allotted, and the desired optimisation direction (such as minimise loss or maximise accuracy). Repeat as many times as necessary, up to the time allotted has been used. Optuna selects a collection of hyperparameters from the specified search space for each trial, passes those selections forward to the objective function, and then logs the evaluation metric. This data is then used by Optuna's optimisation algorithms to decide which set of hyperparameters to test next.

Pseudo code for Optuna

Start Optuna.
1. Input the pre-processed data.
2. Determine the initial optuna parameters
    Population size (P)
    Number of iterations (N).
3. Initialization of hyperparameters and calculate the search ranges
4. For t = 1:N
5. Generate the grid search and random search.
6. For I = 1:P
7. Phase 1: Build and manipulate hyperparameter search (sampling phase).
8. For j = 1:m
9. Calculate new status of the jth dimension
10. End.
11. Update the ith population
12. Phase 2: Optimize the objective function (pruning phase).
13. For j = 1:m.
14. Calculate new status of the jth dimension
15. End.
16. Update the ith population
17. End.
18. Update best optimized solution.
19.End.
20.Output best candidate solution obtained by Optuna.
End Optuna.

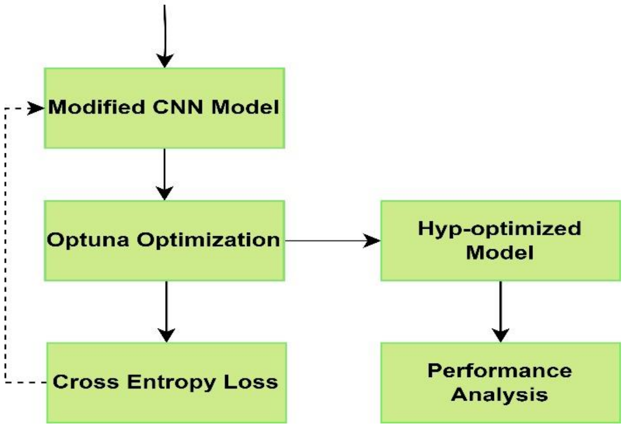

Fig. 3 Hyper-parameter Optimization

The top set of hyperparameters is discovered during the search when the optimisation is finished. Optuna offers a way to access the optimal trial and its related hyperparameters. Retrain the CNN model on the complete training dataset using the optimal hyperparameters discovered during optimisation. This time, the convergence is improved by using more epochs or other methods. Finally, the optimum hyperparameters is used to assess the trained CNN model's performance on a different validation or test dataset. As a result, a fair assessment of the model's performance is received.

IV. Results and Discussion

Experiments were conducted for the evaluation of the proposed model using windows 10, i3 processor, speed 2 ghz, 4GB RAM, system in python environment with pycharm. The performance measures are obtained using the following metrics given in equation 1 to 4 and obtained values are depicted in Table 5.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (4$$

TABLE V
PERFORMANCE METRICS

Metrics	Dataset		
	Cleveland	CHD	HD
Accuracy Score	**97.0**	**97.2**	**96.2**
Precision Score	**98.6**	**99.4**	**94.3**
Recall Score	**97.1**	**94.5**	**89.6**
F1 Score	**97.8**	**96.8**	**91.8**
Sensitivity	**86.5**	**95.4**	**85.4**
Specificity	**89.5**	**89.84**	**88.4**
Area Under the Curve	**97.6**	**98.7**	**95.4**

TABLE VI
MODIFIED CNN HYPERPARAMETERS

Parameter	Search Space	Tuned Value
Weight decay	[0.0001,0.001]	0.0005
Dropout	[0,0.5]	0.25
Pool Method	'max', 'average',	'max'
Kernel Initializer	'glorot', 'he', 'normal', 'uniform'	'glorot'
Optimizer	'SGD', 'Adam'	'Adam
Learning Rate	[0.001,0.1]	0.05
Learning Schedule	'constant', 'step decay', 'exponential decay'	'step decay'
Epochs	[30,80]	50

Table VI depicts the hyperparameter value optimized using the Optuna algorithm. The results of using demographic attributes for heart disease prediction using deep CNNs can vary depending on the dataset, model architecture, and specific attributes being used. However, in general, including demographic attributes in heart disease prediction models can lead to improved accuracy compared to models that only use medical data. For the approach presented in this paper, a deep CNN is developed to predict heart disease based on age, sex, and clinical variables such as cholesterol levels and blood pressure. The model achieved an AUC-ROC of 0.987, outperforming traditional machine learning algorithms. The area under the curve for the proposed method is shown in Figure 4. The top ten attributes ranked by the feature selection method are given in table 7. The feature selection method utilized in the method are chi-square, anova and information gain. By determining which features are most pertinent, feature selection methods such as Chi-square, ANOVA, and

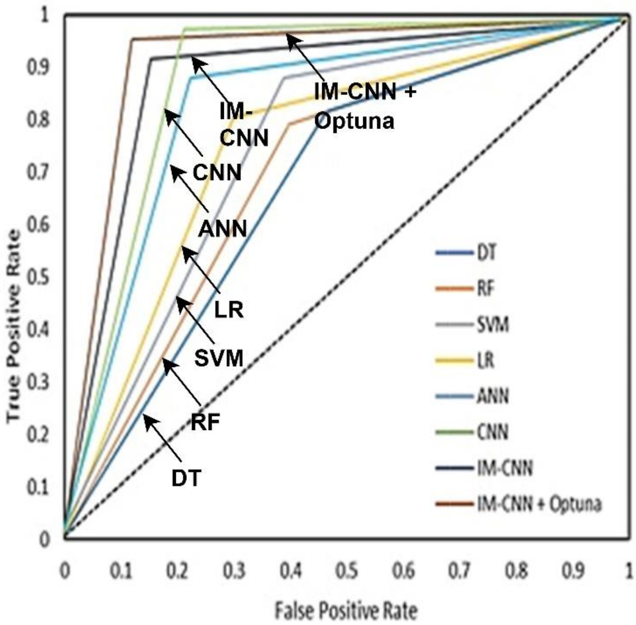

Fig. 4 AUC for proposed method of CHD dataset

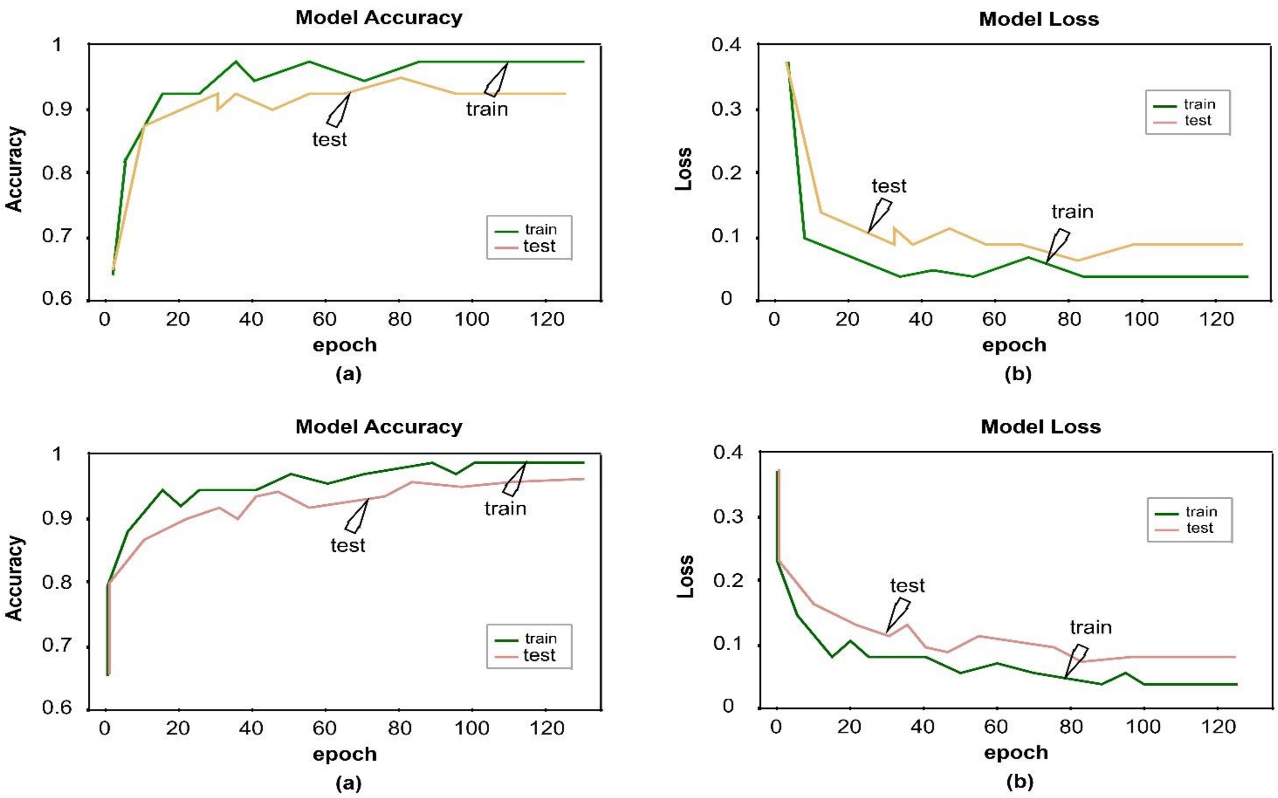

Fig. 5 Model Accuracy and loss for a) & b) Cleveland c) & d) CHD dataset.

TABLE VII
TOP TEN ATTRIBUTES RANKED BY FEATURE SELECTION METHODS ON THREE DATASETS

Method	Metrics	CLEVELAND Dataset								
Chi-Square	***Feature***	Thal	cp	chol	ca	rbp	fbs	ecg	mhr	Old peak
	Score	165.6	108.6	72.8	62.59	31.5	28.9	22.6	11.9	9.6
ANOVA	***Feature***	Thal	rbp	ca	chol	cp	Old peak	eia	fbs	sex
	Score	145.3	89.3	76.2	61.6	35.4	30.5	12.8	9.5	6.7
Information Gain	***Feature***	Thal	chol	cp	ca	fbs	ecg	rbp	mhr	eia
	Score	98.5	82.5	65.4	45.7	39.6	22.6	15.7	8.6	4.5
		CRD Dataset								
Chi-Square	***Feature***	Totchol	Age	P.stroke	Diabetes	Heartrate	Glucose	SysBP	DiaBP	sex
	Score	163.6	143.7	123.8	97.5	67.4	43.2	24.9	12.8	5.6
ANOVA	***Feature***	Diabetes	Age	Heartrate	SysBP	P.stroke	DiaBP	Glucose	sex	BMI
	Score	234.6	157.8	92.6	83.5	45.8	23.6	19.6	8.7	4.3
Information Gain	***Feature***	Totchol	Age	SysBP	Glucose	Diabetes	sex	P.stroke	DiaBP	BMI
	Score	123.6	89.5	65.8	45.7	34.7	19.6	11.8	7.9	3.6
		HD Dataset								
Chi-Square	***Feature***	eia	age	rbp	cp	ecg	mhr	fbs	Old peak	sex
	Score	96.7	88.5	73.7	68.4	53.7	35.6	23.8	16.3	8.9
ANOVA	***Feature***	rbp	age	chol	eia	mhr	ecg	cp	Old peak	fbs
	Score	408.5	320.6	289.6	224.8	127.6	95.7	56.9	32.9	18.6
Information Gain	***Feature***	rbp	cp	age	chol	mhr	eia	ecg	sex	fbs
	Score	134.6	112.5	98.5	75.3	56.7	45.6	27.2	12.3	8.7

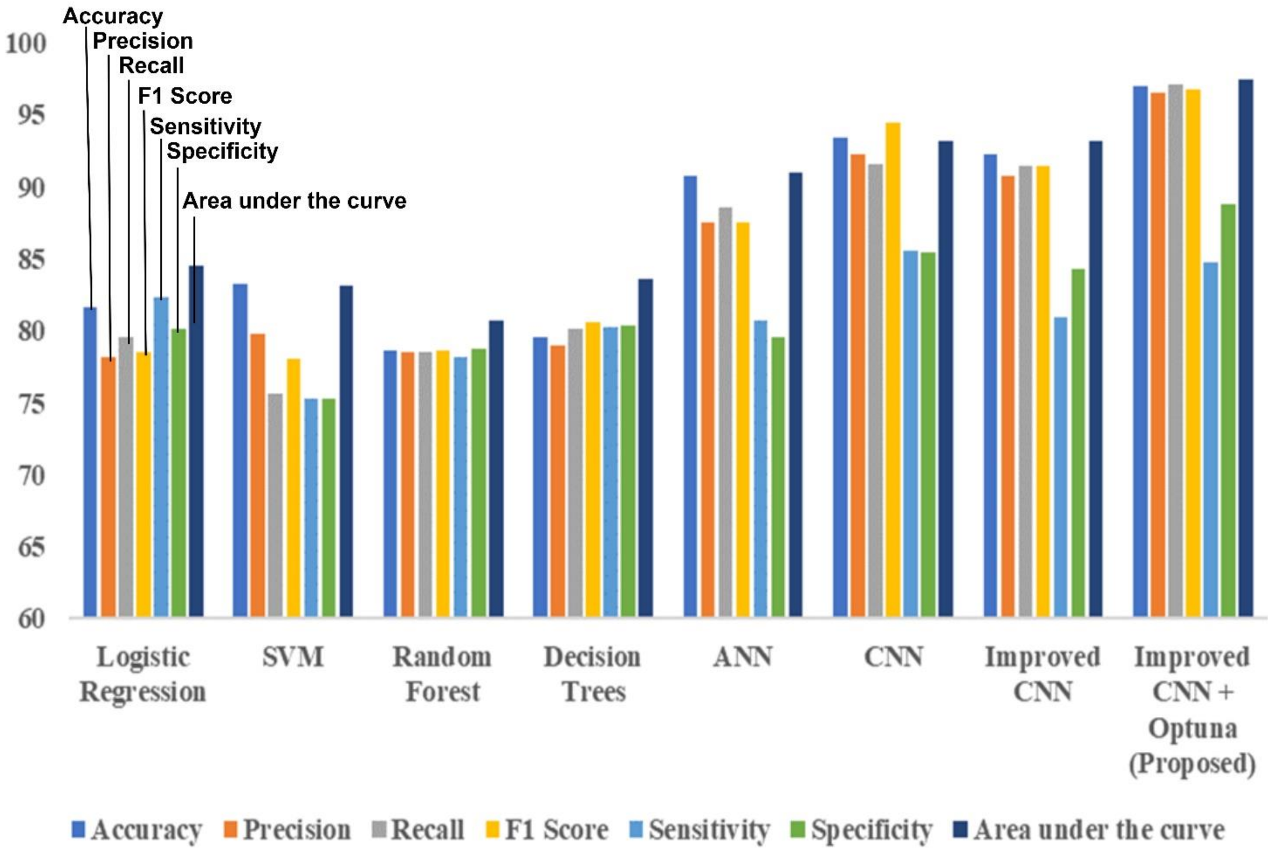

Fig. 6 Graphical Representation of performance evaluation of various methods (Cleveland dataset)

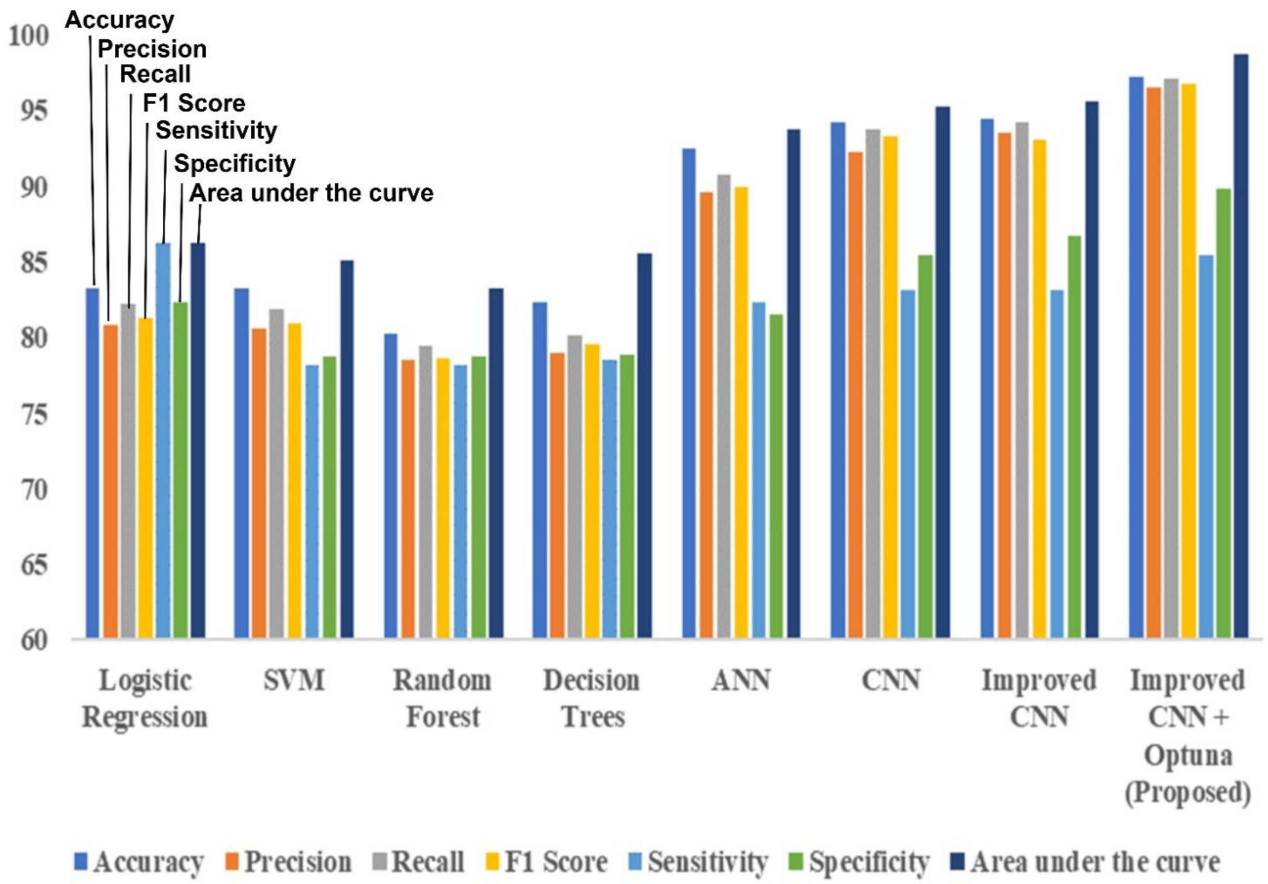

Fig. 7 Graphical Representation of performance evaluation of various methods (CHD Dataset)

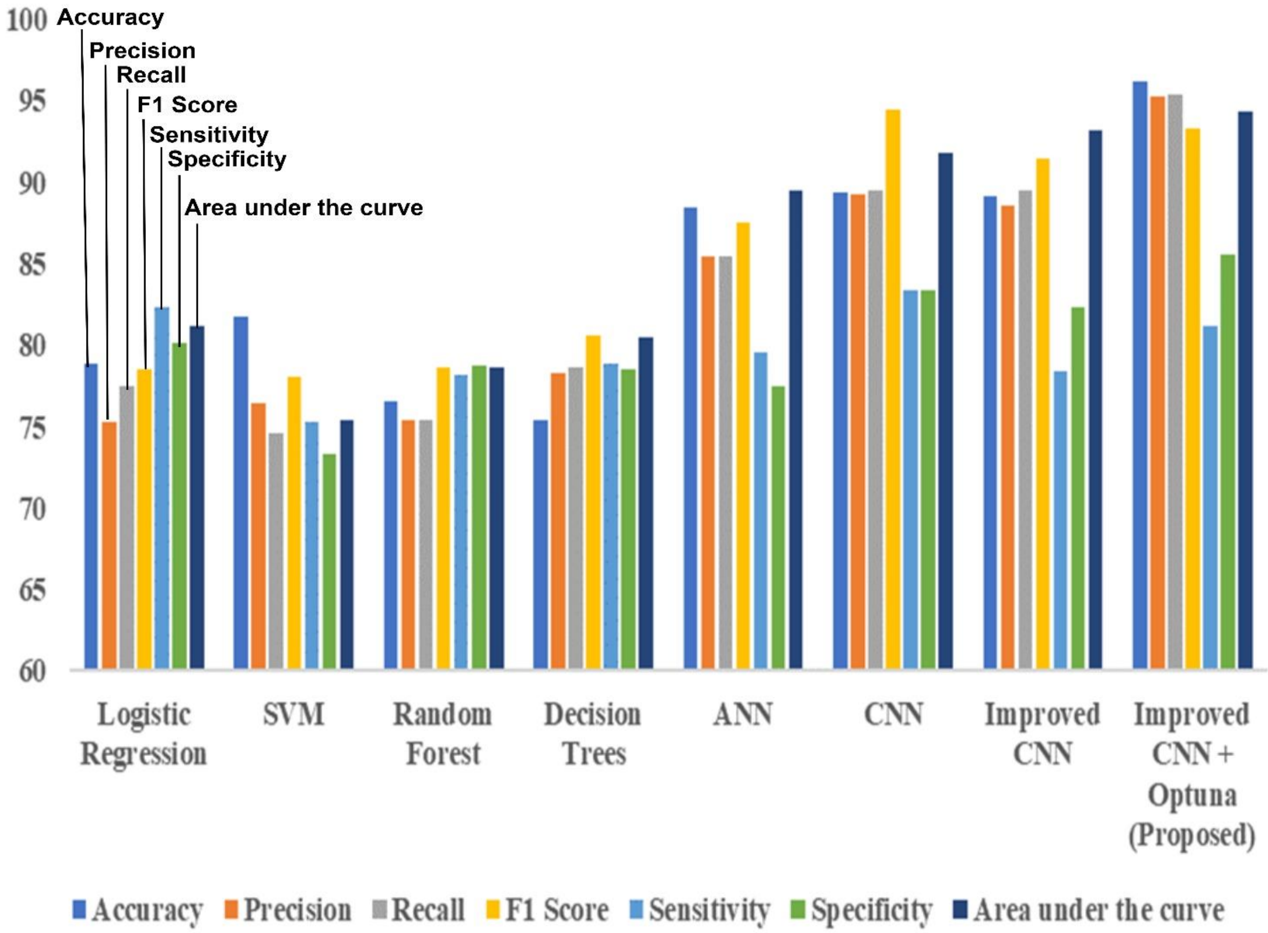

Fig. 8 Graphical Representation of performance evaluation of various methods (HD Dataset)

TABLE VIII
PERFORMANCE ANALYSIS BY COMPARING FEATURE SELECTION METHOD IN CLEVELAND DATASET

Metrics	Without Feature Selection					With feature Selection				
	SVM	ANN	CNN	Improved CNN	Improved CNN + Optuna	SVM	ANN	CNN	Improved CNN	Improved CNN + Optuna
Accuracy	80.3	89.6	90.7	90.3	**94.5**	83.2	92.5	94.2	94.5	**97.0**
Precision	78.5	87.4	89.6	90.5	**93.7**	80.6	89.6	92.3	93.5	**98.6**
Recall	79.4	88.5	89.9	91.6	**95.7**	81.9	90.8	93.8	94.2	**97.1**
F1 Score	78.6	88.5	90.4	89.7	**94.6**	80.9	89.9	93.3	93.1	**97.8**
Sensitivity	78.2	82.3	81.2	81.2	**83.2**	78.2	82.3	83.1	83.1	**86.5**
Specificity	78.7	80.4	83.8	84.1	**84.6**	78.7	81.5	85.4	86.7	**89.5**
Area under the curve	83.3	91.8	92.5	85.4	**92.5**	85.1	93.8	95.2	89.8	**97.6**

TABLE IX
COMPARISON OF PROPOSED METHOD WITH STATE OF THE ART METHODS

Authors	Method	Year	Accuracy	Sensitivity	Specificity	F1-Score
[3]	(FAMD)+RF	2019	93.44	89.28	96.96	-
[15]	HRFLM	2019	88.4	92.8	82.6	90
[7]	WAE	2020	93	91	-	93
[23]	FCMIM+SVM	2020	92.37	89	98	-
[24]	NFR+LR	2020	92.53	-	-	-
[25]	CART	2021	88.33	84.62	-	-
[30]	GA-LDA	2021	93.65	96	-	-
[29]	SMOTE-ENN	2022	90	97.3	-	92.3
[26]	XGBoost	2022	91.8	85.71	96.96	90.56
[27]	MLP-PSO	2022	84.61	88.3	-	84.4
[28]	CART	2023	87.25	84.51	89.74	-
Our Proposed method	**Improved CNN + optuna**		**97.2**	**95.4**	**89.8**	**96.8**

Information Gain are crucial for enhancing model performance. The attributes such as thal, cholesterol, blood pressure, age, heart rate are some of the top attributes which are gained top rank which is proved by the feature selection method. This information can be used to fine-tune the model, adjust the decision threshold, or explore other techniques to improve the performance of the model.

Table 7 summarizes the top ten attributes selected by feature selection algorithms for the three datasets. Feature selection is a crucial step in the process of building machine learning models, especially for medical applications like early heart disease detection. It helps in identifying the most relevant features that contribute to the prediction, thereby improving the model's performance and interpretability. The three popular feature selection methods: Chi-square, ANOVA (Analysis of Variance), and Information Gain are analysed in the proposed work. In practice, the choice of feature selection method can depend on the nature of the features (categorical or continuous) and the specific requirements of the analysis. For early heart disease detection, combining these methods can provide a robust set of relevant features, leading to better model performance and more accurate predictions.

The comparative analysis of the proposed method with some of the machine learning and deep learning methods for Cleveland, CHD and HD datasets are summarized in Figure 6, 7 and 8. The improvement in the performance of the model with the utilization of feature selection method is analysed in Table VIII. The results obtained by feature selection method for various classifier is furnished in the table. The comparison of the proposed method with state-of-the-art method is shown in Table IX. It is evident from the table that deep learning methods produce more results than machine learning methods in the prediction of heart disease using the data attributes. It is also proven that the proposed method produces greater results in different performance metrics than other methods. Also, it has been evident from Figure 6 that the performance of the proposed method is good compared to the various methods evaluated. There is a matching bar for every metric, and the score is displayed on the y-axis. For every metric, the bars show how well the system is performing. For all metrics, the proposed method yields a very good result, and it is perceptible from figure 8.

The use of demographic attributes for heart disease prediction using deep CNNs has several potential benefits and limitations, as discussed earlier. One potential benefit is that demographic attributes are easily obtainable and non-invasive features that can be used to supplement medical data. Additionally, the statement conveys that certain characteristics or qualities can be utilized to recognize people who are more likely to develop heart disease. By identifying such individuals early on, it becomes possible to take preventive measures and minimize the risk of heart disease.

V. Conclusion

In conclusion, heart disease prediction using deep CNNs can be improved by including demographic attributes such as age, sex, and other relevant clinical variables as features in the model. The use of these attributes can provide additional information about the patient and help improve the accuracy of the model. However, it is important to carefully consider the attributes being used, the potential biases and limitations of these attributes, and how they can be combined with other relevant clinical data to improve accuracy. In the proposed work, the model created using Optuna CNN yields a promising result, which is very well observed by the accuracy obtained. Overall, the use of demographic attributes in heart disease prediction and the utilization of deep learning models have great potential for improving patient outcomes by enabling early detection and prevention of heart disease.

References

[1] American Heart Association, Heart Failure, American Heart Association, Chicago, IL, USA, 2020, https://www.heart.org/en/health-topics/heart-failure.

[2] Drożdż, K.; Nabrdalik, K.; Kwiendacz, H.; Hendel, M.; Olejarz, A.; Tomasik, A.; Bartman, W.; Nalepa, J.; Gumprecht, J.; Lip, G.Y.H. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach. Cardiovasc. Diabetol. 2022, 21, 240.

[3] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, ''MIFH: A machine intelligence framework for heart disease diagnosis,'' IEEE Access, vol. 8, pp. 14659–14674, 2020.

[4] S.Shalev-Shwartz and S. Ben-David, "Understanding machine learning," From Theory to Algorithms, Cambridge University Press, Cambridge, UK, 2020.

[5] S.Marsland, "Machine learning," An Algorithmic Perspective, CRC Press, Boca Raton, FL, USA, 2020.

[6] E.Keogh and A. Mueen, "Curse of dimensionality," in Encyclopedia of Machine Learning and Data Mining, C. Sammut and G. I. Webb, Eds., Springer, Cham, Switzerland, 2017.

[7] I. D. Mienye, Y. Sun, and Z. Wang, ''Improved sparse autoencoder based artificial neural network approach for prediction of heart disease,'' Informat. Med. Unlocked, vol. 18, Jan. 2020, Art. no. 100307. Esteva, A. et al. A guide to deep learning in healthcare. Nat. Med. 25, 24–29 (2019).

[8] Liu, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digital Health 1, e271–e297 (2019).

[9] Yazdani, A. Malik and W. Ahmad, "A novel approach for heart disease prediction using strength scores with significant predictors", BMC Med Inform Decis Mak, 2021.

[10] S. Sharma and M. Parmar, "Heart Disease Prediction using Deep Learning Neural Network Model", International Journal of Innovative Technology and Exploring Engineering, 2020.

[11] R. S. Singh, B. S. Saini, and R. K. Sunkaria, "Detection of coronary artery disease by reduced features and extreme learning machine," Medicine and Pharmacy Reports, vol. 91, no. 2, pp. 166–175, 2018.

[12] J. K. Kim and S. Kang, ''Neural network-based coronary heart disease risk prediction using feature correlation analysis,'' J. Healthcare Eng., vol. 2017, Sep. 2017, Art. no. 2780501, doi: 10.1155/2017/2780501.

[13] Z. Du, Y. Yang, J. Zheng, Q. Li, D. Lin, Y. Li, J. Fan, W. Cheng, X.-H. Chen, and Y. Cai, ''Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: Model development and performance evaluation,'' JMIR Med. Informat., vol. 8, no. 7, Jul. 2020, Art. no. e17257, doi: 10.2196/17257.

[14] M. Senthilkumar, C. Thirumalia and G. Srivastava, "Effective Heart Disease Prediction using Hybrid Machine Learning Techniques", IEEE Access.

[15] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande and Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", Hindawi Computational Intelligence and Neuroscience, 2021.

[16] Ebenezer Owusu, Prince Boakye-Sekyerehene, Justice Kwame Appati, and Julius Yaw Ludu, 'Computer-Aided Diagnostics of Heart Disease Risk Prediction Using Boosting Support Vector Machine', Comput Intell Neurosci. 2021; 2021: 3152618. doi: 10.1155/2021/3152618.

[17] Shin H. C., Roth H. R., Gao M., et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging 2016;35(5):1285–1298. doi: 10.1109/tmi.2016.2528162.

[18] Nandy S., Adhikari M., Balasubramanian V., Menon V. G., Li X., Zakarya M. An intelligent heart disease prediction system based on swarm-artificial neural network. Neural Computing and Applications. 2021:1–15. doi: 10.1007/s00521-021-06124-1.

[19] Wafaa N. Al-Sharu, Ali Mohammad Alqudah, Shoroq Qazan, and Amin Alqudah, "Detection of Valvular Heart Diseases Using Fourier Transform and Simple CNN Model," IAENG International Journal of Computer Science, vol. 49, no.4, pp985-993, 2022.

[20] Shumpei Takezaki, and Kazuya Kishida, "Data Augmentation and the Improvement of the Performance of Convolutional Neural Networks for Heart Sound Classification," IAENG International Journal of Computer Science, vol. 49, no.4, pp1033-1040, 2022.

[21] Md. Asfi-Ar-Raihan Asif, Mirza Muntasir Nishat, Fahim Faisal, Rezuanur Rahman Dip, Mahmudul Hasan Udoy, Md. Fahim Shikder, and Ragib Ahsan, "Performance Evaluation and Comparative Analysis of Different Machine Learning Algorithms in Predicting Cardiovascular Disease," Engineering Letters, vol. 29, no.2, pp731-741, 2021.

[22] M. M. Arifin, M. A. Based, K. M. Mumenin, A. Imran, M. A. Azim, Z. Alom, and M. A. Awal, ''OLGBM: Optuna optimized light gradient boosting machine for intrusion detection,'' in Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (IC4ME2), Rajshahi, Bangladesh, Dec. 2021, pp. 1–4.

[23] 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 ehealthcare,'' IEEE Access, vol. 8, pp. 107562–107582, 2020.

[24] S. J. Pasha and E. S. Mohamed, ''Novel feature reduction (NFR) model with machine learning and data mining algorithms for effective disease risk prediction,'' IEEE Access, vol. 8, pp. 184087–184108, 2020.

[25] E.Miranda, M. Aryuni, C. Bernando, and A. Hartanto, ''Application for early heart disease prediction based on data mining approach,'' in Proc. 4th Int. Conf. Comput. Informat. Eng. (IC2IE), Sep. 2021, pp. 375–378.

[26] K. Budholiya, S. K. Shrivastava, and V. Sharma, ''An optimized XGBoost based diagnostic system for effective prediction of heart disease,'' J. King Saud Univ., Comput. Inf. Sci., vol. 34, no. 7, pp. 4514–4523, Jul. 2022.

[27] A. Al Bataineh and S. Manacek, ''MLP-PSO hybrid algorithm for heart disease prediction,'' J. Pers. Med., vol. 12, no. 8, p. 1208, Jul. 2022.

[28] M. Ozcan and S. Peker, ''A classification and regression tree algorithm for heart disease modeling and prediction,'' Healthcare Anal., vol. 3, Nov. 2023, Art. no. 100130.

[29] M. M. Nishat, F. Faisal, I. J. Ratul, A. Al-Monsur, A. M. Ar-Rafi, S. M. Nasrullah, M. T. Reza, and M. R. H. Khan, ''A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset,'' Sci. Program., vol. 2022, pp. 1–17, Mar. 2022.

[30] V. Jothi Prakash and N. K. Karthikeyan, ''Enhanced evolutionary featureselection and ensemble method for cardiovascular disease prediction,'' Interdiscipl. Sci., Comput. Life Sci., vol. 13, no. 3, pp. 389–412, Sep. 2021.

[31] P. Srinivas and R. Katarya, ''HyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost,'' Biomed. Signal Process. Control, vol. 73, Mar. 2022, Art. no. 103456, doi: 10.1016/j.bspc.2021.103456.

[32] D. Dua and C. Graff. (2017). UCI Machine Learning Repository. [Online].Available:http://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[33] UCI Machine Learning Repository. (2018). Heart Disease Dataset (Comprehensive). [Online]. Available: https://archive.ics.uci.edu/ml/ datasets/Heart+Disease.

[34] Kaggle. (2019). Cardiovascular Disease Dataset. [Online]. Available:https://www.kaggle.com/sulianova/cardiovascular-disease-dataset.