Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

Early Prediction of Heart Conditions by K-Means Consensus Clustering and Convolution Neural Network

[1,*]SaiyedFaiayazWaris, [2]S.Koteeswaran
[1,*]Research Scholar, Department of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, India
[2]Associate Professor, Dept. of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, India
***Corresponding Author:** saiyed.cse@gmail.com

Abstract
Today, predicting heart disease is the most challenging task in the medical world.In this generation, due to complications of heart disease, a person dies each minute or so.Data science plays a significant role in collecting large amounts of data in the medical domain.Since the prediction of heart disease is a complex task, it is necessary to anticipate earlier to avoid risks and inform the patient in advance.Machine learning, data mining, and deep learning are the three methods that provide the methodology and technology to leverage the relevant knowledge to make the right decision.Heart and vascular diseases are associated with cardiovascular diseases including Heart Disease (HD).Extensive research into heart disease has been carried out using vast amounts of data on all other aspects of HD.This paper, which is a cardiac dataset, comes out of the available public health dataset.In this article, we proposed a new Conventional Neural Network (CNN) architecture that was used to extract and categorize histopathological images using the K-means Consensus Clustering.The use of the deep neural network, the selection, and recovery of functionalities is the important process carried out before the classification of the dataset.In consequence, the dataset is pre-formed by the training process.The proposed model is associated with temporal data modeling use the earlier CNN prediction of HD.We achieved good results with the cardiac dataset compared to the existing results.The outcome of the proposed work achieves a precision rate of 97%.

Keywords: Conventional Neural Network, Cardio Vascular Disease,Electronic Health Record, Heart Disease, K-means Consensus Clustering.

1 Introduction

Coronary artery disease, also called HD is the most common type and it creates when courses that deliver blood to the heart become stopped with plaque. It makes them solid and restricted. It is filled with cholesterol and a variety of substances. As a result, blood flow is reduced, and the heart gets less oxygen and supplementation. As predicted, the cardiac muscle has weakened the risk of cardiovascular degradation and arrhythmia. As plaque develops in the arteries, it is called atherosclerosis [1].

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

Cardiovascular disease (CVD) [2] is a general term for heart or vein-related conditions. It generally relates to the development of fat reserves in corridors (atherosclerosis) leads to the increased danger of blood blocks. It may also be associated with damage to hallways in organs such as the mind, heart, kidneys, and eyes. Cardiovascular disease is an important driver of death and disability in the United Kingdom. However, it can routinely to a large extent by the conduct of a healthy lifestyle. The development of greasy plates in the conduits, or can harm veins and the heart. Plaque development results in restricted or blocked veins that trigger coronary, thoracic torment (angina) lead to stroke[3].

The symptoms of coronary heart disease can be distinct to people.For example, men suffer inevitable torments on their chests.Women have various signs and manifestations next to chest discomforts, such as nausea and extraordinary exhaustion[4].The medical services industry collects a lot of information on medical services used to locate hidden data for a powerful dynamic.Inspired by the overall increase in deaths of patients suffering from coronary heart disease each year.And the accessibility of enormous measure of patients' information from which to extricate valuable information, analysts have been utilizing information mining strategies to help medical services experts find coronary illness[5].

A healthcare environment that uses smart clothing to observe practical well-being. The heterogeneous frameworks and achieved the best results for minimizing costs. Measurable patient data, test results, and disease history are recorded in the Electronic Health Record (EHR)[6], allowing in identifying the potential information-based responses to reduce the cost of clinical contextual analyses. The heart is a powerful organ that generally consists of cardiovascular muscle. The heart drains the oxygen-rich blood through the body's construction and expansion. Cardiovascular diseases help in managing HD from venous problems[7].

As per the World Health Organization (WHO), 17 million individuals died every year because of Cardiovascular sickness, which represents 31% of the passings worldwide. It attempts to distinguish between cardiovascular infections based on many factors [8].Critical variables that cause cardiovascular infection are high blood pressure, higher lipids in the blood, stress, weight, metabolic disorders, diabetes. In some cases, family and ancestry may also be a factor in illness[9].Certain tests are done before the determination of cardiovascular diseases, such as Electrocardiogram (ECG), pulse, cholesterol, glucose, and auscultation.Artificial intelligence (AI) is used in various domains to respond to complex problems.The AI provides an efficient method for diagnosing disease.AI provides a variety of ways to find secret examples in similarities of information.Since cardiovascular infections are extremely confusing and so care should be taken during treatment [10].

The AI cluster is to localize cardiovascular illness.Irregular sorter to predict heart disease.False neural organizations help to analyze the disease such as cardiovascular infections, malignancies,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

and brain tumors.The model presented in this paper uses 13 credits of the standard coronary artery disease dataset to predict coronary artery disease.In this work, the proposed model provides that CVD using AI approaches.The proposed model is based on weighted Linear Regression dem(LR)[11], K-Nearest Neighbour (KNN)[12], Random Forest (RF)[13], Neural Network activated with ReLU function (NNR)[14], and Gaussian Naive Bayes (GNB)[15].The main expectation is to build a model with improved performance in diagnostics of cardiovascular infection.The proposed design of this work achieved an accuracy of 89% and an accuracy of 91.6%.Compared to other available models, test results from the proposed model capable of anticipating CVD [16].

CVD is a major killer at this point.In 2016, the majority (54%) of deaths were caused by coronary heart disease.Numerous machine learning approaches are inadequate in measuring infectious diseases.As a result, it is necessary to have a framework that provides for infection effectively[17].The Deep Learning approach predicts blocked heart disease proposes that CNN anticipate the disease at an early stage.Cluster review is how objects are grouped into subsets that are significant for a specific problem.Following this, the items are coordinated to describe the inspected population[18].Contrary to the schema, clustering does not depend on predefined classes.Clustering is referred to as a spontaneous learning strategy because no data is provided on the “right answer” for either element.It can find previously unidentified connections in a perplexed informational index.Several cluster review requests are available.For example, in an operational application, the group review may be used to identify and describe customer groups for promotional purposes.In the non-hierarchical grouping, for example, the computation of the k-mean, the connection between the clusters is uncertain.Progressive grouping connects several groups until each piece of information is stored in sequence.These methodologies are to decide the resemblance between two items so that groups can be framed from objects with high comparability to each other.Usually, distance measurements like Manhattan and Euclidean, are used to decide on similarity.Distance work provides more incentive for sets of elements that are less similar to each other.From time to time, approach work is generally used, resulting in superior qualities for more comparable sets [19].

2 Related Work

This study sets up a competent neuronal organization with convolutional layers to characterize clinical information in a fundamentally unbalanced class. While most of the current AI models that have been used in this class of information are helpless against class irregularity even after the change of explicit class loads. The basic two-layer CNN displays the force to inequality with reasonable concordance in the explicit performance of classes.

Item selection technique which creates a small subset without changing the information.A system to determine compound elements.Highlighted selection will produce a subset of information.Many highlights are not useful for AI because it produces complex, competent prediction, and selection. A method of determining elements is a cross-strategy that has a global

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

population struggling with this problem.In examining the mortality rate and countless people with coronary heart disease, we recognize the importance of early CVD findings.Many conventional techniques exist to predict such a disease, but they seem inadequate.There is a need for a clinical determination framework that can foresee the heart analysis at a beginning phase and offers a more exact conclusion than conventional techniques such as LR, Lasso, Elastic Net, and Group Lasso regularization. These days, AI approaches are becoming very important.

The EMR provides a range of important clinical data for patients. For the situation where CVD proves to be more worldwide presaging. Many experts have focused on EMR-dependent cardiovascular diseases to address this aspect of improving the efficacy [20]. The purpose of clustering is to find a single segment consistent with current fundamental models, to the extent reasonably expected. Clustering is perceived as producing strong segments, finding unusual groups, handling clams, exceptions, testing varieties, and incorporating arrangements from various disseminated or deficient information sources. Here, the major parts may be produced a similar grouping calculation with various limits, or by a similar grouping calculation with various salient points [21].

The highlights are classified as a strong, weak, poorly secured component.Given that element, the requirement has been expressed.Agreement groupings expect to find a solitary pattern that is in agreement. It has been broadly perceived that agreement clustering is successful to create strong grouping results, recognize odd groups, handle commotion, anomalies, and test varieties, and coordinate arrangements from various conveyed wellsprings of information or qualities. Not quite like the usual bundling techniques, which direct the information network, bundling contribution agreements is the arrangement of different matching essential parts.So the grouping of agreements is a combined issue, rather than a customary grouping issue. This proposal deal with the complex grouping of agreements by transforming them into another simple issue. As a rule, K-means-based Consensus Clustering (KCC) proposed, which precisely changes the agreement grouping issue into a K-means clustering issue with hypothetical backings, and give them adequate and essential state of KCC utility capacities [22,23].

3 Proposed Methodology

CNN architecture and extractive function with deep neural networks used to predict heart conditions.The main goal of this research is to predict heart disease accurately.The patient data set is derived from the Report on Public Health Care. Then the data are given into a model that makes a prediction of the HD of a patient with the help of the pre-recorded data and makes the patient alert us if they have a probability of having HD in the future. Figure 1 demonstrates the whole process associated with this model.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

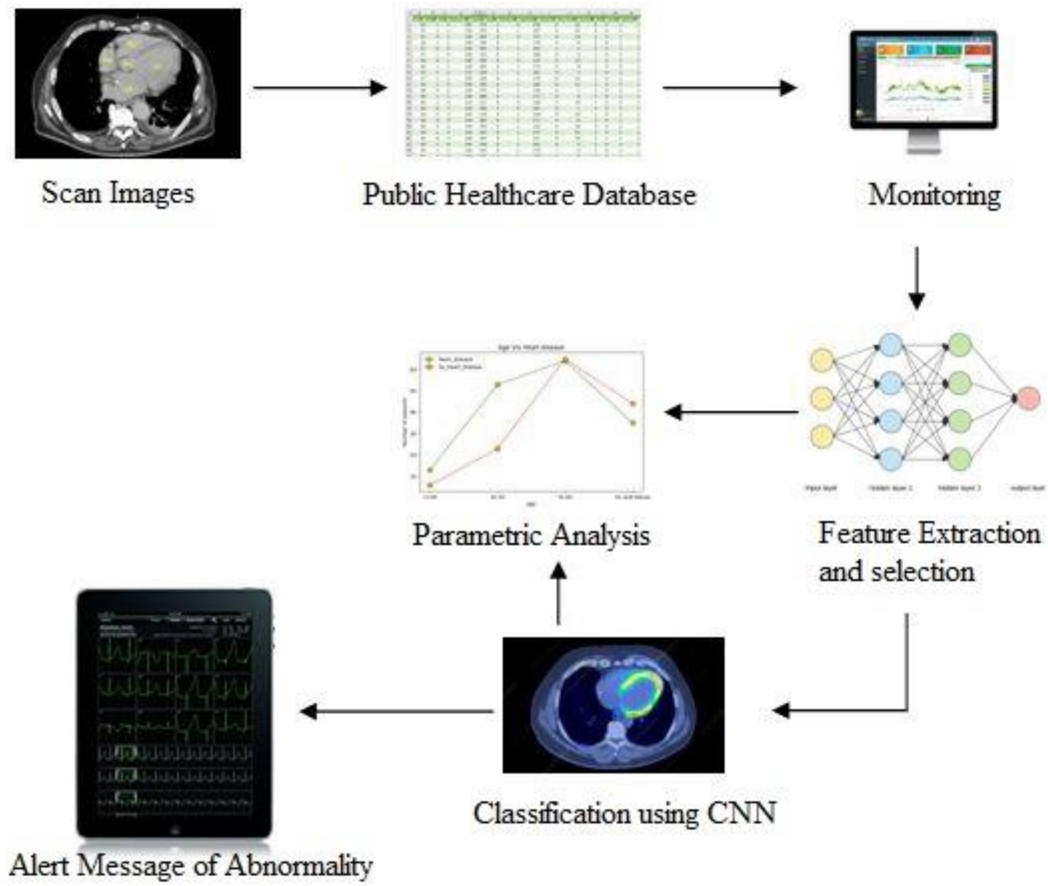

Figure 1: Proposed Architecture

First, pre-recorded medical information is collected from normal people and patients with cardiac complications. In the public healthcare database, the recorded data shall be retained. Histopathological images are collected using the surveillance system and the extraction and selection of features are also performed using CNN. In the overall process, the data is pre-formed and has been intended to be normal and abnormal data. The risk of heart disease is predicted with the help of the CNN architecture classification technique, and the probability of getting the risk of heart disease at the earliest stage.

This paper proposed two different types of neural network models that have a global learning framework, in which the binary classification task is mainly performed by basic deep neural networks. The KCC model is the first architecture model proposed in Figure 1. The NN model comprises three separate blocks of convolutional layers. A method of abandonment is used to avoid the overflow problem and removing random inputs into the training process can lead to the loss function.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

3.1 Feature Extraction and Selection

The selection of characteristics is very important to the exact prediction. The data set with the desired characteristics is taken and used by the data scanning algorithms to select the desired characteristics. For preprocessing, the selected functions or attributes are directly loaded into memory. A subset of own characteristics is selected through a certain process called the characteristic selection process. And the methodology is used to identify the most important attributes and help predict the result. The dimensional reduction form is used for preprocessing. Reducing attributes without changing the data set is the main difference between selecting characteristics and reducing dimensionality (selecting characteristics). And also the reduced complexity of the feature selection method is only due to the attribution with fewer parameters.

In AI, the provision refers to a premonitory display problem in which a class name is provided for a particular illustration of the information.Examples of arrangement issues include: Given a model, characterize in case it is spam or not.Considering a handwritten character, group it as one of the familiar characters.Scores from different statistical correlations are used by filtering methods for feature selection. The wrapper method has a different kind of approach in the selection of the feature and this method evaluates all possible combinations and gives an accurate result for the ML. The integrated method produces the result with a combination of the above two models. Various statistical methods are used in selecting machine learning functions. In general, machine learning with a large dataset uses the filter method when selecting features. The classification algorithm is used as a wrapper method to measure performance. As well, the screening method uses the scientific approach rather than the classification algorithm.

Raw data transmitted through the system may be under any form of numerical data. The pre-processed data are received using the appropriate techniques. Pre-processing of information is a cycle in which raw information is established and made reasonable for an AI model. It is the first and decisive breakthrough while doing a model of artificial intelligence.
When carrying out an AI project, it is not usually a case for us to confess all the information and arrange it.And knowing that doing any activity with information, it is mandatory to clean it up and set it up in an organized way.This is done through the pre-processing task.For the test process, the chosen characteristics and the preprocessed data are passed through the classifier.When the accuracy of the classification is improved, the quality of the extracted features is improved as well.The generic nature of the characteristic retrieval process is verified by testing the system with different types of data sets as shown in Figure 2.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

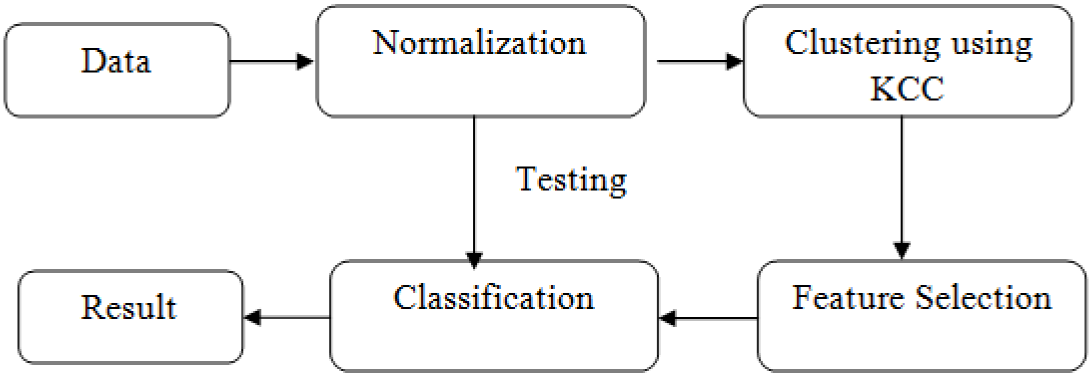

Figure 2: Feature Extraction and Selection Using KCC

3.2 Preprocessing Data

A numeric form of data in the conversion of all attributes is performed in the pre-processing stage which is used by the clustering process. And it's very helpful for the data set reduction dimension. The other way of preprocessing known is normalization. And sometimes it may occur in which some attribute values may get varied to certain ranges, so that its effects in reduction of such attributes, and if it lies in some normal range, then the values of attributes are considered to be normalized.

3.3 Normalization Model

In mining algorithms, a device that is used to pre-arrange data or information is called normalization. By measuring the features of dataset qualities, it is said to be standardized and then the dataset gets fallen inside a less chosen run, e.g., 0.0 to 1.0. Standardization is particularly useful for calculations of social opportunity that includes neural systems, or separation estimates, for example, the solicitation and packaging of nearest neighbors. The primary core dataset which has the right change is called Min-Maximize normalization and it is mapped a value of P to d ranges from [newmin(p), newmax(p)]. The various key values of the core dataset and its relations are also protected by Min-max normalization.

3.4 Cluster Data

The clustering is a necessary precursor for the extraction of the feature. The pre-processed data are input to the function extraction, in which the analyzed labels are used.Clustering is the certain points without knowing the possible attributes label points with other points of clustering which makes the other points similar.Where clusters overlap, the K-mean technique is used and it must be shown that the actual datasets are very common, sometimes also in the usual scenario.A Consensus Clustering algorithm based on K-means is explained in its description.

Algorithm for KCC:

Algorithm KCC ()

{

Initialising the matrix with attributes and objects for the row and column.

Initialize cluster c= 1,2,...,n

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

group matrix g=0
while (g(i)<=n)
{
new consensus clustering coordination
obtain distance for every node
}
}

The membership matrix is updated when cluster centers are calculated and the location of cluster centers is allocated accordingly. The new value of a point is calculated by selecting one specific cluster, the distance between that point from the specified cluster center, and also the distance between that point and all other cluster centersare taken for the calculation. Then we calculate the change in the centroid matrix.The process is stopped, if the calculated value is lower than a predefined threshold, if that is not the case, other cluster centers are updated by the centroid matrix. The membership matrix is to be minimized to avoid continuous iteration.

3.5 Extract Features

The feature extraction is the important process that builds a generic data mining system that is given to various data sets. Only the values of the attributes are considered and the nature of the attributes is not affected by this process. It is clear evidence from the FCM, the k – dimensional space gives the cluster centers, and the attributes are referred to as n, the cluster centers are independent of the natures of the attributes. Consider k attributes from an n-class problem.

The mathematical form of the neural network input is given in Eq. (1). Cross-entropy is considered with the softmax function and the cost function in the output is dependent on the layers.

The adjustment activation function equation is:

$$f(xc) = xc + \max(0, xc) \qquad (1)$$

Where xc is input from the neural network.

The rectifier unit gives the analytical function of the smooth approximation which are tracked in eq 2:

$$f(xc) = \ln[1 + \exp(xc)] \qquad (2)$$

The prediction is made, the new data extracted from the concealed layer are given in eq 3:

$$Xc1 + 1 = H(Wc1Xc1 + A1), 1 = 1,2, \ldots\ldots, n \qquad (3)$$

Wherethe activation function is H, the weight matrix of the cluster is W_c, and A_l is the bias hidden layer and the rectified linear unit (ReLu) is used for the above parameter selection. The input training or test dataset does not provide input directly to the hidden unit.

In the covert layers, the neurons of the t-layer are interconnected to form the neurons of the t-1-layer. Each of the hidden layers generates an output as a function of Eq 4.

The hidden layer of interconnected neurons with the n-1 layer that produces the result which is given in the form of:

$$xc_{i+1} = \sigma\left(\sum_{i=1}^{n}(wc_i xc_i + A_i)\right) \qquad (4)$$

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

Where the bias variable and the weight of the cluster are denoted as wc and A, the input layer of the cluster is denoted as xc and the n is the hidden layer of the neural network. There are ReLU and hidden node layers are used in the network.The function ReLU, f (xc) = maximum (0, xc).To enable the output layer, the sigmoidal function is used.For the loss function, the Mean Square Error (MSE) and to optimize the MSE loss, the Adadelta algorithm is used respectively.

Algorithm for Feature Extraction using KCC

algorithmfeature_extract_kcc()
begin
initialize several clusters as n; the number of attributes as k;
for i=1 to k
for j=1 to n-1
dist=j-i;
end for
min_dist=min(n);
end for
end

3.6 Classification Using Convolutional Neural Network

With the use of the image enhancement technique, the model is refined to obtain CNN models from scratch. Finally, image classification and verification of training and validation data are carried out by the dummy model.Figure 3 demonstrates the implementation of the CNN architecture.

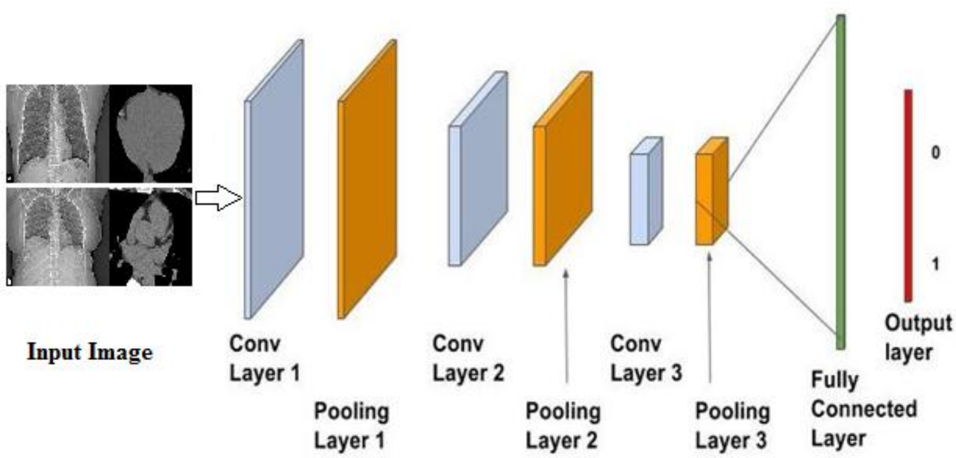

Figure 3: Implementation Architecture of CNN

CNN is a kind of artificial neural network, which is used for the multiple perceptrons which look at the input images.The CNN has the weights and basics that can be learned in many parts of the pictures and that can be separated from each other.It has the advantage of investing the use of local spatial consistency in the input images, few parameters are shared to reach fewer

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

heights.The above process certainly has memory and effective complexity.The following layers are used in the CNN:

Convolution Layer – In this layer, to create the map of the characteristics of the succeeding layer, a kernel matrix is given on the input matrix. The Kernel matrix on the entry matrix is obtained by a mathematical operation called convolution by sliding. The element-by-element matrix multiplication operation is performed and adds the output to the characteristics map in each slot. A convoluted image is calculated according to equation 5 as follows:

$$C(i,j) = \sum_b \sum_a I(b,a)k(i-b,j-a) \qquad (5)$$

Pooling Layer – The feature map of the demerit result has the disadvantage of recording the exact position of the input. This implies that the input image entirely generates a map with different characteristics when rotation and cropping occur or there is a minor change in the input.To prevent this problem, we headed for the CNN layers.The gradually invariant translation of the entry is represented by Pooling.The invariance of the translation means that if a small amount of input is changed, the pooled output values are not changed to some extent.

Let b is the input layer and a be the output layer is fully connected and it is given in eq 6:

$$y\ i = \sigma\ (\ wc\ 1\ xc\ 1 + \cdots + w\ cb\ x\ a\ ) \qquad (6)$$

The final output of the fully connected layer is given as eq 7:

$$y = \sigma\ (\ wc\ 1{,}1\ x\ 1 + \cdots + wc\ 1{,}b\ x\ a\ ) \vdots \sigma\ (\ wc\ n{,}1\ x\ 1 + \cdots + w\ a{,}b\ x\ a) \qquad (7)$$

3.7 Fully Connected Layer

The output of CNN is taken from the preceding layer that is given as input for the fully connected layer and several layers are connected in CNN. The first layer of the connected network is interconnected with the second layer, as shown in Figure 4.

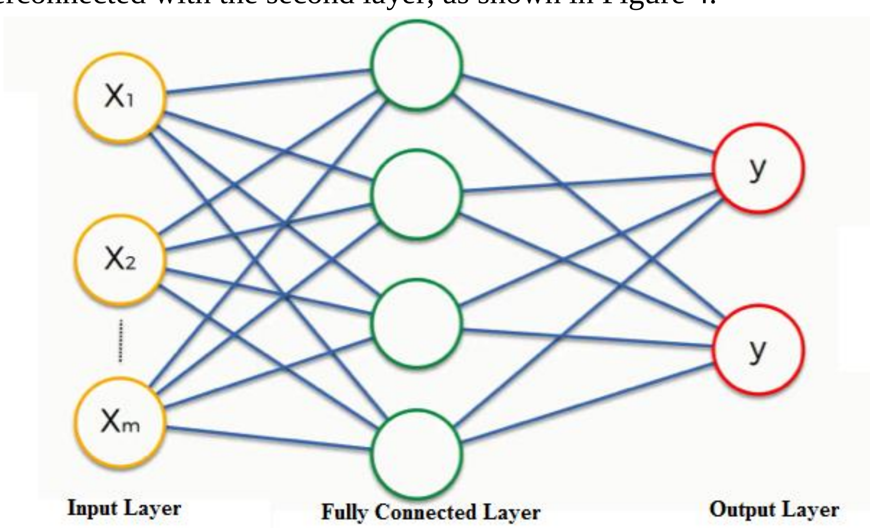

Figure 4: Fully connected layer, where the first layer is connected with the second layer of every node

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

4 Results and Discussions

The proposed system model was formed for the cardiac data set, in which each of the models precalculates the probabilities of the class label of the given patient record.The primary objective of training models is to reduce the binary cross-entropy loss between two labels.An open-source Python software is used for implementing the templates. Firstly, the dataset is classified into two classes, i.e., HD andNo HD. The other classifier that has four classes is formed, which identifies the different data sets. A Python programming language program was written and executed to determine the number of existing types. The program searches for the dataset and groups it into four different classes. Subsequently, the determination of the number of classes, the type of folders, and the closest class results are automatically labeled and processed. With the help of the four classes, the formed template was tested, and outputs were generated along with the confusion matrix. It is expected that the four classes will be 87% accurate, which is as high as the outputs generated.

4.1 Dataset

In the model described, Kaggle, which is a standard data set on cardiac diseases, whose data is collected from the four prestigious hospitals in Switzerland and the United States of America. The datasets include 303 records and 13 attributes. Sufficient values are not present in the dataset. In the dataset, an attribute that is referred to as a target, which indicates whether an individual has heart disease or not. If the indicated target value is 1, then the patient has heart disease, or if the value is 0, then the patient does not have heart disease. In the dataset folder, there are 165 patients with heart disease and 138 patients with no heart disease. Using CNN, a model of a classifier is generated to categorize binary classes from the data and file system that stores the model itself.Table 1 indicates that the calculated accuracy ratio for the binary classification is 97%.

Table 1: Computed Accuracy Rate of classification

Attribute	Description	Type
Age	Years	Num
Sex	Gender	Nom
Chest pain	Cp	Nom
Bp	Blood_pressure	Num
Chol	Cholesterol	Num
Ca	Num of major vessels	Num
Slope	Exercise	Nom
Thal	T3,t6,t7	Num
Target	Disease prediction	Num

Figure 5 shows that the ROC curve for the given dataset during training and validation of the proposed system and prediction did with patient age.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

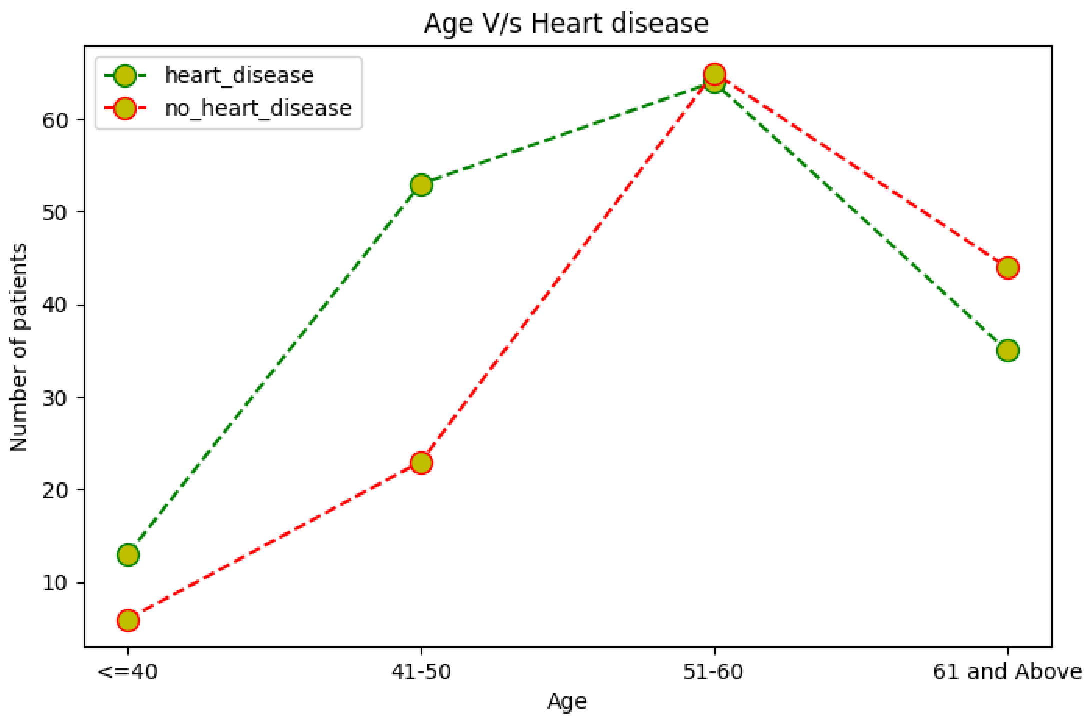

Figure 5: Roc Curve for Proposed system

Figure 6 shows that the patient with gender who having heart disease in the given dataset.

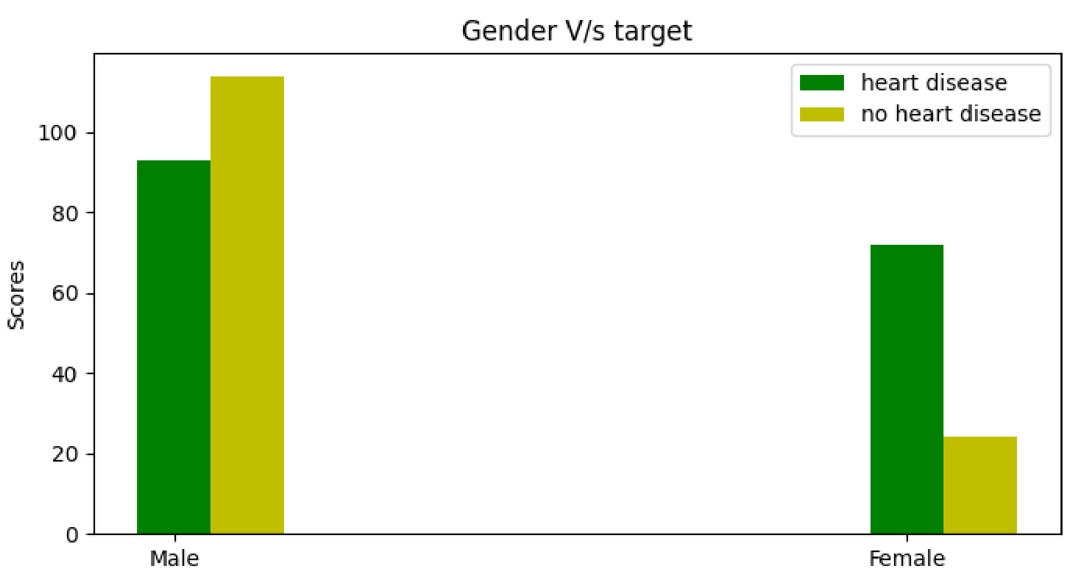

Figure 6: Gender vs target who had heart disease in the given dataset

Blood disorder of Thalassemia status of patients having heart disease and Thalassemia blood disorder status of patients who do not have heart disease are shown in Figures 7 and 8 respectively.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

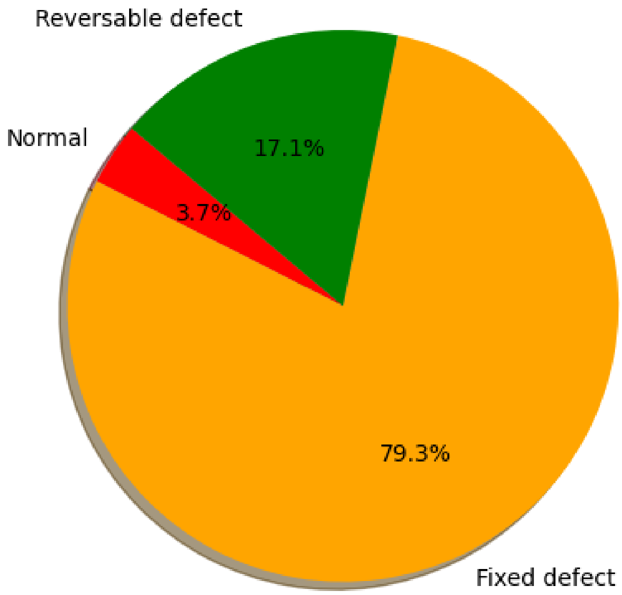

Figure 7: Blood disorder of Thalassemia status of patients having heart disease

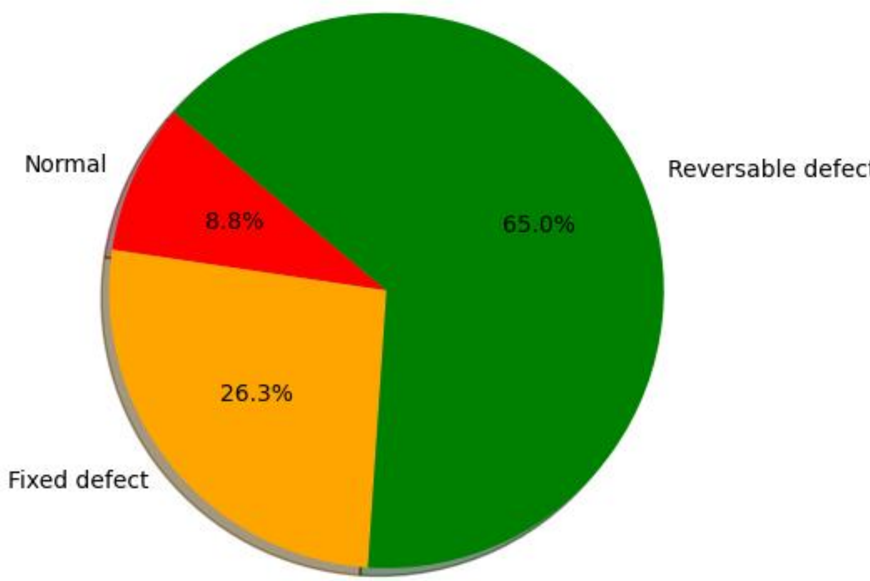

Figure 8: Blood disorder of Thalassemia status of patients who do not have heart disease

The binary classification of diagnosed records of patients and prediction of heart disease is represented in the confusion matrix as shown in figure 9. The prediction is done by two different classes like 0 and 1. where 0 denotes a patient without heart disease and 1 denotes that the patient having heart disease.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

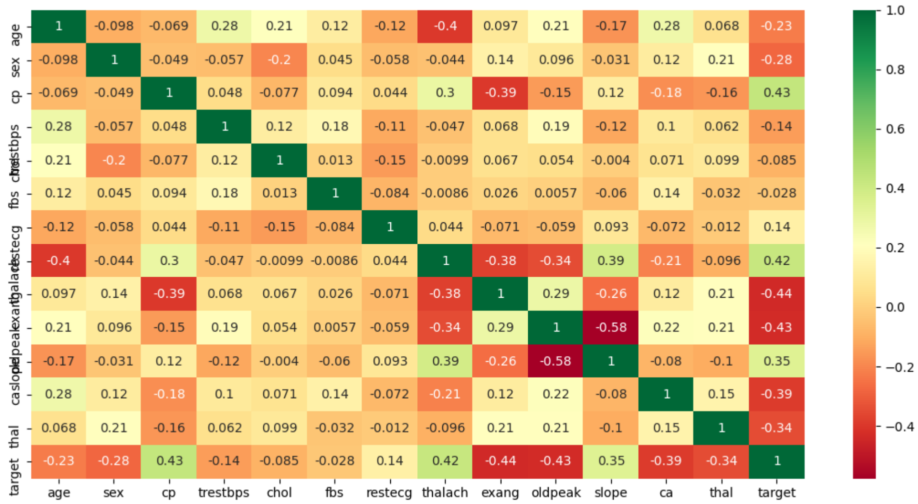

Figure 9: Confusion Matrix for the proposed method

The proposed model's performance concerning the precision, recall, F1 score, and accuracy of 97% is shown in Table 2.

Table 2: Accuracy rate of this proposed model

	precision	recall	f1-score	Accuracy %
0	0.87	0.78	0.82	97
1	0.76	0.85	0.80	97

The binary classification of experiments is reported using results of performancemetricsof evaluation with accuracy (a), precision (p), recall (r), and F1 score are calculated using formulas as shown in eq 8,9,10,11. The parameters using the elements of true positive(tp), true negative (tn), false positive (fp), and false negative (fn).

$$a = (t_p + t_n) / (t_p + t_n + f_p + f_n) \quad (8)$$

$$p = t_p / (t_p + f_p) \quad (9)$$

$$r = t_p / (t_p + f_n) \quad (10)$$

$$f1 = (2xpxr) / (p+r) \quad (11)$$

The predicting accuracy score is calculated and show the results below:

Majority Predicting accuracy score: 0.7912087912087912
Weighted Average accuracy score: 0.8131868131868132
Bagging_accuracy score: 0.8021978021978022
Ada_boost_accuracy score: 0.7362637362637363
Gradient_boosting_accuracy score: 0.8131868131868132

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

The training and validation accuracy model for the proposed system is shown in figure 10.

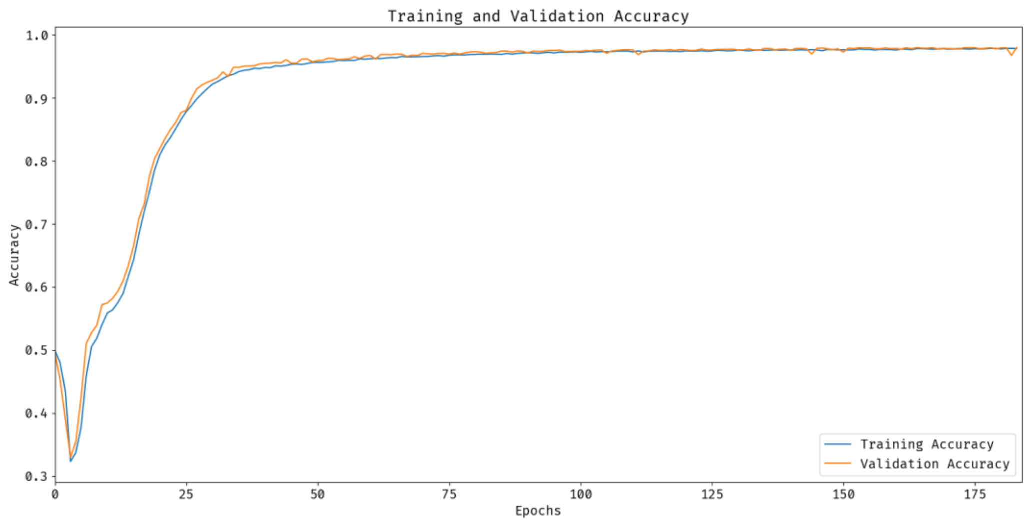

Figure 10: Accuracy of Training and Validation

Below figure 11 shows the results of the proposed system using accuracy percentage, precision, recall, and F-1 score, where 0 is denoted that the patient not having heart disease and 1 denotes that the patient having heart disease.

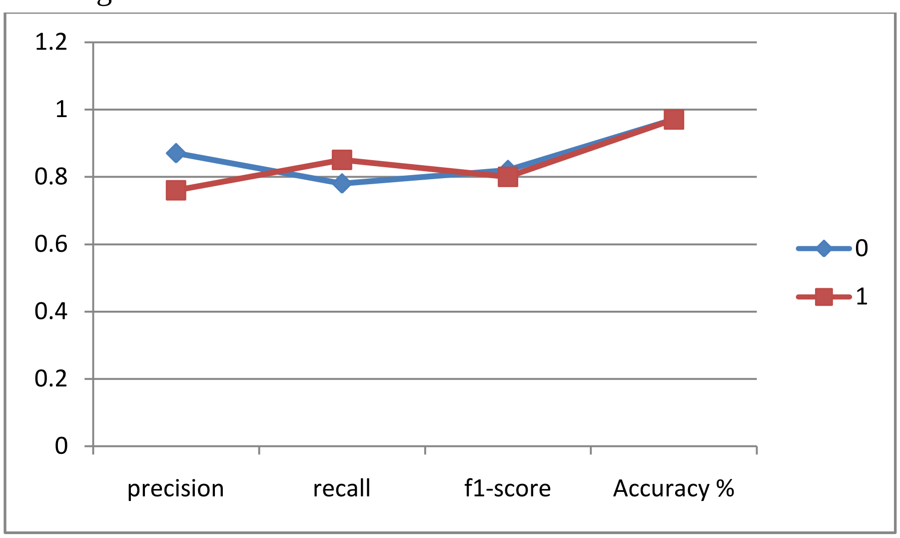

Figure 11:Accuracy rate of this proposed model

K-nearest neighbor (KNN) algorithm is mainly used for clustering of classification and regression. The parametric model having the training data of the non-parametric method. KNN will directly compare with nearer neighbors of the training data. KCC algorithm having a group of K clusters as data points. The shortest distance of cluster with datapoint assigned with parametric which are compared in below figure 12.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6623 - 6640
Received 16 February 2021; Accepted 08 March 2021.

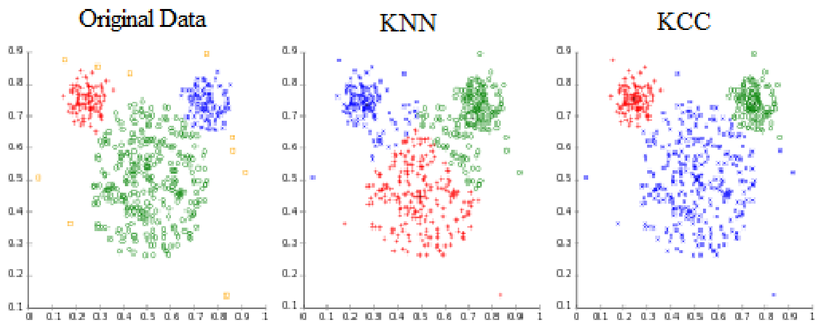

Figure12. Comparison of KNN and KCC algorithm

The comparison of execution time between KNN and KCC is shown in below figure 13.

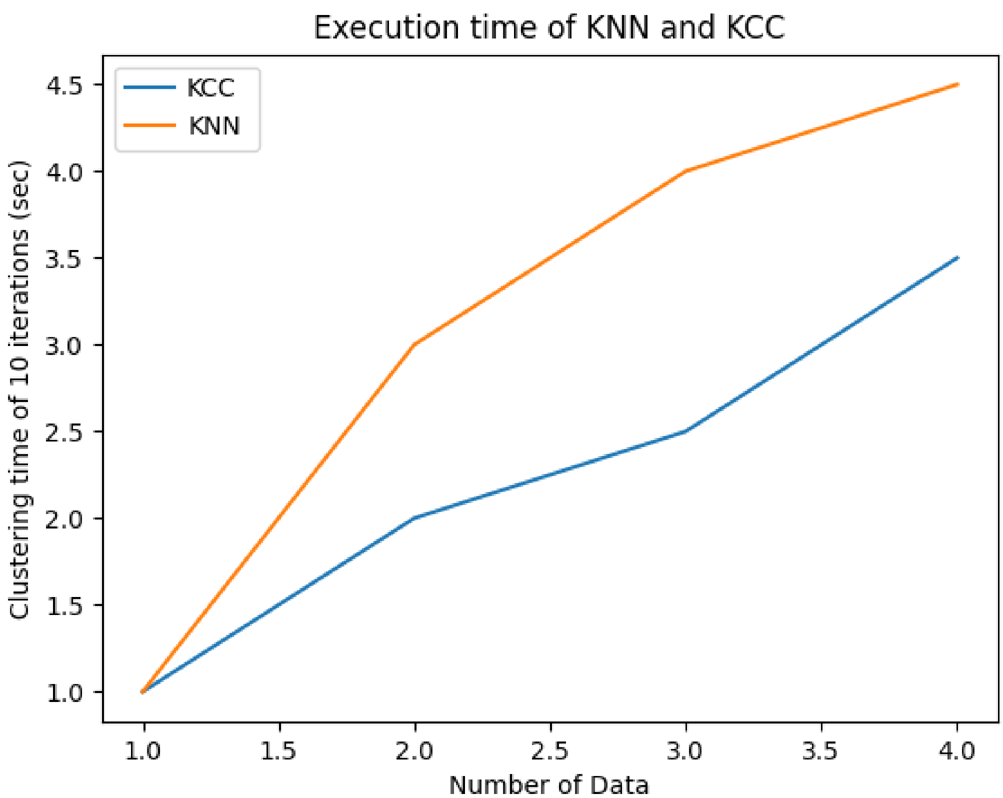

Figure 13. The execution time of clustering data

5 Conclusion

In this paper, we proposed a new CNN with KCC architecture that is designed for the classification of histopathological images. With the assistance of the publicly available Kaggle data set, the training and evaluation of the architecture are implemented. The 15 convolution overlays and two fully connected overlays are used in the proposed architecture. The highest 95.46% accuracy is achieved with this DNN architecture. In this system, we also studied the CNN with KCC performance effects with different activation functions and the effect of locating the network activation function. The purpose of this proposed system is to accommodate a small collection of patient data and advanced models of deep learning that have the clinical domain. As a result of this work, the above-mentioned procedure structure connects a broad range of data and records to a logical framework and then applies the results to the final analysis. Because of the unequal distribution of health records, these tools may be used to make the best decision. Future work will include the implementation of alternative neural organization models such as the Generative Adversarial Network (GAN) and the Attention-based Recurrent Neural Network. It would also provide a better understanding of how to predict various types of heart diseases. Furthermore, the conversion of each output target's binary classification into a multi-label classification task is considered for the solution of its imbalance and control of its structure for our dataset.

References

1. Mohan, S., Thirumalai, C., &Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554.
2. AlGhatrif, M., Cingolani, O., &Lakatta, E. G. (2020). The dilemma of coronavirus disease 2019, aging, and cardiovascular disease: insights from cardiovascular aging science. JAMA cardiology, 5(7), 747-748.
3. Anitha, S., &Sridevi, N. (2019). Heart disease prediction using data mining techniques. Journal of Analysis and Computation.
4. Sharma, M., Singh, G., & Singh, R. (2019). An advanced conceptual diagnostic healthcare framework for diabetes and cardiovascular disorders. arXiv preprint arXiv:1901.10530.
5. Waris, S. F., &Koteeswaran, S. (2021). Heart disease early prediction using a novel machine learning method called improved K-means neighbor classifier in python. Materials Today: Proceedings.
6. Shickel, B., Tighe, P. J., Bihorac, A., &Rashidi, P. (2017). Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics, 22(5), 1589-1604.
7. Spritzer, C. E. (2009). Progress in MR imaging of the venous system. Perspectives in vascular surgery and endovascular therapy, 21(2), 105-116.
8. Domont, F., &Cacoub, P. (2016). Chronic hepatitis C virus infection, a new cardiovascular risk factor?. Liver International, 36(5), 621-627.
9. Wang, X., Ota, N., Manzanillo, P., Kates, L., Zavala-Solorio, J., Eidenschenk, C., ...&Ouyang, W. (2014). Interleukin-22 alleviates metabolic disorders and restores mucosal immunity in diabetes. Nature, 514(7521), 237-241.
10. Blaner, W. S. (2019). Vitamin A signaling and homeostasis in obesity, diabetes, and metabolic disorders. Pharmacology & therapeutics, 197, 153-178.
11. Erdoğan, S. (2010). Modelling the spatial distribution of DEM error with geographically weighted regression: An experimental study. Computers & Geosciences, 36(1), 34-43.

12. Cunningham, P., & Delany, S. J. (2020). k-Nearest Neighbour Classifiers--. arXiv preprint arXiv:2004.04523.
13. Belgiu, M., &Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.
14. Lin, G., &Shen, W. (2018). Research on convolutional neural network based on improved Relu piecewise activation function. Procedia computer science, 131, 977-984.
15. Jahromi, A. H., &Taheri, M. (2017, October). A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. In 2017 Artificial Intelligence and Signal Processing Conference (AISP) (pp. 209-212). IEEE.
16. American Diabetes Association. (2016). 8. Cardiovascular disease and risk management. Diabetes care, 39(Supplement 1), S60-S71.
17. Ettehad, D., Emdin, C. A., Kiran, A., Anderson, S. G., Callender, T., Emberson, J., ...&Rahimi, K. (2016). Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. The Lancet, 387(10022), 957-967.
18. Mohan, S., Thirumalai, C., &Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554.
19. Zhu, H., Samtani, S., Brown, R., & Chen, H. (2020). A deep learning approach for recognizing activity of daily living (ADL) for senior care: Exploiting interaction dependency and temporal patterns. Forthcoming at MIS Quarterly.
20. Mammen, J. R., Java, J. J., Halterman, J., Berliant, M. N., Crowley, A., Frey, S. M., ...&Arcoleo, K. (2019). Development and preliminary results of an Electronic Medical Record (EMR)-integrated smartphone telemedicine program to deliver asthma care remotely. Journal of telemedicine and telecare, 1357633X19870025.
21. Wang, H., Yang, Y., Liu, B., & Fujita, H. (2019). A study of graph-based system for multi-view clustering. Knowledge-Based Systems, 163, 1009-1019.
22. Li, X., & Liu, H. (2018). Greedy optimization for K-means-based consensus clustering. Tsinghua Science and Technology, 23(2), 184-194.
23. Liu, H., Zhao, R., Fang, H., Cheng, F., Fu, Y., & Liu, Y. Y. (2017). Entropy-based consensus clustering for patient stratification. Bioinformatics, 33(17), 2691-2698.