Explainable Artificial Intelligence Method for Identifying Cardiovascular Disease with a Combination CNN-XG-Boost Framework

J Chandra Sekhar[1], T L Deepika Roy[2], Dr. K. Sridharan[3],
Dr. Natrayan L[4], Dr.K.Aanandha Saravanan[5], Ahmed I. Taloba[6]

Professor in CSE, NRI Institute of Technology, Guntur, India[1]
Assistant Professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh, India[2]
Department of IT, Panimalar Engineering College, Chennai, India[3]
Department of Mechanical Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India[4]
Associate Professor, Department of ECE, VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India[5]
Department of Computer Science, College of Computer and Information Sciences, Jouf University, Saudi Arabia[6]
Information System Department, Faculty of Computers and Information, Assiut University, Assiut, Egypt[6]

Abstract*—Cardiovascular disease (CVD) is a globally significant health issue that presents with a multitude of risk factors and complex physiology, making early detection, avoidance, and effective management a challenge. Early detection is essential for effective treatment of CVD, and typical approaches involve an integrated strategy that includes lifestyle modifications like exercise and diet, medications to control risk factors like high blood pressure and cholesterol, interventions like angioplasties or bypass surgery in extreme cases, and ongoing surveillance to prevent complications and promote heart function. Traditional approaches often rely on manual interpretation, which is time-consuming and prone to error. In this paper, proposed study uses an automated detection method using machine learning. The CNN and XGBoost algorithms' greatest characteristics are combined in the hybrid technique. CNN is excellent in identifying pertinent features from medical images, while XGBoost performs well with tabular data. By including these strategies, the model's robustness and precision in predicting CVD are both increased. Furthermore, data normalization techniques are employed to confirm the accuracy and consistency of the model's projections. By standardizing the input data, the normalization procedure lowers variability and increases the model's ability to extrapolate across instances. This work explores a novel approach to CVD detection using a CNN/XGBoost hybrid model. The hybrid CNN-XGBoost and explainable AI system has undergone extensive testing and validation, and its performance in accurately detecting CVD is encouraging. Due to its ease of use and effectiveness, this technique may be applied in clinical settings, potentially assisting medical professionals in the prompt assessment and care of patients with cardiovascular disease.

Keywords*—*Cardiovascular disease; CNN; XGBoost; traditional approaches; explainable AI

I. Introduction

In 2019, CVD will be responsible for 32% of all fatalities worldwide [1]. If CVD is identified early on, the untimely deaths caused by it can be avoided. Ongoing surveillance of a person's heart health and functioning can aid in early identification of CVD. The circulatory system's primary job is to circulate clean blood throughout the body via electrical waves produced within the organ. When a blood clot blocks the flow of blood to a portion of the heart, it can cause a heart attack. The portion of the cardiac muscles fed by the arteries starts to perish if this clot totally stops the blood supply. The majority of people recover from their initial cardiac events and go on to lead regular lives, engaging in constructive activities for a long time afterwards. However, suffering from a cardiac event does force you to adjust. Coronary arteries disease symptoms include discomfort in the chest, chest pressure, tightness in the chest, and pain in the chest (angina), Breathlessness Returning upper abdomen, throat, jaw, or neck pain discomfort, tingling, numbness, or coolness in the arms or legs if the coronary arteries there are constricted. Depending on how severely the coronary artery was injured and what level of heart disease precipitated the heart attack, the doctor will recommend different drugs and lifestyle modifications [2]. The word "cardiomyopathies" refers to conditions affecting the cardiac muscles. Usually, people just refer to them as enlarged hearts. Hearts that are abnormally large, thick, or rigid are found in people with various diseases. Their hearts' ability to circulate blood is compromised. When left untreated, cardiomyopathies worsen. Heart problems and erratic heartbeats may result from them. Although it may additionally be brought on by infections, metabolic disorders, diabetes, obesity, hypertension, and other factors, cardiomyopathy may occur in family. An issue with one or more blood artery or heart components is known as congenital heart disease. It affects roughly eight out of one thousand kids. Some individuals with it may exhibit symptoms from birth, but others may not show signs until later in adolescence or early adulthood.

Most of the time, there is no idea why it occurs. Genetics might be involved, or it could occur if a new-born is given drugs, alcohol, or viral illnesses prior to birth. Heart failure indicates that the heart isn't pumping as hard as it ought to. The body will keep both salt and water as a result, this will make you swollen and breathless [3]. Over 6.7 million Americans suffer with heart failure, making it a serious health issue. It is the main reason why adults over 65 ends up in hospitals. The American Heart Association (AHA) predicts that by 2030, 8.5 million Americans will have been given the diagnosis of coronary artery disease [4]. Regretfully, coronary artery disease cannot be cured, and after it has been identified, it cannot be reversed. However, people can alter the way of living to lower chance of experiencing more health issues, like as heart attack. Wearable technology such as electronic watches, Fit-bits, collar traps, and others are being utilised to continuously track the health of the heart. Such devices don't require hospital-grade medical equipment to read an individual's ECG readings in real time [5]. The precise diagnosis of CVD requires analysis of the ECG data captured using these gadgets. Wearable technology is battery-operated and has processing power limitations when compared to medical technology. As a result, wearable technology is unable to identify the kind of CAD with greater precision. Thus, wearable technology is limited to the monitoring of cardiac function and the assessment of abnormality frequency. They serve as alert systems, and anyone interested in learning more about a particular kind of arrhythmia or heart problem should speak with a physician. Unfortunately, a cardiologist's workload increases along with the number of individuals with heart rhythm disorders. For this reason, it is essential to detect abnormalities using machine learning algorithms. Medical devices that use Machine Learning algorithms can identify the type of heartbeat with greater accuracy. In medical situations, machine learning methods are applied for the advantage of healthcare providers, organisations, and patients [6].

A classic statistical approach for binary classification problems is logistic regression. It is frequently used to detect the probability of CVD depending on a number of risk variables, including age, gender, cholesterol, blood pressure, and so forth. Common artificial intelligence methods for categorisation tasks include random forests as well as decision trees. It can manage non-linear interactions among data and are comprehensible. When compared with single decision trees, random forests especially are more resilient and are not as susceptible to excessive fitting. SVM is a method for supervised training that finds the appropriate hyperplane in a space with high dimensions to divide classes in order to accomplish the task of classification. CVD tasks for prediction have successfully used SVM. Because machine learning techniques, especially neural networks, are capable of learning structures from raw data, they have showed promise in CVD predicting applications. Although Recurrent Neural Networks or models based on transformers are capable of processing sequential information like EHR or clinical notes, CNN are able to be employed using images from medical imaging like X-rays, MRIs, or CT images. One kind of RNN that performs well for consecutive analysis of information is a type of Long Short-Term Memory (LSTM) network [7]. Through the processing of time-series data, such as electronic health records (EHRs), which are collections of occurrences over a period of time they can be used for predicting the possibility of CVD.

For the purpose of making simulations more accurate, attentive techniques were implemented in deep learning architectures in order to concentrate on pertinent portions of the input information. Models based on transformers have been successful in analysing medical records or free-text information for CVD diagnosis. One of these models is the well-known Bidirectional Encoder Representations from Transformers. Multiple frameworks are combined in ensemble methods to increase prediction precision and generalisation. For CVD prediction problems, methods like as bagging (Bootstrap Aggregating) or boosting (e.g., AdaBoost, Gradient Boosting Machines) can be used on a variety of core learners, including decision trees and artificial neural networks. Because labelled data is scarce in medical applications, learning through transfer entails applying large-scale, trained algorithms to problems using smaller datasets. Medical imaging data can be used to fine-tune models that have been trained, such as those generated on ImageNet, for the diagnosis of CVD. These techniques can be used separately or in combination, based on the CVD forecasting task's specific needs and the data that is available. When implementing machine learning and deep learning models in healthcare settings, other crucial factors to take into account are modelling interpretability, confidentiality of data, and compliance with regulations [8].

Among the research's noteworthy achievements is the creation of a novel hybrid CNN-XGBoost technique for the detection of CVD. By combining CNN's power for feature extraction from medical image collections with the resilience and understanding of a gradient boosting algorithm XGBoost, this approach provides an extensive solution for accurate CVD diagnosis. When it comes to identifying minor subtleties that point to cardiovascular issues, the CNN component is excellent at extracting complex patterns as well as characteristics from medical images like MRIs and X-rays. Following that, XGBoost expertly combines the learned traits with additional pertinent clinical data to improve prediction accuracy and offer valuable information about the relative significance of each feature to support medical decision-making. Better patient outcomes and earlier identification may emerge from this exciting method of enhancing CVD diagnostics via the combining of deep learning and traditional machine learning approaches. The key contributions of the suggested model are listed below:

1) The study presents an innovative method to anticipate CVD that combines CNN with the gradient boosting technique XGBoost.

2) The paper discusses the significance of normalization approaches such as Decimal Scaling, Min-Max, and Z-Score for pre-processed information. By implementing information within a specified range, this step improves the model's effectiveness and ensures information homogeneity.

3) The research automatically extracts pertinent information from medical images, such as images of the heart, using CNN in order to identify significant trends and traits associated with CVD.

4) Using the merged dataset, the CNN-XGBoost algorithm is trained in the research, and the model variables and hyper parameter settings are optimized. In order to enhance its effectiveness and fit the information more accurately, the algorithm's parameters are adjusted in this stage to increase the effectiveness and generality of the prediction.

5) To merge the predicted results of individual CNNs, XGBoost is utilized in ensemble learning approaches.

The paper is structured as follows: Section II comprises relevant material designed to help readers comprehend the proposed paper using existing methodologies, while Section III elaborates on the problem description. Section IV displays the proposed CNN-XGBoost architectures. Section V includes tabular and graphical representations of the results and performance indicators. Discussion in Section VI. Finally, in Section VII, the conclusion and future works are discussed.

II. Related Works

Mathur et al., [9] suggested the use of artificial intelligence in cardiovascular care, namely machine learning as well as deep learning. Artificial intelligence (AI) programs have helped us better comprehend cardiac failure and hereditary coronary artery disease. These uses led to better topics covered include treatment techniques for multiple cardiovascular conditions, novel approaches to medication therapy, and a post-marketing assessment of prescription pharmaceuticals. Applications that utilize AI face difficulties with medical application and comprehension, such as confidentiality of information, obsolete information, choice bias, and inadvertent perpetuation of prehistoric biases/stereotypes, leading to incorrect inferences. However, artificial intelligence (AI) is a disruptive invention with tremendous promise in wellness. Compared the detection skills of four healthcare structures: Dxplain, Iliad, Meditel, and QMR. Recommended that such initiatives be employed. Doctors that can effectively utilize the data offered by such platforms. The network-based model displayed 86.3% sensitiveness, 85.7% precision, and 85.7% reliability, in that order. Employed conventional coronal images automated categorization to real-time heart function assessment, achieving a 95% reliability rate.

Deshai et al., [10] suggested a two-stage strategy for accurately predicting serious heart problems. Involved training the neural network with an improved sparse auto encoder (SAE), an unstructured neural network that operates forecasting of a wellness for operation. The artificial neural network (ANN) focuses on the gathered materials. The SAE was properly designed to ensure a successful model. The recommended method outperformed the ANN classification's reliability as well as efficacy. A deep learning approach that works in a transudative way is suggested for sparse demonstration-based categorization. The system includes a fully connected layers and a convolution-based an autoencoder. An increased approach. Limited autoencoder ANN may accurately forecast serious cardiac conditions in a trustworthy and effective method. A sparse auto encoder identifies the most effective database demonstrations, while the ANN makes forecasts based on acquired attributes. The SAE can be improved through an Adam method and sequential normalization techniques. The classifier's accuracy on the studied datasets is 91%. Compared to traditional machine learning algorithms as well as artificial neural networks, the method that was suggested yielded better results.

Wang et al., [11] proposed networks to distinguish between Breast Artery Calcification (BAC) and non-BAC and use a pixelwise, patch-based method for identifying BAC. To evaluate the efficacy of the system, conducted a readership study with qualified physicians to offer reliable input. Evaluated the outcome of 840 full-field digitized radiographs across patients utilizing FROC plus calcium density characterization analyses. The FROC study indicates that deep learning reaches a degree of identification comparable to experienced specialists. Calcium density measurements the estimated calcium density closely matches the earth's truth, using a linear correlation producing an error of measurement of 96.24%. These findings show that deep learning techniques can be utilized to construct a computerized system for detecting BAC in mammograms, assisting in identifying and assessing individuals with risk factors for CVD. A computerized technique for detecting BAC and estimating calcium in mammogram was studied as potential risk factors for coronary artery disease (CAD) sickness. Investigated the relationships among the two viewpoints (CC vs. MLO) in the anticipated and ground-truth BAC regions, and calcium standards, and Within the R and L breasts. While these occurrences have not been thoroughly examined in the literature, it seems plausible to anticipate an extensive level of interaction between both perspectives and those of both the right as well as left breast.

Kelvin et al., [12] suggested the flexible characteristics of deep learning (DL) for CVD picture categorization, division, and identification for Effective control of challenges for implementing deep learning in the medical field. Recent advancements in the fields of computing and neuroscience have resulted in the invention of complex perceptron systems backward propagation neural network representations, and CNN. Classical models include neural network (CNN) a deep belief network (DBN) among others. The results of this research have sped up the growth of deep learning algorithms, enabling their widespread use throughout medical disciplines. DL approaches are used to recognize and locate surgical videos and have proven effective in practical practice. To identify and categorize the various forms of myocardial plaque, the use of CNN and recirculating neural networking mixture was suggested. This approach did not involve manual extraction of features and took into account both spatial and temporal data found in multiplanar reformatted images involving the coronary arteries. A 3D CNN could be used to extract characteristics in the coronary arteries. Such collected features can then be aggregated by running recurrent neural networks for two concurrent multi-class task classifications. The technique just requires the collection of the coronary artery baseline from coronary CT angioplasty as input from the user, in contrast to most current techniques that depend on myocardial epithelial fragmentation to identify and describe myocardial plaques and constriction. The technique effectively divides individuals into two groups: those lacking coronary plaque and those who need a second CV assessment because they have both constriction and myocardial blockage.

Kuang et al., [13] proposed the LSTM with reducing uneven duration among treatment phases to generate a time-dependent vector of features. Enhanced LSTM by normalizing uneven duration among treatment periods to produce longitudinal vectors of features. The memory loss limit uses a spatial vector of features to efficiently handle unpredictable time. The space that separates multi-period information improves the algorithm's ability to predict. The idea put forward improves the internal mechanism for forgetfulness gate input. Smoothing the uneven period of time yields the time variable vectors, which is subsequently sent into the forgetful gate to solve the problem. The uneven time gap creates an impediment to forecasting. The suggested changing forecasting approach outperformed the classic LSTM approach to accuracy in classification, demonstrating its efficacy. Modified the threshold value arrangement in the LSTM unit to acquire behavioural characteristics related with CVD progression at various time intervals before using LSTM to handle sequence information with irregular time intervals. Next, suggest use the objective to repetition a technique for predicting the hidden layer's output at every step that can make developing a model with varying time series lengths easier. In order to anticipate the patient receiving many diagnosing tag as results, a sigmoid function is ultimately used as the resulting layer of the framework as the activation element for the multi-tag result.

Komal et al., [14] suggested the use of neural network tree classifiers to predict CVD. The device Various training tree classification algorithms, including Random Forest, Decision Tree, Logistic Regression, support vector machine, and the k-nearest-neighbours algorithm, have been evaluated based on accuracy and AUC ROC ratings. The Random Woodland Machine Learning classification performed well in predicting CVD, with an 85% accuracy, ROC area under the curve of 0.8675, with implementation duration of 1.09 seconds. In this study, machine learning classifiers including K-nearest neighbours, Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine had been utilized for the intended use Heart illness. Prognosis. The suggested approach, that classified people with heart failure utilizing the random forest machine learning classification algorithm, beat every other classifier examined in terms of precision, achieving a higher 85.71% as well as a ROC average area under the curve of 0.8675. When compared to the remaining classifier in the investigation, the random forest classifier produced an inaccurate classification rate of 85.71% for the examples that are larger.

Jian et al., [15] suggested a machine learning-based technique which is simultaneously precise and effective in detecting heart problems. The system was created using classification methods, which comprise Artificial neural networks, Logistic regression, Support vector machine, Although common selection techniques like Relief, Minimal redundancy maximum significance, Least relative shrinking selection manager, as well as Local learning were used to remove unnecessary and redundant characteristics, K-nearest neighbour, Naïve bays, and Decision tree have also been utilized. In order to address the feature choice challenge, suggested a unique fast conditionally mutual data feature choosing approach. The characteristic selection methods are employed to pick elements in order to improve the precision of classification and shorten the overall time of operation. In addition, tweaking hyperparameters and learning the best techniques for model evaluation have been accomplished through the application of the leaving one topic out the cross-validation technique. The classifiers' abilities are evaluated using performance measurement measures. The participants' contributions have been examined in relation to the characteristics that the selection of features techniques choose. The experimental findings demonstrate the viability of using a classifier support vector machine in conjunction with the suggested feature selection algorithm to create an advanced neural network which can detect coronary artery disease. Comparing the recommended diagnosis methodology to other approaches that had been offered, excellent accuracy was attained. Furthermore, the suggested approach is simple to use in the medical field to identify cardiac illness.

Several research investigated the potential for the use of AI and deep learning in cardiovascular care, with a focus on applications such as comprehending heart failure and heart disease, therapeutic methods, pharmacological therapy, and pharmacological assessment. Problems like as privacy and bias were noted, but AI was rated beneficial for medicine. Strategies such as artificial neural networks with sparse autoencoders and deep learning algorithms were presented for reliable heart condition estimation, exceeding standard methodologies. Deep learning algorithms shown efficacy in detecting BAC and myocardial plaque, which aids in cardiovascular risk assessment. LSTM networks were improved to manage irregular time intervals, resulting in higher prediction accuracy. Neural network tree classifiers, notably Random Forest, were highly accurate for estimating CVD outcome. Furthermore, methods based on machine learning based on multiple classifiers and methods for choosing features were proposed for effective cardiac diagnosis, resulting in good accuracy and ease of application in hospitals. Overall, this research demonstrates the enormous potential of AI and deep learning to transform cardiac healthcare through better diagnostics and prediction abilities.

III. Problem Statement

Traditional machine learning methods, such logistic regression or support vector machines, are not particularly effective at diagnosing CVD because they depend on artificial features that may not accurately capture the complex patterns found in medical data. Moreover, autoencoders may be prone to over fitting, especially in situations with little training data, even if they are capable of learning models from raw data. A unique approach built on a hybrid CNN and XGBoost framework was created to overcome these shortcomings. The CNN-XGBoost method, in contrast to conventional machine learning approaches, makes use of CNNs' hierarchical characteristic acquisition capabilities, allowing it to identify and eliminate significant characteristics from medical images without the need for further advancement of features. Moreover, the model can effectively handle tabular data pertaining to patient data, health information, and clinical factors by combining CNNs with XGBoost's collaborative learning technique, which enhances the model's predicting accuracy and generalization. By integrating CNN capabilities with XGBoost, this hybrid technique addresses the shortcomings of previous approaches and provides a more accurate and long-lasting cardiovascular detection solution [15].

IV. PROPOSED HYBRID CNN-XGBOOST MODEL

The methodology commences by sourcing input data from a Kaggle dataset housing objective, examination, and subjective parameters pertinent to cardiovascular health. Pre-processing ensues, employing normalization techniques like Min-Max, Z-Score, and Decimal Scaling for consistency and heightened model efficacy. Feature extraction is executed utilizing Convolutional Neural Networks (CNNs), autonomously gleaning relevant features from clinical images, notably cardiac pictures, capturing pivotal patterns and characteristics. Following this, pre-processed and feature-extracted data undergoes ingestion into an XGBoost classifier for CVD prediction, capitalizing on its robustness in classification tasks. Hyperparameter tuning is subsequently conducted to refine model parameters, augmenting efficiency and generalization. Ensemble learning techniques are then deployed, amalgamating predictions from separate CNN and tabular algorithms. Ultimately, the framework's functionality is scrutinized and verified with appropriate metrics, evaluating its efficacy in predicting CVD on previously unseen data. The proposed methodology integrates an array of techniques, spanning from data pre-processing to ensemble learning, culminating in the development of a dependable and precise predictive model for CVD identification. Fig. 1 illustrates the sequential flow of the methodology, showcasing the cohesive integration of data processing, feature extraction, classification, and validation steps. This systematic approach ensures a comprehensive and robust framework for accurately identifying cardiovascular disease, thereby contributing to enhance clinical decision-making and patient care.

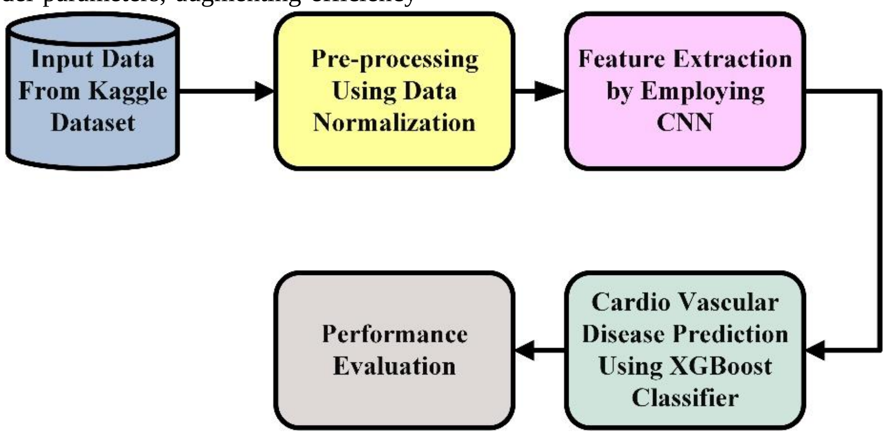

Fig. 1. Workflow of proposed methodology.

A. Data Collection

Through the use of a Kaggle dataset, it investigates the potential applications of artificial intelligence for medical prediction. Three types of data entry parameters are available: objective, examination, and subjective. It offers an alternative viewpoint on the issues people have with their health. It looks for patterns and predictions for a range of diseases using sophisticated modelling and assessment methodologies. By using statistical techniques and artificial intelligence (AI), the helpful details from this study can assist scholars and medical professionals in estimating difficult patient data [16]. Table I depicts the characteristics of the dataset.

B. Data Pre-Processing

Data mapping to a variety of scales is the aim of normalization strategies. There are many different kinds of normalizing methods in the literature. To preprocess the research data, three normalization methods are employed: Min-Max Normalization, Z-Score Normalization, and Decimal Scale Normalization. Min-Max Normalization linearly alters the initial dataset, ensuring that the normalized outcomes fall within a specified range, thereby enhancing consistency and model performance. Z-Score Normalization, also known as Zero Mean Normalization, utilizes the mean and standard deviation of the data to normalize it, effectively standardizing the distribution and facilitating comparison between different variables. Decimal Scale Normalization involves adjusting the numeric scale based on the smallest value of the characteristic, shifting the decimal point of values accordingly to maintain consistency across the dataset. These preprocessing methods confirm that the information is standardized, consistent, as well as suitable to be used for feature extraction as well as prediction modelling tasks.

TABLE I. DATASET CRITERIA

Feature	Type	Variable Name	Description
Age	Objective Feature	age	Integer (days)
Height	Objective Feature	height	Integer (cm)
Weight	Objective Feature	weight	Float (kg)
Gender	Objective Feature	gender	Categorical code
Systolic blood pressure	Examination Feature	ap_hi	Integer
Diastolic blood pressure	Examination Feature	ap_lo	Integer
Cholesterol	Examination Feature	cholesterol	1: normal, 2: above normal, 3: well above normal
Glucose	Examination Feature	gluc	1: normal, 2: above normal, 3: well above normal
Smoking	Subjective Feature	smoke	Binary (0 or 1)

Alcohol intake	Subjective Feature	alco	Binary (0 or 1)
Physical activity	Subjective Feature	active	Binary (0 or 1)
Presence or absence of CVD	Target Variable	cardio	Binary (0 or 1)

1) Min-max normalization: The initial data set is altered linearly by min-max normalization. The normalized outcomes fall into the specified range. The calculation is provided by, for translating a v value of a property from range $[min_{\ Z}, max_Z]$ to a new range of $[new\_min_Z, new\_max_Z]$.

$$\frac{v - min_{\ Z}}{max_Z - min_Z}(new\_min_Z - new\_max_Z) + new\_min_Z \quad (1)$$

where, v is the updated value inside the necessary range. Min-Max normalization has the advantage of annealing every value in a specific range[17].

2) Z-Score normalization: Another name for Z score normalization is Zero mean normalization. In this case, the difference between the standard deviation and mean are used to normalize the data. Next, the equation is

$$d' = \frac{d - mean(X)}{std(X)} \quad (2)$$

where, Mean(X) = sum of the all-attribute values of X; Std(X) =Standard deviation of all values of X.

3) Decimal scale normalization: The process of normalization of the numeric scale determined by the change of the characteristic's smallest value. The greatest absolute amounts of each property determine how the decimal point of values are shifted. A normalization equation for the decimals scale is,

$$d' = \frac{d}{10^n} \quad (3)$$

where, n is the smallest integer that max (| d′ |) <1.

C. Feature Extraction and Classification using CNN-XG-Boost Model

The suggested approach for predicting CVD makes use of an innovative model that combines XGBoost and CNN. First, normalization of data is used to preprocess the Kaggle dataset in order to guarantee consistency and improve model performance. The CNN model undergoes training on normalized data after pre-processing in order to extract pertinent characteristics gathered from cardiac pictures. The collected characteristics are then passed into an XGBoost algorithm to enhance accuracy and further fine-tune predictions. This hybrid strategy aims to improve the predictive potential for the identification of cardiovascular illness by utilizing the advantages of both CNN for feature extraction from images and XGBoost for the classification of images [18].

In the method of machine learning, a map of features is made for the data, and the classifier is then used to address the issue. Additionally, each challenge has a different set of facts, and the approaches used to solve it vary depending on the issue. CNN is therefore utilized to automatically produce characteristics and integrate them into the classifiers in order to prevent it. Among the CNN classifier's benefits is that, of all the classification algorithms, the method's list of layers that convert the input amount to output amount is the easiest. These aren't many distinct layers, and each of them uses an identifiable function to convert the input to the output. One drawback is that their projections do not take the object's direction and placement into account. Either forward or reversed, the convolution process operates far more slowly than, example, max pool. Every training stage will take considerably more time if the network is large.

Convolutional, ReLu, and max pooling layers are among the numerous layers that make up a CNN architecture. AlexNet served as inspiration for the design. Conv2D, ReLu, Max-pooling, and a completely connected layer make up its six layers. Additional layers, such as dropout, are incorporated into the network to improve training success. The dropout component is only turned on when you're training. In the forward motion (input to the function), the dropout stage arbitrarily removes a certain number of neurons and retains the remaining neuron after the forward transit. Only the drop after the backwards changes the non-dropped. One element which contributes to normalization is the dropouts. By teaching the algorithm resilient characteristics that do not rely on the neurons, the dropout layer helps the model prevent excessive fitting in its training process [19]. Fig. 2 represented the CNN-XGBoost Classification Network.

CNN usually consist of multiple layers, each with a specific function. Accessible filtering is used by the convolutional layer in order to acquire attributes from the information that is input. Activation layers introduces non-linearity and aids in capturing complicated trends. Layer pooled reduces computing load and preserving vital information by downsampling features networks. The resultant layer, which generates the final estimates, is the result of completely interconnected components that combine retrieved information for tasks involving extrapolation or categorization. All of these parts come together to create the framework of the CNN , making it possible to identify connections and extract details from intricate data sources like text as well as pictures [20].

An activating function (f), a selection term, and an array of filters that can be learned or kernel (K) are applied to the feature map(X) that is input within the convolutional layer of a CNN network. The neural layer's mathematical formula is expressed as:

$$Y = f\big((X \times K) + a\big) \quad (4)$$

where, *X* represents the input feature map, K represents the set of learnable filters/kernels. ‘a' represents the bias term, $*$ denotes the convolution operation represents the output feature map after applying the convolution operation, adding the bias, and applying the activation function f.

Following the convolution process, the feature map X is subjected, element-by-element, to an activation function. Tanh, sigmoid, and ReLU (Rectified Linear Unit) are examples of common activation functions.

$$Y = f(X) \quad (5)$$

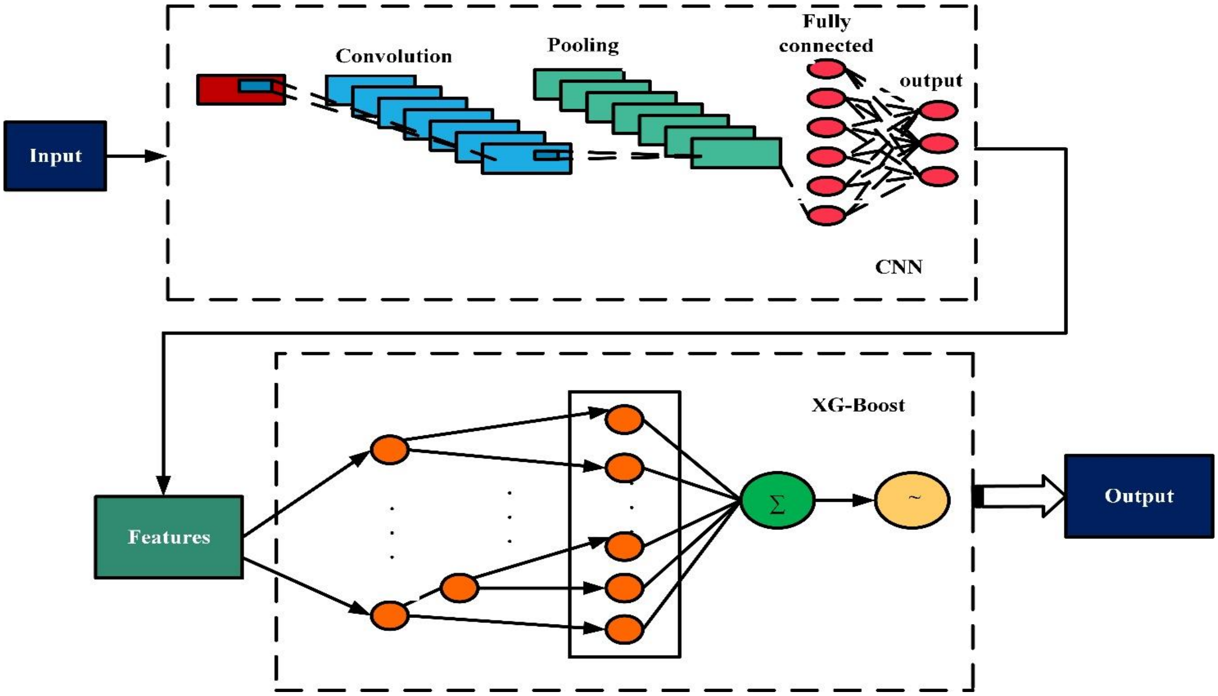

Fig. 2. CNN-XGBoost classification network.

Pooling layers is employed for down-sampling the map attributes, preserving significant data while reducing the number of dimensions. A popular method is called "max pooling," in which the highest value found inside a window is used. Assigning a letter P to the pool procedure

$$Y = P(X) \tag{6}$$

The maps of features have been flattened and put into a number of completely linked layers following one or more convolutional and pooling layers. The resultant value O can be computed as follows if we designate the flattening feature map as X, K as the weights, and s as the term representing the bias of the layer that is completely connected.

$$O = f(X.K + s) \tag{7}$$

The final estimates are produced by the output layer. The type of issue determines this layer's activation function. One frequent function of activation used for identifying binary issues is the sigmoid.

XGBoost is an effective technique for classification, in addition regression [21]. It functions as a set of applications that have won machine learning contests on Kaggle. Drawing from a gradient enhancing framework, XGBoost continually creates new decision trees in order to increase learner performance and efficiency while fitting a value using residual many rounds. With a better trade-off between bias and variance, XGBoost approximates the loss function using a Taylor expansion. It often requires less trees of decisions to achieve a greater accuracy. XGBoost uses the reduced model complexity to add normalization to the standard function. The residual error is fitted by utilizing the initial and second derivatives. Additionally, columns selection is supported by this technique in lowering excessive fitting and cutting down on processing. Thus, compared to the gradient boosting decision tree, higher improvements result in more the hyper-parameter. Still, it is challenging to adjust the extreme parameters in a reasonable way. In addition to the researchers' past knowledge and experience with tuning parameters, a reasonable setup necessitates a significant amount of time. Hyper-parameter optimisation works well in solving this issue. The equation for the prediction $\hat{y_i}$ using XG-Boost can be represented as follows:

$$\hat{y_i} = {}^{\Phi}(x_i) = \sum_{n=1}^{N} f_k(x_i) \tag{8}$$

where, $\hat{y_i}$ is the predicted output for the i^{th} observation, Φ (x_i) is the final prediction function, N is the number of weak learners, $f_k(x_i)$ is the prediction of the k^{th} weak learner for the input x_i

Established enhancing is how XGBoost actually manages excessive fitting and complexity of models. The process of regularization factors is employed to punish difficulty, and a function with losses is often included in the function of objectives utilized by XGBoost to assess the disparity between the actual and predicted values. In order to determine the perfect combination of learners with weaknesses, the model optimizes the aforementioned objective function while training. In general, the algorithm's additive feature—multiple weak learners merged to build a learner who is strong by weighing the total of their separate predictions—is encapsulated in the XGBoost predictions solution. XGBoost, also known as eXtreme Gradient Boosting, being a collaborative learning strategy that emphasizes correcting mistakes in previous models while concurrently building numerous decision trees to continuously reduce a predetermined loss parameter. The total prediction power is increased by building each new tree with the intention of capturing the error or residuals of the preceding ones. a number of the most popular and efficient machine learning methods for tasks such as classification and regression, XGBoost uses regularization terms in the goal function to prevent excessive fitting and integrates an innovative gradient boost method that maximizes both performance and accuracy of models.

The fundamental component of XGBoost is the goal function, which is created to penalize complexity while minimizing loss L. Usually, it is composed of two components: the regularization term Δ(f), which reflects what is predicted model, and the corresponding loss function [22].

$$L(x_i, \widehat{x_i}) + \Delta(f) \quad (9)$$

where, x_i is the true label of the i-th sample, $\widehat{x_i}$ is the predicted value for the i-th sample. To modify the model, one computes the slope of the impairment function in relation to the expected scores. The change in the gradient g_i in relation to the projected score for an overall distinguished function of loss L can be calculated as follows:

$$g_i = \frac{\Delta L(x_i, \widehat{x_i})}{\Delta \widehat{x_i}} \quad (10)$$

The goal of XGBoost's tree boosting technique is to minimize the total objective function by adding new models, or trees, to the system. The following provides the update procedure for appending an additional tree to the model is depicted in Eq. (11). The Algorithm 1 illustrates the mechanism of CNN-XGBoost method.

$$\widehat{z_i}^{(t)} = \widehat{z_i}^{(t-1)} + \Delta . \mathrm{h}(y_i) \quad (11)$$

Algorithm 1: CNN-XG-Boost mechanism	
Input: *Data from Kaggle dataset*	
Output: *Identification of CVD*	
Load input data (Age, Height, Weight, Gender, Systolic blood pressure, Diastolic blood pressure, Cholesterol, Glucose, Smoking, Alcohol intake, Physical activity.	
Y={y_1,y_2,y_3,…y_n}	*// data acquisition*
Pre-processing of data	
Linearly alters the initial dataset	*//minmax normalisation*
Utilizes the mean and standard deviation of the data	*//Z-Score Normalisation*
Adjusting the numeric scale	*//Decimal Scale Normalisation*
Feature Extraction	*// CNN*
Initialize the CNN model with random weights.	
Forward pass through the CNN layers	
Apply convolution operation using learned filters.	
Apply activation function to introduce non-linearity.	
Apply pooling operation to down sample the features.	
Extract the output feature maps from desired intermediate layers of the CNN model	
Classification	*//CNN-XG-Boost*

V. Results and Discussion

The use of a CNN-XGBoost model for cardiovascular illness diagnosis in Python produced good results, demonstrating its efficacy as a successful diagnostics tool. By combining the capabilities of CNN and the XGBoost boosting approach, the model displayed strong performance in distinguishing between both positive and negative cases of heart failure. This fusion approach successfully extracted complicated patterns from complicated cardiovascular data, resulting in improved diagnosis accuracy. The findings imply that the CNN-XGBoost model has potential for early identification and intervention in coronary artery disease, making it a significant tool for doctors in maximizing the treatment of patients and management. Further refining and confirmation of the algorithm on bigger data sets and various demographics is necessary to ensure its usefulness and usefulness in real-world healthcare environments.

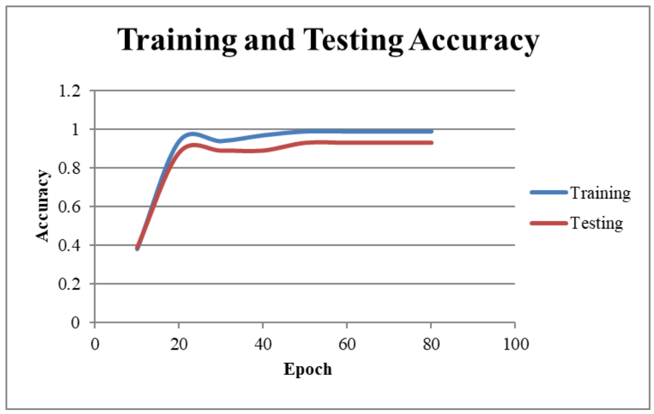

Fig. 3. Graphical representation for training and validating the accuracy of the proposed approach.

The training-testing accuracy graph for CVD detection using the CNN-XGBoost model depicts how well it performs during various phases of training and testing rounds. Initially, as the model is trained, its accuracy gradually improves, showing good training and extraction of features from cardiac data. In testing, the graph illustrates the model's capacity to generalize to previously unknown information, with accuracy declining at a high level, showing strong performance in diagnosing cardiovascular illnesses. This graph demonstrates the gradual improvement of training and testing accuracies, indicating little over fitting and a well-generalized framework. In general, the curve depicts the model's learning dynamics and accuracy in diagnosing cardiovascular disorders at different phases of growth and analysis. Fig. 3 shows Graphical representation for training and validating the accuracy of the proposed approach.

Training-testing deficit graph for CVD detection employing the CNN-XGBoost model demonstrates the model's minimization and adaptation potential. Initially, in training, the loss gradually lowers as the algorithm learns to eliminate mistakes and increase its prediction accuracy on the used training data. When training advances, the loss gradually decreases until it reaches a minimum, signifying optimal convergence. During testing, the loss graph illustrates the model's capacity to generalize to previously unseen data while keeping the loss reasonably low, indicating that it's able to generate correct predictions on new data points. The resulting curve is a visual representation of the model's learning dynamics, demonstrating the balance between complexity of the model and generalization how they perform, and highlighting its effectiveness in detecting cardiovascular illnesses while reducing overfitting from occurring. Fig. 4 shows Graphical representation of loss in proposed CNN-XG Boost.

Fig. 4. Graphical illustration of loss in suggested CNN-XGBoost.

TABLE II. EXPERIMENTAL RESULT ANALYSIS FOR DIFFERENT PARAMETERS WITH OTHER METRICS

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-sore (%)
Extra Tree Classifier [23]	90	87	91	89
Logistic Regression [23]	88	79	93	85
DNN [24]	87.59	97.77	76.27	65.54
LR [25]	87.60	87.05	90.98	88.97
SVM [25]	81.82	77.30	94.74	85.14
Proposed CNN-XG-Boost	98.7	98	97.9	96.98

Table II provides a comprehensive comparison of performance metrics across various machine learning models, including Logistic Regression (LR), Deep Neural Network (DNN), Support Vector Machine (SVM), Extra Tree Classifier, and the proposed CNN-XG-Boost approach. Each row corresponds to a specific model, while columns represent assessment criteria such as accuracy, precision, recall, and F1-score, expressed as percentages. For instance, the Logistic Regression model achieved accuracy ratings of 87.60%, with corresponding precision, recall, and F1-score values of 87.05%, 90.98%, and 88.97%, respectively. On the other hand, the DNN model garnered accuracy of 87.59%, but displayed higher precision at 97.77%, albeit lower recall at 76.27%, resulting in an F1-score of 65.54%. Similarly, the SVM and Extra Tree Classifier models underwent evaluation using these criteria, with respective performance scores recorded. However, it's noteworthy that the suggested CNN-XG-Boost model outshines its counterparts significantly. With impressive accuracy, precision, recall, and F1-score values of 98.7%, 98%, 97.9%, and 96.98%, respectively, it demonstrates unparalleled efficiency in the task at hand compared to conventional techniques. This comparative analysis underscores the superior performance of the CNN-XG-Boost model, indicating its potential as a highly effective tool for CVD identification and classification. By surpassing existing models across all key metrics, the suggested approach establishes itself as a frontrunner in the field, offering heightened accuracy and reliability in cardiovascular health assessment. The clear advantage exhibited by the CNN-XG-Boost model highlights its suitability for real-world applications, where precise and timely diagnosis is paramount. As such, its implementation could lead to enhanced patient outcomes and streamlined healthcare processes. Moreover, the robustness of the proposed model suggests its adaptability to diverse datasets and clinical scenarios, further solidifying its position as a valuable asset in the realm of cardiovascular disease management and prevention.

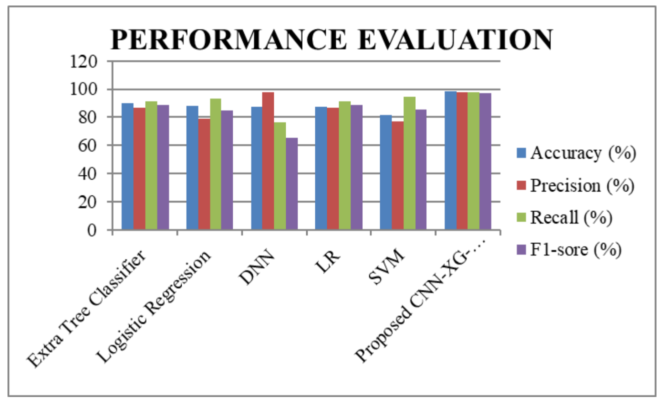

Fig. 5. Performance evaluation for different methods of classification.

The performance assessment of the suggested hybrid CNN-XGBoost model is presented in tabular format, highlighting its superiority over other types of classifiers, particularly in terms of accuracy. While the median recall and precision of existing approaches may surpass those of the suggested model, the average accuracy and recall serve as crucial indicators of categorization, alongside the F1-score. In Fig. 5, a pictorial representation of the performance evaluation elucidates the model's strengths and areas for improvement. The visual depiction aids in understanding how the suggested hybrid model fares against competing classifiers, emphasizing its superior accuracy and potential for enhancing categorization accuracy. Despite potential variations in individual metrics such as recall and precision, the overall performance of the CNN-XGBoost model shines through in terms of its accuracy, as indicated in the tabular assessment. This comprehensive evaluation approach ensures a nuanced understanding of the model's capabilities, allowing for informed decisions regarding its implementation in real-world scenarios. As the field progresses, continued refinement and optimization of the CNN-XGBoost model will likely further improve its performance metrics, potentially bridging any gaps in recall and precision observed in comparison to existing approaches. This iterative process of evaluation and enhancement ensures that the suggested model remains at the forefront of CVD classification research, driving advancements in diagnostic accuracy and patient care. Table III provides the comparison of various dataset with the proposed dataset used in the study. It shows the proposed work dataset achieves higher accuracy when compared with another dataset.

TABLE III. DATASET COMPARISON

Dataset	Accuracy
AptaCDSS-E [26]	66.00
GitHub PCG – CNN [27]	86.57
Pascal Challenge [28]	96.25
Proposed CVD Dataset	98.7

VI. DISCUSSION

The use of a CNN-XGBoost model for CVD detection offers an exciting opportunity to improve diagnostic accuracy and predictive capacities in this essential healthcare domain. This hybrid model, which combines the CNN for feature extraction with the XGBoost algorithm for classification, can successfully capture subtle correlations and trends in cardiac data, allowing for more precise estimations. Compared to other techniques such as Extra Tree Classification algorithm [23], logarithmic regression [24], DNN [24], LR [25], and SVM, the CNN-XGBoost model has significant advantages. The proposed CNN-XGBoost hybrid method for cardiovascular disease (CVD) detection achieves superior accuracy (98.7%) and balanced precision (98%), recall (97.9%), and F1-score (96.98%) compared to existing methods like Extra Tree Classifier, Logistic Regression, DNN, LR, and SVM. This approach integrates CNN feature extraction with XGBoost classification, offering robustness and precision in CVD prediction. First, the CNN component enables the automatic extraction of features from raw input data, avoiding the requirement for feature engineering by hand and possibly enhancing the accuracy of the model. Furthermore, the XGBoost algorithm, which is well-known for its durability and effectiveness when working with huge datasets, refines the retrieved features and increases classification accuracy. The combined use of deep learning and gradient boosting approaches provides a potent approach to CVD identification that outperforms current techniques.

Furthermore, the CNN-XGBoost model's capacity to incorporate time and space variables from heart data sets might offer more extensive insights into the underlying physiological processes that contribute to CVD. By efficiently collecting complicated trends and interconnections in the data, this model has the potential to uncover subtle signals and early warning signals for cardiovascular problems, allowing for prompt treatment and proactive measures. In addition, the XGBoost algorithm's interpretability enables physicians to obtain insights into the major aspects influencing CVD prediction, resulting in more informed decision-making and individualized patient care plans. Overall, the CNN-XGBoost model is a promising approach to CVD detection, with enhanced accuracy, interpretability, and the possibility for early detection and intervention, resulting in improved patient outcomes and healthcare oversight.

VII. CONCLUSION AND FUTURE WORKS

The proposed method represents a significant advancement in cardiovascular disease (CVD) identification and classification, offering superior accuracy and efficiency compared to existing classifiers, making it a promising avenue for future research. With an impressive accuracy rate of 98.7%, recall of 97.9%, precision of 98%, and an F1 score of 96.98%, the provided model outperforms previous approaches, showcasing its robustness and reliability. In contrast to the Deep CNN-SVM approach, the chosen classification methodology of the proposed model demonstrates exceptional efficiency. By leveraging the hybrid technique of CNN and XGBoost, the model achieves remarkable performance in detecting CVD, providing clinicians with a reliable and accurate method of diagnosis. This fusion of CNN and XGBoost capitalizes on the strengths of both methods, effectively identifying subtle patterns within cardiovascular data for early detection and management. The high predictive capability of the CNN-XGBoost model underscores its potential to revolutionize cardiovascular healthcare, empowering clinicians to make timely and informed decisions for patient care. As research progresses, further refinement and validation of this model on diverse datasets will enhance its efficacy and applicability in real-world clinical settings. Continued experimentation with various topologies and refinement strategies holds the key to further improving the efficiency of the CNN-XGBoost classifier. By optimizing model parameters and exploring novel approaches, the model can achieve even higher levels of accuracy and performance, ultimately leading to better medical outcomes and more efficient treatment strategies. Future validation of the suggested approach across different geographical locations and medical disciplines will further validate its effectiveness and generalizability. By testing the model on a diverse range of datasets, researchers can ensure its robustness and reliability across various populations and healthcare settings. In conclusion, the CNN-XGBoost model represents a powerful tool in the realm of CVD detection, offering a comprehensive and accurate approach to diagnosis. With ongoing refinement and validation, this model has the potential to significantly impact cardiovascular healthcare, improving patient outcomes and advancing medical practice.

REFERENCES

[1] "Cardiovascular diseases (CVDs)." Accessed: Mar. 22, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] G. Battineni, G. G. Sagaro, N. Chintalapudi, F. Amenta, D. Tomassoni, and S. K. Tayebati, "Impact of obesity-induced inflammation on cardiovascular diseases (CVD)," Int. J. Mol. Sci., vol. 22, no. 9, p. 4798, 2021.

[3] "Cardiovascular Disease: Types, Causes & Symptoms," Cleveland Clinic. Accessed: Mar. 22, 2024. [Online]. Available: https://my.clevelandclinic.org/health/diseases/21493-cardiovascular-disease.

[4] "What is Cardiovascular Disease?," www.heart.org. Accessed: Mar. 22, 2024. [Online]. Available: https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease.

[5] O. Cheikhrouhou, R. Mahmud, R. Zouari, M. Ibrahim, A. Zaguia, and T. N. Gia, "One-Dimensional CNN Approach for ECG Arrhythmia Analysis in Fog-Cloud Environments," IEEE Access, vol. 9, pp. 103513–103523, 2021, doi: 10.1109/ACCESS.2021.3097751.

[6] M. Ouyang et al., "A review of biosensor technologies for blood biomarkers toward monitoring cardiovascular diseases at the point-of-care," Biosens. Bioelectron., vol. 171, p. 112621, 2021.

[7] M. M. Ahsan and Z. Siddique, "Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review." arXiv, Dec. 13, 2021. Accessed: Mar. 22, 2024. [Online]. Available: http://arxiv.org/abs/2112.06459.

[8] K. Saikumar and V. Rajesh, "A machine intelligence technique for predicting cardiovascular disease (CVD) using Radiology Dataset," Int. J. Syst. Assur. Eng. Manag., vol. 15, no. 1, pp. 135–151, 2024.

[9] P. Mathur, S. Srivastava, X. Xu, and J. L. Mehta, "Artificial Intelligence, Machine Learning, and Cardiovascular Disease," Clin. Med. Insights Cardiol., vol. 14, p. 117954682092740, Jan. 2020, doi: 10.1177/1179546820927404.

[10] N. Deshai and S. Ramya, "Prediction of Heart Disease with Autoencoder based ANN," Int. J. Eng. Res., vol. 9, no. 5, 2021.

[11] J. Wang et al., "Detecting Cardiovascular Disease from Mammograms With Deep Learning," IEEE Trans. Med. Imaging, vol. 36, no. 5, pp. 1172–1181, May 2017, doi: 10.1109/TMI.2017.2655486.

[12] K. K. L. Wong, G. Fortino, and D. Abbott, "Deep learning-based cardiovascular image diagnosis: A promising challenge," Future Gener. Comput. Syst., vol. 110, pp. 802–811, Sep. 2020, doi: 10.1016/j.future.2019.09.047.

[13] K. Junwei, H. Yang, L. Junjiang, and Y. Zhijun, "Dynamic prediction of cardiovascular disease using improved LSTM," Int. J. Crowd Sci., vol. 3, no. 1, pp. 14–25, May 2019, doi: 10.1108/IJCS-01-2019-0002.

[14] N. K. Kumar, G. S. Sindhu, D. K. Prashanthi, and A. S. Sulthana, "Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers," in 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India: IEEE, Mar. 2020, pp. 15–21. doi: 10.1109/ICACCS48705.2020.9074183.

[15] F. Li, Z. Liu, H. Chen, M. Jiang, X. Zhang, and Z. Wu, "Automatic Detection of Diabetic Retinopathy in Retinal Fundus Photographs Based on Deep Learning Algorithm," Transl. Vis. Sci. Technol., vol. 8, no. 6, p. 4, Nov. 2019, doi: 10.1167/tvst.8.6.4.

[16] "Cardiovascular Disease dataset." Accessed: Mar. 21, 2024. [Online]. Available: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset.

[17] H. Henderi, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," IJIIS Int. J. Inform. Inf. Syst., vol. 4, no. 1, pp. 13–20, Mar. 2021, doi: 10.47738/ijiis.v4i1.73.

[18] "Detection of Cardiac Arrhythmia from ECG Using CNN and XGBoost," Int. J. Intell. Eng. Syst., vol. 15, no. 2, pp. 414–425, Apr. 2022, doi: 10.22266/ijies2022.0430.38.

[19] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning," in 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India: IEEE, May 2018, pp. 2319–2323. doi: 10.1109/RTEICT42901.2018.9012507.

[20] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," in 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India: IEEE, Feb. 2019, pp. 1–7. doi: 10.1109/ICECCT.2019.8869364.

[21] "Detection of Cardiac Arrhythmia from ECG Using CNN and XGBoost," Int. J. Intell. Eng. Syst., vol. 15, no. 2, pp. 414–425, Apr. 2022, doi: 10.22266/ijies2022.0430.38.

[22] W. Jiao, X. Hao, and C. Qin, "The Image Classification Method with CNN-XGBoost Model Based on Adaptive Particle Swarm Optimization," Information, vol. 12, no. 4, p. 156, Apr. 2021, doi: 10.3390/info12040156.

[23] R. Shafique, A. Mehmood, S. Ullah, and G. S. Choi, "Cardiovascular Disease Prediction System Using Extra Trees Classifier," In Review, preprint, Sep. 2019. doi: 10.21203/rs.2.14454/v1.

[24] A. Alqahtani, S. Alsubai, M. Sha, L. Vilcekova, and T. Javed, "Cardiovascular Disease Detection using Ensemble Learning," Comput. Intell. Neurosci., vol. 2022, pp. 1–9, Aug. 2022, doi: 10.1155/2022/5267498.

[25] E. K. Hashi and Md. Shahid Uz Zaman, "Developing a Hyperparameter Tuning Based Machine Learning Approach of Heart Disease Prediction," J. Appl. Sci. Process Eng., vol. 7, no. 2, pp. 631–647, Oct. 2020, doi: 10.33736/jaspe.2639.2020.

[26] Z. Alkayyali, S. A. B. Idris, and S. S. Abu-Naser, "A Systematic Literature Review of Deep and Machine Learning Algorithms in Cardiovascular Diseases Diagnosis," J. Theor. Appl. Inf. Technol., vol. 101, no. 4, pp. 1353–1365, 2023.

[27] "GitHub - vinayakumarr/Signal-Processing-and-Pattern-Classification: Signal-Processing-and-Pattern-Classification - Atrial fibrillation & PCG classification." Accessed: May 27, 2024. [Online]. Available: https://github.com/vinayakumarr/Signal-Processing-and-Pattern-Classification.

[28] "Cardiovascular Disease pascal dataset." Accessed: May 27, 2024. [Online]. Available: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset.