Prediction of Heart Disease Using Deep Convolutional Neural Networks

Awais Mehmood[1] · Munwar Iqbal[1] · Zahid Mehmood[2] · Aun Irtaza[1] · Marriam Nawaz[1] · Tahira Nazir[1] · Momina Masood[1]

Abstract

Heart diseases are currently a major cause of death in the world. This problem is severe in developing countries in Africa and Asia. A heart disease predicted at earlier stages not only helps the patients prevent it, but I can also help the medical practitioners learn the major causes of a heart attack and avoid it before its actual occurrence in patient. In this paper, we propose a method named CardioHelp which predicts the probability of the presence of cardiovascular disease in a patient by incorporating a deep learning algorithm called convolutional neural networks (CNN). The proposed method is concerned with temporal data modeling by utilizing CNN for HF prediction at its earliest stage. We prepared the heart disease dataset and compared the results with state-of-the-art methods and achieved good results. Experimental results show that the proposed method outperforms the existing methods in terms of performance evaluation metrics. The achieved accuracy of the proposed method is 97%.

Keywords Health care system · Heart disease · Machine learning · Convolutional neural networks

1 Introduction

The world is currently facing a significant number of deaths mainly caused by heart attacks. Developing countries, especially Asian and African regions, face a significant amount of failures to save human lives just because of the late determination of the severity of the attack [1, 2]. Detection of a heart attack at an early stage may significantly help prevent the attack. Day by day practice by medical practitioners generated a treasure of datasets that can be analyzed to determine the important attributes while diagnosing a heart attack [3]. Unfortunately, currently, these datasets are not being effectively to serve the purpose. The main goal of the research is to use those real-life datasets in a way that may help timely predict a potential heart attack. Various data analysis and data mining techniques are there to serve this purpose [4]. A lot of people die experiencing symptoms that were previously undiscovered or simply ignored. It is time to predict heart disease before its actual occurrence. There are several main causes of heart disease. Some of them may be high cholesterol levels, blood pressure, smoking, use of alcoholic drinks, high sugar, lack of physical activities, cardiovascular disease (CVD), and a hypertensive heart [5].

The importance of data science and Big Data in terms of online shopping, search, and multimedia is not much less than the Internet in the current era. People want to know more about their surroundings. The most important questions that come in a typical user's mind are what is happening, why it happened, and what is likely to happen in the future. An analyzer may even want to know what will happen through ages. A user may be curious about a loan being offered by a bank, but he/she may want to first analyze their services and potential cost and benefits involved in getting a loan from that bank. Another user may want to invest his money and want to know the best way to invest it. Someone may be unsatisfied with their daily routine and want to optimize it and even do this optimization in a few moments. He may also want to know the trends regarding this matter. Another major application of Big Data is involved in online shopping where the consumer is always confused that which product is better to buy from a haystack of brands. Some researchers and big data professionals claim that IoT may one day outperform or take-over big data as the most glorified technology in the world [6]. It may be true someday but one thing to be con-

✉ Tahira Nazir
  tahira.nazir77@gmail.com

1 Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan

2 Department of Computer Engineering, University of Engineering and Technology, Taxila, Pakistan

sidered that IoT cannot come alive without big data. Gartner defines Big Data as a set of information assets attributing to large volumes, huge velocity, and height of variety, at the same time demanding cost-effective and innovative shapes of information processing to support the advanced insight and decision-making capabilities [7]. In real life, most of the data are voluminous. As an instance of examples, social media, especially Facebook has 800 Million active members. There are around four billion phones and more than 25% are smartphones, and there are billions of RFID tags as well.

1.1 Big Data

The term Big Data refers to a collection of large datasets that are impossible to process with traditional computing techniques [7]. Normally, there is a centralized server in traditional enterprise systems for the storage and processing of data. Here comes the need for MapReduce, which, basically is a programming model used to write applications with the ability to process Big Data on multiple nodes in a parallel fashion. MapReduce is responsible to provide analytical capabilities to analyze very large volumes of data with increased complexity. The volume, variety, and velocity became known as the 3 V's of big data. Further, Oracle added another V i.e., "Value" of big data meaning that how much data is valuable and important to analyze. The variety of data can be structured, un-structured, and semi-structured. The velocity of data can be determined by the incoming data rate. The incoming rate can be as slow incoming data and fast incoming data. On the other hand, the volume of data can be in the form of Terabytes and Exabyte.

In the past few years, Big Data analytics is being used by different companies in a more and more meaningful way. Without analysis, data are pretty much useless. Recently, large companies have understood the worth of data and have tried to give better facilities and services to their clients. Like, Google works by giving relevant advertisements to user's web surfing and amazon suggests the products according to one's interest and YouTube recommender system that displays different videos relevant to the user's previous video search history.

The processing of big data is an exhaustive task. One may face different challenges during data loading, storing, and processing. The first challenge is "Large Volume". We must store a large dataset on disk. Hence, more hardware resources/disks are required to store because single hardware is not enough. Our physical structure should be vertically scalable as we have to add more hardware like RAM, CPU, and memory, etc. The physical structure should also be horizontally scalable to increase the number of nodes/machines to distribute the data among them. Other challenges include

high throughput, heterogeneity, relating & linking, and data complexity.

1.2 Machine Learning

The practical implementation of artificial intelligence, in which the user can program intelligent algorithms to get more accurate results and predict the output within an acceptable range [8]. ML algorithms can be divided into supervised and unsupervised learning; in supervised learning, input data along with target values are given to the algorithm to train on and predict the output values with a certain accuracy. The algorithm train from the given data and can apply the learned model on new datasets. The concept is similar to predictive modeling and data mining techniques in which we look deep into datasets and find the patterns. Unsupervised learning, on the other hand, is used for more complex jobs and doesn't require to be trained by giving the required outcome data. The target of unsupervised learning is to group those datasets into sensible classes [9, 10].

1.3 Deep Learning

The term 'deep learning' (DL) is also called hierarchical or deep-structured learning [11–13]. Unlike, task-based methods, DL is a type of ML technique that is based on learned-data representation and here the learning can either be supervised, unsupervised, or semi-supervised. The models of DL are vaguely encouraged from the working of biological nervous systems like how the information is processed and communicated in it. However, these DL techniques are structurally and functionally different from human brains. These differences make them incompatible with neuroscience evidence. The architectures of DL such as convolutional neural networks (CNN), DL networks, recurrent NN, and deep belief networks have been employed to various research areas including recognizing human speech, computer vision (CV), audio recognition, manipulation of natural language, machine translation, filtering social sites, drug design, bioinformatics, processing of the medical image, board game programs, and material examination. These advanced machine learning models have generated equal to and, in some scenarios better results than humans.

1.4 Applications of Deep Learning

Recently, DL techniques have regenerated the NN models. Researchers have represented the stacked restricted Boltzmann machines and autoencoders, which are exhibiting the remarkable performance in the field of digital image processing. The great advancement in the variation of NN models

has allowed the DL techniques to be employed in the field of CV, audio/visual content processing, and many other research areas. RNN techniques work well in processing the sequenced-data; therefore, several RNN models' variations are developed and employed in the field of sequence-based tasks like long short-term memory (LSTM) and gated recurrent unit (GRU). These methods are showing improved performance in numerous areas like in the field of recognizing handwriting, translating and modeling languages, and acoustic speech modeling.

1.5 Applications in Healthcare

Because of the promising results of DL approaches in various fields, now researchers are employing these methods in the field of medicine as well [14]. Some of the main research areas of DL in the field of healthcare are the detection of phenotypic patterns from serum uric acid measurement, determination of physiologic patterns, and predicting the severity level of various diseases [15]. The proposed work is based on detecting the HF at the early stages by employing the available HER information of patients.

To understand the associations between various clinical procedures or among the Unified Medical Language System (UMLS), now, DL [4, 8, 11] approaches are employing the textual data of healthcare units by utilizing the Skipgram method [16, 17]. Skip-gram works by identifying the low-dimensional HER data representation like process, diagnostic, medication codes, etc. We employed this concept in our proposed technique to obtain the same data representation. Our work is concerned with temporal data modeling by utilizing CNN for HF prediction at its earliest stage.

The main contributions of the proposed method are as follows:

- The proposed method (named as CardioHelp) predicts the probability of the presence of cardiovascular disease in a patient by incorporating a state-of-the-art deep learning algorithm called convolutional neural network.
- To our best knowledge, this is the first-time deep learning model applied in the medical field for predicting a coronary heart disease (CHD) which works with just 14 attributes.
- We prepared the heart disease dataset and compared the results with state-of-the-art methods and achieved good results.

The rest of the paper is organized as follows; Sect. 2 is a brief description of the research work already carried out in the prediction of heart diseases. Section 3 elaborates on the proposed framework. Section 4 is a description of the incorporated dataset and its attributes that are used in this research work. Section 5 is a detailed description of the achieved results.

2 Related Work

A significant amount of research work in this field has already been carried out by various researchers. This section performs an overview of the research work already carried out in the prediction of heart disease with the help of advanced ML algorithms and big data technologies. Various pattern recognition and data analysis methods for model prediction to diagnose cardiovascular diseases have been used by different research groups. State-of-the-art machine learning algorithms like K-nearest neighbor algorithm, Naïve Bayes classifier, genetic algorithm, decision trees, ANNS, and deep neural network was used for carrying out the experimental work. It is noted that in the majority of the researches performed when increase the features and combinations of the above-mentioned techniques are used the accuracy calculated is high [16]. A significant amount of research has already been carried in the prediction of cardiovascular disease.

Let us now briefly review some of the important and relevant work done by various research groups in this area. In a similar context, Qrenawi, Mohammed et al. [18] have successfully applied an ontology-based data mining technique on a special dataset of diabetics with cardiovascular disease. This research work was carried out to identify any relationship between the history of 'type two' diabetic patients and their laboratory tests advised by a medical practitioner. A later phase of the research work focuses on using some frequent pattern discovery and ontology-based, rule induction methods incorporating data mining algorithm named as RMonto. The key outcome of this research is that the usage of an ontology-based technique lessens the number of attributes/properties in the phase of preprocessing and aids in most of the data mining phase. Additionally, the authors successfully achieved an accuracy of 90%. Nguyen et al. [19] proposed a real-time deep learning framework for heart disease classification in an IOT-based medical environment. The heartbeat signals collected from ECG devices are decomposed into the wavelet coefficient using WPD (Wavelet Packet Decomposition) algorithm and features are extracted by applying a WPCA (wavelet-based kernel PCA). A deep neural network based on backpropagation using three hidden layers having 80 40 20 nodes used to classify the heart disease. In [17], an expert system using deep neural networks is developed by H. Van Pham. The knowledge base is represented using fuzzy rules and updated according to the doctor's preference to improve the database and help them in making the right decision based on heart disease risk level.

Various researchers have successfully applied intelligent machine learning algorithms in decision support systems. For instance, we may refer to the research work carried out by Isra'a Ahmed et al. [20], who developed an intelligent medical decision support system based on data mining techniques. This work incorporates a total of five data mining algorithms

with large datasets. The main purpose of this incorporation was to access and analyze the risk factors related to heart diseases in the context of statistical analysis for the identification of heart disease. This made it easy to compare the performance of different implemented classifiers, i.e., Decision Tree, Random Forest, Naïve Bayes, Discriminant, and support vector machine. The selected classification models were implemented on two different datasets to show that their approach is practically viable. The research paper concludes that all classification algorithms are fairly predictive and can give a nearly correct answer. However, the authors noted that the decision tree outperforms other classifiers. The random forest algorithm stands second in this categorization. Modern research is now able to predict the development of a heart attack or a heart disease before its occurrence. A risk factor-based approach proposed by Kanchan More et al. [21] can predict the risk of developing heart attack by using smartphone technology. They developed an android application which is integrated with a clinical database constructed by data of more than five hundred patients admitted in a cardiac hospital, with the final diagnosis. A correlation of the presence of ischemic heart disease (IHD) with available data is performed and various risk factors like diabetes, hypertension, smoking, dyslipidemia, family history, stress, obesity, and existing clinic symptoms were considered. The data mining technology was used to mine the data and a nominal score was generated. Three classes of risk were created i.e., low, medium, and high for IHD to the score. The data of 89 participants with acute coronary syndrome (ACS) were used to test the performance of the android app. While comparing the patients with data generated by scores, they found a substantial correlation of having a cardiac incident when a comparison of the high and low category was $p = 0.0001$. 89% of patients with high category had IHD while only 12.5% of patients with low category were found to have IHD. Similarly, a significant difference between high and medium was found with $p = 0.0001$. An observation demonstrated that 86.7% of patients with ACS had high scores.

We have another example of a software application that successfully predicts the occurrence of a cardiac attack. Application software was developed by Ammar Aldallal et al. [22]. Medical practitioners could be able to predict the occurrence of non-communicable diseases (NCDs) by using this software. The proposed software application was examined by using the patients' records gathered from the Bahrain Defense Force Hospital. Practitioners were asked to execute and test this application in the mentioned hospital. The research work claimed that the proposed prediction model can instantly predict NCDs with a satisfactory level of effectiveness, and efficiency. The application is claimed to be able to help a medical practitioner to make proper decisions toward patient health risks. In addition to the software-based approaches, some researchers also focused on developing a

sensor-based and hardware-oriented approach to analyze the factors involved in the development of heart disease. A similar sensor-based approach is proposed by Johanna O'Donnell et al. [23]. The authors proposed that the data collected from multi-sensor patch data can be used for analyzing heart failure. These data were collected by multi-sensor patches from patients suffering from heart failure. The authors claim that this data can be helpful initially analyzing the sleep patterns and activities of the heart failure patients by the severity of heart disease. Chest-worn multi-sensors were provided to thirteen heart failure patients, and they were asked to wear the devices for almost seven consecutive days. Multi-sensor data gather from eleven out of those thirteen patients were found to be of high quality, and included in the analysis. A potential difference between heart rate, sleep angle, wake-time activity was found among heart failure patients with different severities. The authors proposed that larger studies are important to coherently analyze the role of activity and sleep as markers for the occurrence of potential heart failure. Mehmood et al. [24] performed prediction of a potential heart attack by using the attributes extracted from the dataset taken from the UCI repository. The authors insisted on the importance of attribute extraction techniques for mining information for the prediction. They added that various patterns can be derived to predict heart disease earlier by utilizing attribute extraction techniques. Several techniques in Artificial Neural Network (ANN) are explained in this research work. The paper shows that ANN gives an accuracy of 94.7, but the principal component analysis improved the Accuracy to 97.7%. The utilization of the data mining technique is also performed by Alizadeh-dizaj et al. [25]. The authors investigated the performance of data mining algorithms and predicted the risk of stroke in suspected stroke patients using a decision tree based on the risk factors that affect it. The main database contains 1184 records. In the modeling phase, the Classification Tree, Naïve Bayes, Neural Network, SVM, and KNN algorithms have been used. This study states that physical inactivity, high cholesterol, cardiovascular disease, history of transient ischemic attack, history of the previous stroke, and high blood pressure were the most effective variables while predicting potential heart disease. This research work established a decision tree that developed some rules that can be used as a model for predicting the risk of stroke in patients. The claimed accuracy of the model was 95.52%. The research work shows that it is possible to determine the stroke risk for a new sample with specific characteristics by applying the rules. A method to investigate the performance of different classification algorithms such as DT, NB, K-NN, and NN on heart disease dataset was proposed by T. John Peter et al. [26]. They classify the patients' records and predict who is having heart diseases. After applying various classification algorithms, the authors found that the Naïve Bayes classifier gives better accuracy than other

classifiers. For the sake of efficiency, the authors reduced the dimensionality of data using the attribute selection methods.

A *k*-means clustering algorithm has been used by researchers [27] incorporating a warehouse for heart diseases. They applied MAFIA (Maximal Frequent Itemset Algorithm) for calculating the overall significance of the most occurring patterns leading to heart attacks. Another neuro-fuzzy algorithm incorporating a layered approach was proposed in [28] which helped predict the occurrences of coronary heart diseases. The approach was simulated in MATLAB tool and shown significant results with a very low error rate and a higher efficiency. Another technique incorporating the association rule mining algorithms was introduced in [29] which makes use of transactional clustering datasets with a sequence number for predicting the likelihood of potential heart disease. The said technique was implemented in the C programming language. The key idea of the technique was to use smaller cluster sizes which helped it increase memory usage efficiency. There is a vast variety of research works carried out for predicting potential heart disease, for instance, a research group incorporated some advanced machine learning and data mining techniques like genetic algorithm and fuzzy logic to predict and analyze a potential heart attack [30]. In addition to working on single and standalone machine learning algorithms, some researchers worked on hybrid models as well [31]. They introduced a technique which makes use of ANNs and machine intelligent hybrid algorithm to predict the occurrences of a heart attack. Another research [32] introduces a prototype that makes use of Naïve Bayes and Weighted Associative Classifier (WAC) which helps in efficiently predicting the probability of the patients receiving a heart attack. A research work proposed a solution to the heart attack prediction problem by developing a web-based intelligent system that uses Naïve Bayes classifier to diagnose heart disease by answering a set of complex queries. A technique incorporating Association Rules was presented in [33], incorporating advanced data mining techniques to improve the overall accuracy for predicting heart disease. The research work proposed in [34] incorporates an algorithm with search constraints. The ideas were to minimize the total association rules and validate the approach used for training and testing of the model. A comparative study [35] of different data mining algorithms like support vector machines, artificial neural networks, and decision trees was presented which developed a model that predicted more than 500 cases. A popular dataset with 15 attributes was incorporated in [36] which uses Naïve Bayes classifier and ANNs for prediction of heart disease. Additionally, the main of focus of some recent researches is to predict a CVD or CHD by incorporating decision trees [37], significant features, and ensemble learning (SFEL) model [38],

and sparse autoencoder-based approach [39]. In this regard, a recent survey on the techniques for predicting a CVD has also been presented in [40].

3 Methodology

This work primarily focuses on the prediction of heart disease with the help of a well-established dataset and a state-of-the-art machine learning algorithm called Convolutional Neural Networks. We name the proposed framework as CardioHelp incorporating a state-of-the-art dataset available at [41]. The following subsections contain a detailed description of the proposed technique.

3.1 Majority Voting with LASSO Shrinkage

The least absolute shrinkage or LASSO is a technique involving the regression method which is effectively used for the regularization and selection of variables to improve the prediction accuracy and interpretability of the produced model. The LASSO technique shrinks the data values to a central point P_c and helps in the elimination of parameters and selection of variables. This type of regression technique is well-suited for highly multicollinear models. A penalty equal to the absolute value of coefficients magnitudes is added by the LASSO regression, out of that, some coefficients are eventually eliminated from the model after becoming zero, resulting in a model with fewer coefficients due to variable elimination. Due to their quadratic nature, the LASSO solutions work toward a unique goal, i.e., to minimize:

$$\sum_{j=1}^{n}\left(y_j - \sum_j x_{jk}\gamma_k\right)^2 + \lambda\sum_{k=j}^{q}|\gamma_k| \tag{1}$$

As mentioned in Eq. (1), these results in an interpretation easily understandable by the regression model because a subset of values (denoted by γ) becomes zero after completion of the shrinking process. The above equation includes a parameter λ, which is a tuning parameter denoting the amount of shrinkage. Keeping that in mind, we can say that no parameters are removed from the model when $\lambda = 0$. More coefficients are set to zero and eliminated from the model concerning the increase in λ. The variance is increased with a decrease in λ, and bias is increased with an increase in λ. The importance of variables in terms of their contribution to the underlying variation is interpreted as γ value for a variable or factor. The variable with $\gamma = 0$ is considered unimportant and resultantly ignored. It should be noted that lack of balance in the dataset (data imbalance) results in misleading

results by LASSO regression, which may lead us to incorrect selection of important variables with LASSO performed on the whole dataset.

The effect of imbalance can be mitigated if a strategy is adopted which randomly subsamples the dataset and iterates the LASSO several times. Majority voting is incorporated on γ values to select the nonzero variables in most of the iterations. Let us understand it with the help of a scenario. Assume that N randomly subsampled dataset is used to perform LASSO N times in which every instance contains equal examples of CHD or non-CHD. If we have 45 variables, we can get $\gamma_j = [\gamma_{j1}, \gamma_{j2}, \gamma_{j3}, \ldots \ldots, \gamma_{j45}]$ at a jth instance. We decide the inclusion of a variable v in further analysis by simply counting several nonzero instances of that variable and a manually set threshold value. Alternatively,

$$x(\gamma) \begin{cases} 0 & \text{if } \gamma = 0 \\ 1 & \text{otherwise} \end{cases}$$

$$[x(\gamma_{1,d}), x(\gamma_{2,d}), x(\gamma_{3,d}) \ldots \ldots x(\gamma_{N,d})]1 \geq \frac{m}{\alpha} \Rightarrow d \text{ is selected} \tag{2}$$

3.2 Convolutional Neural Networks (CNN)

Prediction of a CHD in a patient should be referred to as a binary classification task. While talking about supervised learning, the neural network has been proven to be an effective classifier under certain conditions [11]. In recent research, neural networks with application-specific settings like several hidden layers, have given significant improvement in several areas. The application areas of neural networks with significant results include image processing, speech processing, and time series prediction [42]. Relatively bigger datasets have been incorporated into rigorous training and fine-tuning of various deep learning architectures. The working of an artificial neural network consists of the transformation of input data over hidden layers and estimation of error at the output layer [43–45]. A gradient descent algorithm then uses the backpropagated error by the output layer for the iterative update of the layer weights. Several improvements in the gradient descent algorithm have been proposed by various experimentations and analyses including reduction in overfitting, scheduling the training process, making the layers nonlinear, visualization of the hidden layers, and other modifications. Despite the notable success in its applications, the working of deep neural networks is still insufficiently understood. It is also found that the training networks are easily overfitted owing to millions of parame-

ters in a deep architecture. The problem gets worse when the examples are insufficient.

Various algorithms have been proposed to solve this issue. Data augmentation [46, 47] is one of those widely used techniques that artificially populates new small datasets depending on the existing examples. Although this technique generates relatively better examples, however, such procedures are not credible while we are specifically talking about a biological application like clinical datasets. An example scenario can be the augmented measurements of a CHD phenotype, such as platelet count might not correspond to the possible range of readings of the patient. The reason behind this scenario is the fundamental difference between the foundations of platelet count readings and the principles of the statistical generation. Small or skewed datasets lead to poor training, which resultantly leads to poor and inaccurate classification. A wrong prediction in medical research accompanies a significantly bigger penalty as compared to other applications like semantic labeling, image synthesis, chat-bot configuration, etc.

Due to poor prediction strategy, a patient having CHD may be left untreated which may resultantly lead to wrong therapeutic medication. Because the accuracy of a prediction model is critical in medical application, so one of the prime objectives of this research work is to improve the accuracy of the classification i.e., predicting the presence or absence of CHD in a subject more accurately. We propose a shallow CNN to overcome these problems. As depicted in Fig. 1, every two fully connected layers of the mentioned shallow CNN have sandwiched the convolution layers.

3.3 Convolutional Neural Network Architecture

As depicted in the figure below, the architecture of the proposed convolutional neural network (CNN) is a feedforward network which works on a sequential single–input–single–output fashion. For binary classification experimentation, we assume that patients with the presence of CHD will be classified as '1', and others (with CHD absent) will be classified as '0'. An experiment multi-class classification is also performed will be discussed later. As mentioned earlier, the number of active CHD attributes (phenotypes) obtained from the majority voting algorithm is 14. Let us proceed with an assumption that the number of training examples is N, so the input layer indicated in Fig. 1 has $R^{N \times 14}$ dimension.

The 14 selected variables by LASSO-M voting and bias are combined in the fully connected 'Dense' layer with 64 neurons. This layer effectively normalizes various variable types before the nonlinear transformation, which is done by
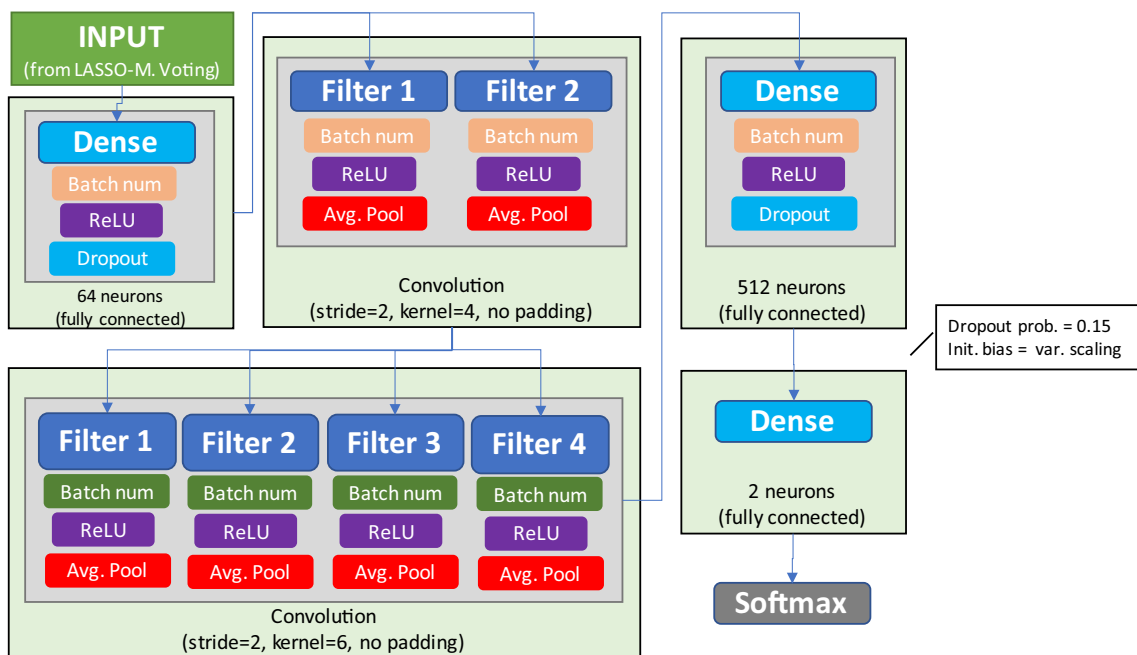
Fig. 1 Proposed CNN architecture

a rectified linear unit or ReLU. Overfitting is reduced by 15% dropout. A cascaded set of convolution layers is followed by the fully connected 'dense' layer. Two filters with kernel size 4 and stride 2 are introduced in the first convolutional layer. The external zero paddings are not provided in this layer. To find the best average pooling size under all constraints, various experiments with different pooling strategies are rigorously performed in the pooling layer. The output of the fully connected layer block $\in \mathcal{R}^{N \times 64}$ is converted into a tensor of $\mathcal{R}^{N \times 64 \times 1}$ dimension in the first convolution layer. The converted tensor is then further involved in various manipulation activities like nonlinear transformation, average pooling, and batch normalization to generate an output tensor of $\mathcal{R}^{N \times 31 \times 2}$ dimension.

The last convolutional layers with no zero paddings contain 4 filters that are taken with kernel width 6 and stride 2 to generate an output tensor of $\mathcal{R}^{N \times 13 \times 4}$. This output tensor is then delivered to average pooling layers, which forwarded it to the next dense layer after average pooling. The loss function is set as the categorical cross-entropy loss in the 'softmax' layer in which we can observe the categorical output as well. In each layer, we initialize bias with some random numbers that are drawn from a normal distribution with variance $\frac{1}{2\sqrt{n}}$, where n is the number of connections to the layer coming from the previous one. The Adam optimizer is used with $\partial_1 = 0.1$, $\partial_2 = 0.988$, learning rate = 0.005,

and decay = 0. Experiments with various hyperparameters associated with the proposed model are performed to obtain accurate consistency of classification. During the training phase, the results are provided by varying epochs, the number of neurons, class weights, and subsampling of input data. The said manipulation is done in each dense layer except the last one. Additionally, the number of filters also varies during the training phase.

3.4 Training Schedule

Although the dropout layers are used in the proposed CNN models, this training schedule is also used to improve the classification accuracy and further reduce overfitting. The concept of penalty helps the algorithm know its deficiency resultantly improve its working. Here, the class weight ratio is adjusted as a penalty because of imbalance in a class. We define it as a ratio to CHD and a Non-CHD dataset. For instance, a class weight ratio of 5:1 shows that we penalize a wrong classification of a CHD training sample 5 times more than a wrong classification on a Non-CHD sample during the calculation of error after each epoch before the backpropagation stage. This is done for initial training of the model with 1: N to a large number of epochs and then gradually increasing the weight ratio with a sudden decline in epochs.

Let us assume that the actual class weight is one, which is taken as a factor φ_0.

African countries. We incorporate only 14 attributes from this dataset to predict the presence of a CHD in a subject. Table

---

Algorithm 1. Fitting the proposed CNN model, D, with varying number of epochs (e) and weight ratio (φ)

1. Initialize $\varphi = N$, $e = N$, $D=1$, end $= 1$, $i=1$
2. While $\varphi \leq p_0$

　　　D.fit(inputData, e, φ)

　　　$\varphi \leftarrow$ floor $\left(\frac{\varphi}{2}\right)$

　　　$e \leftarrow$ floor $\left(\frac{e}{2}\right)$

3. While ($i \leq end$ and Train_loss(i) $\leq$ Train_loss($i - 1$)

　　　D.fit(inputData, i, φ_0)

Although we use dropout layers in the proposed CNN model. We also used this training schedule to further reduce possible overfitting. The intuition is to initially train the model with a 1: N weight ratio for a sufficiently large number of epochs and then, gradually increase the weight ratio with a steady decline in the number of epochs. Let the actual class weight ratio is ρ0: 1, which we take as a factor ρ0. Figure 1 is a pictorial view of the training layer, while Fig. 3 depicts the prediction layer of the proposed framework. We describe both the layers in detail in the following sub-sections.

The main focus of the training layer is to prepare a trained model that can be later used for predicting heart disease. Training of the learning model can be done by incorporating the output labels of the available training data. We use 30% of the rows for training purpose and the remaining 70% of the records shall be used for testing and validation purpose. As Fig. 1 shows, the training layer consists of three distinct steps named as leading and cleansing, model training, and storage of the trained model for later use. Below is a short description of each step. The prediction layer focuses on the testing of datasets for the prediction of potential heart disease. The remaining 70% of the dataset (without output labels) is provided to the training model for the prediction and testing step. The main focus of the training layer is to prepare a trained model that can be later used for predicting heart disease.

4 Experimental Results and Discussions

4.1 Dataset

In this section, we briefly describe the dataset incorporated in the experimental work of this study. As mentioned in previous sections, we make use of a state-of-the-art dataset especially available at [41] for this purpose. This set of attributes is a subset of a dataset compiled by medical practitioners in

1 depicts a list of attributes used in the algorithm with their short description and the possible range of values wherever applicable.

4.2 Experimental Setup, Results, and Discussions

This section elaborates on the experimental setup, underlying platform, along with experimental parameters with hardware and software platforms. We tried to correlate the features found in the dataset to predict the heart health of the patients. Interesting facts are found during the experimentation that is discussed in the subsections.

4.2.1 Experimental Setup

The experimental setup is established with the following parameters to observe and explore the performance of convolutional neural network (CNN) with the standard dataset available at UCI standard repository [41].

Hardware Requirements To carry out this research work and analyze the results, we establish an experimental environment on a personal computer. We equipped the experimental workstation with an Intel Quad-Core i7 4th generation processor, working at a clock rate of 2.3 GHz with an L1 cache of 32 KB, L2 of 256 KB and L3 cache memory with 4 MB of size. Sixteen gigabytes of DDR3 RAM are installed in the workstation and a total of a SATA hard disk having 1 TB capacity, rotating at 7 K RPM is installed.

Software Requirements This work incorporates Microsoft Windows 10 Pro as the base operating system. MATLAB version 2015a.

Table 1 List of incorporated attributes

Attribute	Description	Possible values
Age	Age (in years) when admitted to the hospital	Valid numbers
Sex		1: male/0: female
Cp	Type of chest pain (anginal pain)	1: typical 2: atypical 3: non-anginal 4: asymptomatic
trestbps	Blood pressure while resting (when admitted to hospital) in mm Hg	90–200
Chol	Level cholesterol in mg/dl	125–565
Fbs	Blood sugar (fasting) is fbs > 120 mg/dl?	1: true; 0: false
restecg	Results of electrocardiography while resting	0: normal 1: with ST-T 2: shows left ventricular hypertrophy (with Estes' criteria)
thalach	Max heart rate	71–202
exang	Angina induced by exercise	1 = true; 0 = false
old peak	ST induction of depression by exercise relative to rest	0–7
Slope	The slope of the peak exercise AST segment	1 depicts upsloping 2 depicts flat 3 shows a down-sloping
Ca	Number of major vessels (0–3) colored by fluoroscopy	0–3
Thal		3 = normal; 6 = fixed defect; 7 = reversable defect
Num	Finding for the existence of heart disease, status	0 corresponds to less than 50% narrowing diameter, 1 corresponds to greater than 50% narrowing diameter

4.2.2 Performance Metrics

One can evaluate the performance of the classifier by using some performance metrics. In machine learning, various criteria are available for evaluating the classifier's performance. Some of these criteria are explained as follows.

Precision Precision is the measure of exactness for evaluating the performance of a classifier. If the precision is high, it means there are fewer false positives. In a model with lower precision means, there are more false positives.

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{3}$$

Here in Eqs. (3) and (4), TP stands for true positive, whereas FP denotes false positive. The false positive will be denoted by FN in the subsequent subsection.

Recall It is a measure to determine the completeness of the classifier. Higher the recall, lesser the false negatives, and lower the recall, higher the false negatives. Improvement in recall often results in a decrease in precision.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

F-Score The combination of accuracy and recall is called F-score which can be calculated by the following formula:

$$F1\ \text{Score} = \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{5}$$

As depicted in Eq. (5), the F1 score can be obtained by dividing the product of recall and precision with the sum of recall and precision.

4.2.3 Data Preparation

As already mentioned in the previous section, the preparation of datasets according to the nature of experimentation to be performed is necessary before initiating the experiments. Data preparation step duly involves fetching the dataset into the file system by a third-party program or is directly saved into the file system with the help of a Python Library named SKLearn.

Table 2 Performance of the binary classifier

Performance metric	Value
Precision	0.9706
Recall	0.9635
F1-score	0.9670
Accuracy	0.9700

Table 3 Performance overview of the model with four classes

Performance metric	Value
Precision	0.8669
Recall	0.8174
F1-score	0.8414
Accuracy	0.8667

A noticeable amount of noise and some inconsistencies were observed in the dataset. This required the data to be cleansed and made consistent, which is done by writing various programs in Python programming language. Various programs are written for different tasks like data cleansing, replacement of anomalies, calculation of mean values and normalizers, etc. to translate the data into a cleansed feature matrix. These steps are already discussed in the Proposed Framework section of the previous section.

4.2.4 Data Classification Using CNN

Heart Disease Classification Once the dataset is presented in the form of a feature matrix, the next step is to distribute this dataset into different classes. In our first case, we classified the dataset into two classes i.e., Heart_Disease and No_Heart_Disease. A classifier model is then generated by using CNN to classify the data into binary classes and the model is stored on the file system. The computed accuracy for binary the classification is 97% as depicted in Fig. 2.

Table 2 depicts the performance of the proposed model concerning the precision, recall, F1 score, and accuracy. The overall accuracy of the proposed model is 97% which is seen in Table 2. Another classifier was trained for four classes as well, representing the types of heart disease present in the stored dataset. A program written in Python programming language was executed to determine the number of existent types. The program searched the data and classified the dataset into 4 classes i.e., Type 1 Disease, Type 2 Disease, Type 3 Disease, Type 4 Disease. After determining the number of classes from the dataset, the job automatically labeled the records into the nearest related type of class. The trained model is tested for 4 classes, and the results were represented with the help of the confusion matrix, as predicted in Fig. 4. The predicted accuracy of 4 classes is 87% which is higher than the results presented in [37]. The confusion matrix presented in Fig. 4 shows that a total of 114 records were classified into type 1 class. Similarly, 77 records were classified into type 2 class, and a total of 22 records were classified into type 3 class (Fig. 3).

Additionally, a total of 47 records are classified into type 4 class. The total false classified records for type 1 disease are 8, for type 2 disease, a total of 26 records were false classified. The false classification concerning type 3 disease is 6, whereas type 4 disease has no false classifications. The values of precision, recall, F1-score, and accuracy of 4 classes are shown in Table 3:

4.2.5 Correlation Among Independent Variables

As discussed earlier, variable selection is an important step in our classification model preparation. We started by investigating of correlation among 14 predictor variables with continuous values. A high correlation (0.58) has been found between the *old peak* and *slope* (description of the variables is given in Table 1) as depicted in Fig. 4). However, it is reported in the literature that AST (slope) has a significantly higher value in patients with CHD, so it can be used as a good predictor or biological marker to predict the severity of a CHD in a patient. A high correlation (0.50) has been found between *old peak* and num (percentage of narrowing in diameter) meaning that increase in ST while exercise indicates that the patient may have 50% or more narrow in diameter of arteries which seems quite logical.

Talking about narrowing in diameters of arteries, a positive correlation coefficient (0.52) between *Thal* and *num* is also observed, which shows that these two parameters are positively related to each other. Meaning that higher *Thal* values increase the probability of more narrow arteries in the subject. It is also observed that there is a negative correlation (−0.33) between *Cp* (chest pain) and *thalach* (max heart rate). A positive correlation (0.28) has been found between *trestbps* (blood pressure) and age, depicting that the blood pressure of a CHD may increase with the increase in age. A positive correlation (0.17) between *chol* (cholesterol level) and *restecg* (the result of electrocardiography) is found, which means that CHD patients with higher cholesterol values may have abnormal ECG values as well. Because of their importance of incorrectly determining the predictor variables, we performed LASSO regression because of the significance and association of risk factors with CHD. Because of the significance of risk factors and their association with correlated variables, while predicting a CHD, we performed LASSO regression to correctly identify the predictor variables for further analysis.

4.2.6 Comparison with Other Techniques

As the proposed framework suggests, the trained classifier is then used for the prediction of the whole testing dataset.

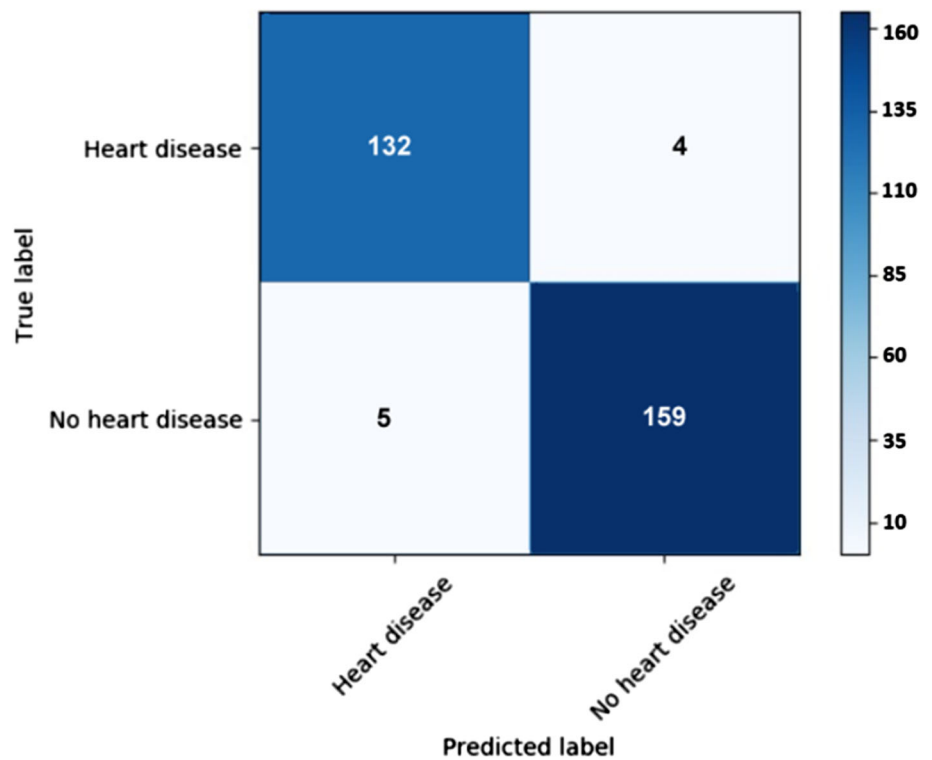**Fig. 2** Confusion matrix for binary classification

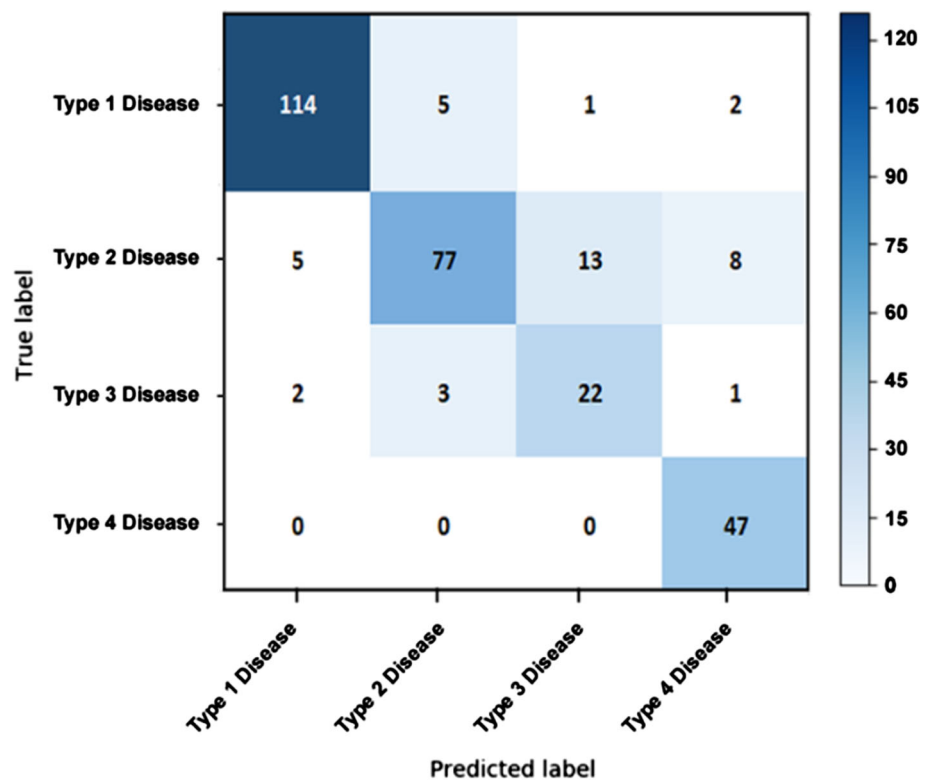**Fig. 3** Confusion matrix for multi-class classification

Table 4 and corresponding Fig. 5 depict a performance comparison between the proposed framework and other works concerning the accuracy outcomes. It can be observed from the said figure that the proposed approach outperforms the other mentioned techniques concerning performance i.e., the overall accuracy. It can also be observed that the proposed work is followed by the decision tree-based approach proposed in [25].

🖄 Springer

Fig. 4 Correlation matrix among predictor variables

	Age	Sex	Cp	trestbps	Chol	Fbs	restecg	thalach	exang	oldpeak	Slope	Ca	Thal	Num
Age	1.00	-0.10	0.10	0.28	0.21	0.12	0.15	-0.39	0.09	0.20	0.17	0.36	0.13	0.22
Sex	-0.10	1.00	0.01	-0.06	-0.20	0.05	0.02	-0.05	0.15	0.10	0.03	0.09	0.38	0.22
Cp	0.10	0.01	1.00	-0.04	0.07	-0.04	0.07	-0.33	0.38	0.20	0.15	0.23	0.25	0.41
trestbps	0.28	-0.06	-0.04	1.00	0.13	0.18	0.15	-0.05	0.06	0.19	0.11	0.10	0.13	0.16
Chol	0.21	-0.20	0.07	0.13	1.00	0.01	0.17	0.00	0.06	0.05	0.00	0.12	0.03	0.07
Fbs	0.12	0.05	-0.04	0.18	0.01	1.00	0.07	-0.01	0.03	0.01	0.06	0.14	0.05	0.06
restecg	0.15	0.02	0.07	0.15	0.17	0.07	1.00	-0.08	0.08	0.11	0.14	0.13	0.02	0.18
thalach	-0.39	-0.05	-0.33	-0.05	0.00	-0.01	-0.08	1.00	-0.38	-0.34	-0.39	-0.26	-0.26	-0.42
exang	0.09	0.15	0.38	0.06	0.06	0.03	0.08	-0.38	1.00	0.29	0.26	0.14	0.32	0.40
oldpeak	0.20	0.10	0.20	0.19	0.05	0.01	0.11	-0.34	0.29	1.00	0.58	0.30	0.34	0.50
Slope	0.17	0.03	0.15	0.11	0.00	0.06	0.14	-0.39	0.26	0.58	1.00	0.10	0.29	0.38
Ca	0.36	0.09	0.23	0.10	0.12	0.14	0.13	-0.26	0.14	0.30	0.10	1.00	0.26	0.52
Thal	0.13	0.38	0.25	0.13	0.03	0.05	0.02	-0.26	0.32	0.34	0.29	0.26	1.00	0.50
Num	0.22	0.22	0.41	0.16	0.07	0.06	0.18	-0.42	0.40	0.50	0.38	0.52	0.50	1.00

Fig. 5 Performance comparison in terms of accuracy
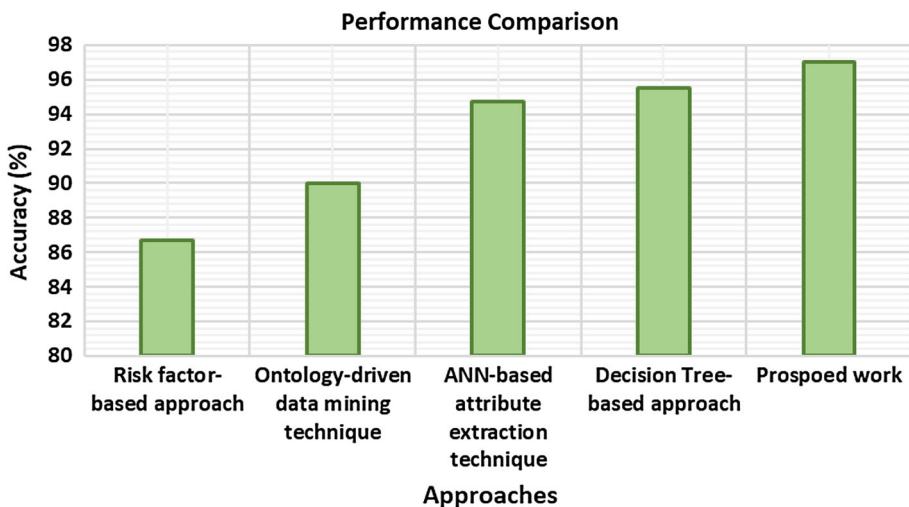
5 Conclusion and Future Work

Prediction of heart diseases at earlier stages may prevent possible deaths due to heart attacks. A good classification algorithm may help the physician predict the presence of cardiovascular disease before its actual occurrence. This research focuses on predicting a possible heart disease by incorporating a state-of-the-art dataset available at UCI repository, and convolutional neural networks (CNN). This dataset consists of some cardiac test parameters as well as general human habits. The results show that the proposed model outperforms the existing techniques referred to in this paper. The overall accuracy of the proposed model is 97%. In the future, we aim to further enhance this research work by

Table 4 Performance comparison of the proposed model with existing methods

Authors/technique	Accuracy
Kaanchan More et al. [21] (Risk factor-based approach)	86.7%
Qrenawi, Mohammed et al. [18] (Rmonto ontology-driven data mining approach)	90.0%
Shahid Mehmood et al. [24] (ANN-based attribute extraction technique)	94.7%
Alizadeh-dizaj et al. [25] (Decision tree-based approach)	95.5%
Proposed approach	97.0%

predicting the occurrence of other major diseases like cancer and other brain-related diseases.

Authors Contributions All the authors contributed equally to this work.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

References

1. Ueshima, H.; et al.: Cardiovascular disease and risk factors in Asia: a selected review. *Circulation* **118**(25), 2702–2709, 2008. https://doi.org/10.1161/CIRCULATIONAHA.108.790048.
2. Bewu, A.; Mbanya, J.C.: Cardiovascular Disease. In: Jamison, D.T;, Feachem, R.G.; Makgoba, M.W.; et al. editors. Disease and Mortality in Sub-Saharan Africa. 2nd edition. Washington (DC): The International Bank for Reconstruction and Development/The World Bank; 2006. Chapter 21. https://www.ncbi.nlm.nih.gov/books/NBK2294/ (2006)
3. Ramirez-Bautista, J.A.; Hernández-Zavala, A.; Chaparro-Cárdenas, S.L.; Huerta-Ruelas, J.A.: Review on plantar data analysis for disease diagnosis. *Biocybern. Biomed. Eng.* **38**, 342–361, 2018
4. Nalluri, S.; Saraswathi, R.V.; Ramasubbareddy, S.; Govinda, K.; Swetha, E.: Chronic heart disease prediction using data mining techniques. In: Data Engineering and Communication Technology, ed: Springer, pp. 903–912 (2020)
5. (2017, 25 March). WHO: World Health Organization, Media Centre, cardiovascular diseases fact sheet webpage. https://www.who.int/mediacentre/factsheets/fs317/en/ (2017)
6. Sharma, R.; Agarwal, P.; Mahapatra, R.P.: Evolution in big data analytics on internet of things: applications and future plan. In: Multimedia Big Data Computing for IoT Applications, ed: Springer, pp. 453–477 (2020)
7. Gartner: The Importance of 'Big Data': A Definition. https://www.gartner.com/doc/2057415/importance-big-data-definition (2012)
8. Jordan, M.I.; Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. *Science* **349**, 255–260, 2015

9. Gupta, R.; Khari, M.; Gupta, D.; Crespo, R.G.: Fingerprint image enhancement and reconstruction using the orientation and phase reconstruction. Information Sciences (2020)
10. Gupta, R.; Khari, M.; Gupta, V.; Verdú, E.; Wu, X.; Herrera-Viedma, E.; González Crespo, R.: Fast single image haze removal method for inhomogeneous environment using variable scattering coefficient. *Comput. Model. Eng. Sci.* **123**(3), 1175–1192, 2020
11. LeCun, Y.; Bengio, Y.; Hinton, G.: Deep learning. *Nature* **521**, 436–444, 2015
12. Khari, M.; Garg, A.K.; Crespo, R.G.; Verdú, E.: Gesture recognition of RGB and RGB-D static images using convolutional neural networks. Int. J. Interact. Multimed. Artif. Intell. 5(7) (2019)
13. Robinson, Y.H.; Vimal, S.; Khari, M.; Hernández, F.C.L.; Crespo, R.G.: Tree-based convolutional neural networks for object classification in segmented satellite images. Int. J. High Perform. Comput. Appl. 1094342020945026 (2020)
14. Dua, M.; Gupta, R.; Khari, M.; Crespo, R.G.: Biometric iris recognition using radial basis function neural network. *Soft. Comput.* **23**(22), 11801–11815, 2019
15. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; et al.: A guide to deep learning in healthcare. *Nat. Med.* **25**, 24–29, 2019
16. Rathnayakc, B.S.S.; Ganegoda, G.U.: Heart diseases prediction with data mining and neural network techniques. In: 2018 3rd International Conference for Convergence in Technology (I2CT), pp. 1–6 (2018)
17. Van Pham, H.: A proposal of expert system using deep learning neural networks and fuzzy rules for diagnosing heart disease. In: Frontiers in Intelligent Computing: Theory and Applications, ed: Springer, pp. 189–198 (2020)
18. Qrenawi, M.I.; Al Sarraj, W.: Identification of cardiovascular diseases risk factors among diabetes patients using ontological data mining techniques. In: 2018 International Conference on Promising Electronic Technologies (ICPET), pp. 129–134 (2018)
19. Nguyen, T.-H.; Nguyen, T.-N.; Nguyen, T.-T.: A deep learning framework for heart disease classification in an IoTs-based system. In: A Handbook of Internet of Things in Biomedical and Cyber Physical System, ed: Springer, pp. 217–244 (2020)
20. Isra'a Ahmed Zriqat, A.M.; Altamimi, M.A.: A comparative study for predicting heart diseases using data mining classification methods
21. More, K.; Raihan, M.; More, A.; Padule, S.; Mondal, S.: A12176 Smart phone based "heart attack" risk prediction; innovation of clinical and social approach for preventive cardiac health. *J. Hypertens.* **36**, e321, 2018
22. Aldallal, A.; Al-Moosa, A.A.A.: Using data mining techniques to predict diabetes and heart diseases
23. O'Donnell, J.; Velardo, C.; Shah, S.A.; Khorshidi, G.S.; Salvi, D.; Rahimi, K.; et al.: Physical activity and sleep analysis of heart failure patients using multi-sensor patches. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6092–6095 (2018)
24. Awan, S.M.; Riaz, M.U.; Khan, A.G.: Prediction of heart disease using artificial neural network. *VFAST Trans. Softw. Eng.* **13**, 102–112, 2018
25. Alizadeh-dizaj, G.: Risk Prediction and Stratification of Patients with Stroke Using Data Mining Techniques. Tabriz University of Medical Sciences, School of Management and Medical Informatics (2018)
26. Somasundaram, T.J.P.K.: An empirical study on prediction of heart disease using classification data mining techniques, presented at the IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), (2015)
27. Liu, Y.: Environmental Health Nexus: Designing Predictive Models for Improving Public Health Interventions (2018)

28. Deepika, M.; Kalaiselvi, K.: A empirical study on disease diagnosis using data mining techniques. In: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 615–620 (2018)

29. Malav, A.; Kadam, K.: A Hybrid Approach for Heart Disease Prediction Using Artificial Neural Network and K-means

30. Wankhar, I.; Wriang, I.; Debnath, P.; Bordoloi, P.; Tripathi, R.; Bhatia, D.; et al.: Comparison and Performance Evaluation of ECG Classification Techniques Trained with Shorter Database, ed: IJSRST, (2018)

31. Ordonez, C.: Association rule discovery with the train and test approach for heart disease prediction. *IEEE Trans. Inf. Technol. Biomed.* **10**, 334–343, 2006

32. Xing, Y.; Wang, J.; Zhao, Z.: Combination data mining methods with new medical data to predicting outcome of coronary heart disease. In: 2007 International Conference on Convergence Information Technology (ICCIT 2007), pp. 868–872 (2007)

33. Srinivas, K.; Rao, G.R.; Govardhan, A.: Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. In: 2010 5th International Conference on Computer Science & Education, pp. 1344–1349 (2010)

34. Liu, J.-L.; Hsu, Y.-T.; Hung, C.-L.: Development of evolutionary data mining algorithms and their applications to cardiac disease diagnosis. *IEEE Congress on Evolutionary Computation* **2012**, 1–8, 2012

35. Chandra, P.; Deekshatulu, B.: Prediction of risk score for heart disease using associative classification and hybrid feature subset selection. In: 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 628–634 (2012)

36. Amin, S.U.; Agarwal, K.; Beg, R.: Genetic neural network based data mining in prediction of heart disease using risk factors. In: 2013 IEEE Conference on Information & Communication Technologies, pp. 1227–1231 (2013)

37. Maji, S.; Arora, S.: Decision Tree Algorithms for Prediction of Heart Disease. Information and Communication Technology for Competitive Strategies, pp. 447–454. Springer, Singapore (2019)

38. Alim, M.A.; et al.: Robust heart disease prediction: a novel approach based on significant feature and ensemble learning model. In: 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). IEEE, (2020)

39. Mienye, I.D.; Yanxia, S.; Zenghui, W.: Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. Informatics in Medicine Unlocked, 100307 (2020)

40. Dwivedi, A.K.: Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Comput. Appl.* **29**(10), 685–693, 2018

41. https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/

42. LeCun, Y.; Bengio, Y.: Convolutional networks for images, speech, and time series. *Handbook Brain Theory Neural Netw.* **3361**, 1995, 1995

43. Iandola, F.: Exploring the design space of deep convolutional neural networks at large scale, arXiv preprint arXiv:1612.06519 (2016)

44. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence, (2017)

45. He, K.; Zhang, X.; Ren, S.; Sun, J.: Deep residual learning for image recognition. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, ed: IEEE (2016)

46. Krizhevsky, A.; Sutskever, I.; Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)

47. Radford, A.; Metz, L.; Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)