International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Building a Novel Ensemble Learning – Based Prediction Framework for Diagnosis of Coronary Heart Disease

Abdullah Al Gahtani*[1], Hoda El-Sayed[2]

Submitted: 22/07/2022 Accepted: 25/09/2022

Abstract: The newer technologies such as data mining, machine learning, artificial intelligence and data analytics have revolutionized medical sector in terms of using the existing big data to predict the various patterns emerging from the datasets available in the healthcare repositories. The predictions based on the existing datasets in the healthcare sector have rendered several benefits such as helping clinicians to make accurate and informed decisions while managing the patients' health leading to better management of patients' wellbeing and health-care coordination. The millions of people have been affected by the coronary artery disease (CAD). There are several machine learning including ensemble learning approach and deep neural networks-based algorithms have shown promising outcomes in improving prediction accuracy for early diagnosis of CAD. This paper analyses the deep neural network variant DRN, Rider Optimization Algorithm-Neural network (RideNN) and Deep Neural Network-Fuzzy Neural Network (DNFN) with application of ensemble learning method for improvement in the prediction accuracy of CAD. The experimental outcomes showed the proposed ensemble classifier achieved the highest accuracy compared to the other machine learning models.

Keywords: *Heart disease prediction, Deep Residual Network (DRN), Ensemble classifiers, coronary artery disease.*

1. Introduction

Heart diseases have been considered as fatal and lifestyle-affecting disease among all disease. Healthy life of human is solely dependent upon the effective functioning of heart. The symptoms of heart disease are weakness of body, difficulties in breath, swollen feet and fatigue [14]. Based on World Health Organization (WHO) released in 2019, approximately 17 million people lost their life every year across the world because of heart diseases [1]. Typically, there are different categories of heart diseases are existed, namely coronary artery diseases (CAD), congenital heart disorder, arrhythmia, and so forth. The patients with heart disease show different symptoms whuch may involve dizziness, chest pain, and heavy sweating [2]. It is observed that nearly 17 million people are losing their life every year because of heart diseases and it is very necessary to predict the heart diseases at earlier stage. By detecting the heart disease at an early stage and providing better treatments based on this prediction can save patient's life thereby reducing the mortality rate [2].

Effective prognosis of heart disease provides personalized treatments to patients and this needs a thorough evaluation of cardiovascular assessment of human that having the symptoms, like chest tightness, pain, pressure, numbness, shortness in breath, and so on [3]. The decision making is a complicated process to the clinicians since a slight ignorance may drive the patient to death edge [4]. To provide accurate diagnosis about heart disease, it is necessary to establish an automated model for heart disease detection that facilitates the task of clinicians in making proper decisions in accordance with the symptoms and medical history of a patient [5]. There are several computational models, such as machine learning technique utilized to detect the people at high-risk stage.

So far, researchers have designed numerous machine learning systems utilizing existing heart disease databases and achieved inconsistent results [6]. Machine-learning techniques have been employed in wide arenas of medical services but, researchers are always attempting for improvisation of such models. Ensemble learning is one of the methods that are scientifically proven to improve the tasks of machine learning [7]. An ensemble classifier is nothing, but the combination of individual classifiers and researchers prove that ensemble classifiers usually provide better performance than traditional classifiers [8]. However, the members may change in framework, while heterogeneous ensemble learning consists of members having various base learners [9].

For past few years, Artificial Intelligence (AI), deep and machine learning are techniques that embedding clinical care model and industries. Deep learning offers clinical care area with the capability to assess information at very high speeds without decreasing the accuracy level. Deep learning functions like brain, wherein it utilizes statistical information in a system [10]. Deep learning is comprised with diverse neurons and layers that lie on a Neural Network (NN) that provides micro-analysts to produce a desired output [11]. Deep learning is very substantial to provide accurate results for heart disease prediction in early diagnosis of patients. Computer-aided diagnostic models depending on machine learning detective methods that are non-invasive and can also assist physicians to provide accurate results and thus mitigate the sufferings of patients [12]. Different types of predictive models

[1] *PhD candidate, Department of Computer Science, Bowie state university, USA.*
[2] *Professor, Department of Computer Science, Bowie state university, USA.*
* *Corresponding Author Email: Algahtania0407@students.bowiestate.edu*

[30] have been introduced and highly utilized for decision making [4]. The deep learning methodologies efficiently boost the process in managing the large-scale data, which is superior to other conventional learning models [13]. Deep learning models have served an important part in healthcare area for diseases categorization, such as diabetes, heart disease using gathered biomedical information that represents diverse kinds of clinical applications utilizing deep learning methodologies and observed certain constraints and requires little improvisation [14].

The main aim of this research project is to develop a novel and robust intelligent computational solution using the ensemble learning for achieving the increased accuracy in terms of predictability of coronary artery disease risk in different context of datasets. The key characteristic of our approach is to build the deep learning based predictive models tested on a different size of datasets to examine the predictive accuracy of the proposed intelligent computational solution.

This paper contributes to the development of a novel computationally intelligent solution for prediction of coronary heart disease using the state-of-the-art machine learning and artificial neural networks. Another contribution of this research work is to utilize increase the predictive performance of the classifiers using the novel feature selection and ensemble method for improving the predictive outcomes of different classifiers employed in this study.

The rest of the paper in organized into four main sections. Section 2 present and discuss the existing literature in the area of ensemble learning and machine learning methods for prediction of heart diseases; section 3 puts forth the methodological overview for achieving the aim of this paper; section 4 presents the results and analysis, while section 5 gives conclusion.

2. Related work

Chandrika and Madhavi [15] employed different machine learning techniques such as random forest with a linear model, generalized linear model, Naïve Bayes, Logistic regression, decision tress and support vector machine in order to predict the heart diseases. They trained and tested their models on the Cleveland dataset in Python Jupyter Notebook platform. They found that hybrid random forest with a linear model (HRFLM) performed better with accuracy of 88.4% compared to the other machine learning classifiers such as Deep learning, random forest and logistic regression.

Ghosh et al [16] tried to develop efficient prediction of cardiovascular disease using hybrid machine learning classifiers with the LASSO and Relief as feature selection methods. They used UCI-based heart disease datasets involving Cleveland, Long Beach VA, Switzerland, Hungarian and Statlog datasets. The classifiers tested by authors Decision-tree bagging method (DTBM), Random Forest bagging method (RFBM), K-Nearest Neighbors Method (KNNBM), AdaBoosting Method (ABBM), and Gradient Boosting Method (GBBM)

They reported 99.05% accuracy for Relief-RFBM in predicting the risk of heart disease. Though it achieved the 99.05% accuracy, but it was only tested on datasets with small number of instances and attributes, so cannot generalized for other feature selection algorithms and other sources of heart disease datasets. The work of Ghosh and his colleagues [16] suggested the promise of hybrid classifiers in different formats to develop the innovative intelligent solution for diagnosing the heart disease in practical settings.

Reddy et al [17] developed the machine learning-based model for prediction of heart diseases using decision trees and support vector machine classifiers, and experiments were conducted in Python. They used unspecified datasets. They tested performance of five classifiers: KNN, SVM, RF, Naïve Bayes and Neural Network; and showed that RF outperformed support vector machine with 90-95% accuracy. However, they concluded that performance of the classifier varies with the number of features, addition of different risk factors in the feature list, features selected and the optimization strategies.

Mustafa Jan, et al. [18] designed an ensemble model for agglomerating the detective capability of diverse classifiers for high prediction accuracy. In order to diagnose CAD, ensemble method utilized five classifiers. Moreover, this model also presented a smart heart disease detection model, which was scalable and highly desirable. The system model only utilized limited datasets was a major disadvantage.

Farman Ali, et al. [19] designed a healthcare model for disease detection by employing ensemble deep learning and feature fusion models. Initially, feature fusion model merged refined features from sensor data as well as electronic medical records to produce essential healthcare information. In second step, the information gain model removes the unnecessary features and chose the significant ones. At last, the ensemble deep learning system was trained for disease detection. The method was inefficient to handle numerous features and high capacities of healthcare records.

Qi Zhenya, and Zuoru Zhang, [20] introduced a cost-effective ensemble method to enhance effectiveness of the prognosis and minimized the misclassification expenses. The developed system included five heterogeneous classifiers. The developed ensemble model achieved significant results in medical decision making. Although the misclassification cost of this system was low, the time complication of this method was relatively high.

Xiao-Yan Gao, et al. [21] modeled an ensemble learning methods that were employed to enhance the classification accuracy while predicting the heart disease. Here, two feature selection techniques, like linear discriminant analysis (LDA) and principal component analysis (PCA) were utilized to choose the significant features from dataset. The method achieved superior performance than any other conventional methods but, computational complexity of this system was high.

Bayu Adhi Tama, et al. [22] developed a two-tier ensemble system for coronary heart disease (CHD) detection based on machine learning, such as classifier ensembles. A stacked framework was designed to merge the class label prediction of three ensemble learners, like gradient boosting machine, and random forest. A particle swarm optimization-based feature selection was performed to select the important feature. The results of this method were extremely high performance. However, it demanded large time for processing the tasks.

3. Research Methods

3.1 Description of datasets

Four datasets derived from the UCI repository were used in this study, which include Cleveland, Hungarian, Switzerland, and VA Long beach dataset. The Cleveland has 303 instances and 13 attributes and Hungarian dataset includes 294 instances and 14 attributes. On the other hand, Switzerland dataset consists of 123 instances and 13 attributes, whereas VA long beach dataset includes 200 instances and 13 attributes. Among them, Cleveland dataset is the most commonly used dataset by ML researchers.

The attributes utilized in this research work for prediction of CAD are presented in the Table 1:

Table 1: The attributes utilized for prediction of CAD

Serials	*Abbreviated names*	*Full names*	*Description*
1	Sex	Sex	1 = male, 0 = female
2	Age	Age in years	Age of patients in years
3	Cp	Chest pain type	Chest pain type help diagnosis of CAD from other different heart diseases
4	Trestbps	Resting Blood Pressure (mm Hg)	Normal range: 120/80; Higher than normal range may reflect the risk of CAD
5	Chol	Serum cholesterol level	Cholestrol – type of triglycerides in blood, normal range: 170mg/dL
6	Fbs	Fasting blood sugar	If Fbs>120 mg/dL, risk of CAD, Fbs<100 mg/dL normal
7	Restecg	Resting electrocardiographic results	These results show the risk of CAD
8	Thalach	Maximum Heart rate	Maximum heart rate achieved
9	Exang	Exercise-induced angina	Binary classification, 0 = no, 1 = yes
10	Oldpeak	The ST depression caused by the peak level of exercise compared to rest	The ST depression caused by the peak level of exercise compared to rest
11	Slope	Slope	Slope of the peak caused by the ST segment
12	Ca	Number of vessels	Number of vessels (0-3) colored by fluoroscopy
12	Thal	Thalasemia	Normal = 3, fixed defects = 6, reversible defects = 7

3.2 Pre-processing of the Dataset using quantile normalization

Pre-processing is the significant step as it removes the redundant data from the acquired input data and makes it feasible for further processing. Here, pre-processing is performed using data transformation based on Quantile normalization [23]. The data transformation is the process that converts the data into a meaningful form. Here, the input data D_i is passed to pre-processing module and process is effectively carried out using Quantile normalization.

Quantile normalization is the significant normalization method adopted mostly for analyzing the high-dimensional data. The steps followed by Quantile normalization are described as follows:

Step 1: Initially, the features in each dataset are ranked by magnitude.

Step 2: Compute the mean value of features holding the equivalent rank.

Step 3: Replacing the measures of all features exhibiting that specific rank with this mean value.

Step 4: The final process is to rearrange features in an individual sample in a hierarchy manner.

The result achieved from pre-processing step is considered as P_i and it is the forwarded to feature selection module for further processing.

3.3 Feature selection using Kendall tau distance

The pre-processed data P_i is forwarded to feature selection module, where desired or significant features are chosen using Kendall tau distance [24]. Once the distance is computed using Kendall tau distance, the top m maximum feature is selected. The feature selection mechanism includes minimization of values by choosing such features that facilitate final detection and considering remaining attributes as eliminated and these eliminated features could not be considered for further evaluation procedure.

When more input rank lists $(\tau_1, \tau_2, ...\tau_m)$, are existed, the Kendall tau distance among a list τ and $(\tau_1, \tau_2, ...\tau_m)$ is computed as,

$$K(\tau_1, \tau_2, ..\tau_j, ...\tau_m) = \frac{1}{m}\sum_{j=1}^{m} K(\tau, \tau_j) \tag{1}$$

Here, K is the Kendall's tau distance and m denotes the total samples. The reason for selecting Kendall tau distance for feature selection process is that it concurrently accomplishes many significant properties in social choice disciplines. The selected feature is represented as F_i.

3.4 Ensemble learning-based CAD prediction model

The approach proposed for processing and analyzing the dataset involved the processing and analysis of selected datasets, followed by application of different hybrid classifiers involving deep learning and neural networks. The hybrid classifiers employed for classification of CAD included the DRN, RideNN and DNFN.

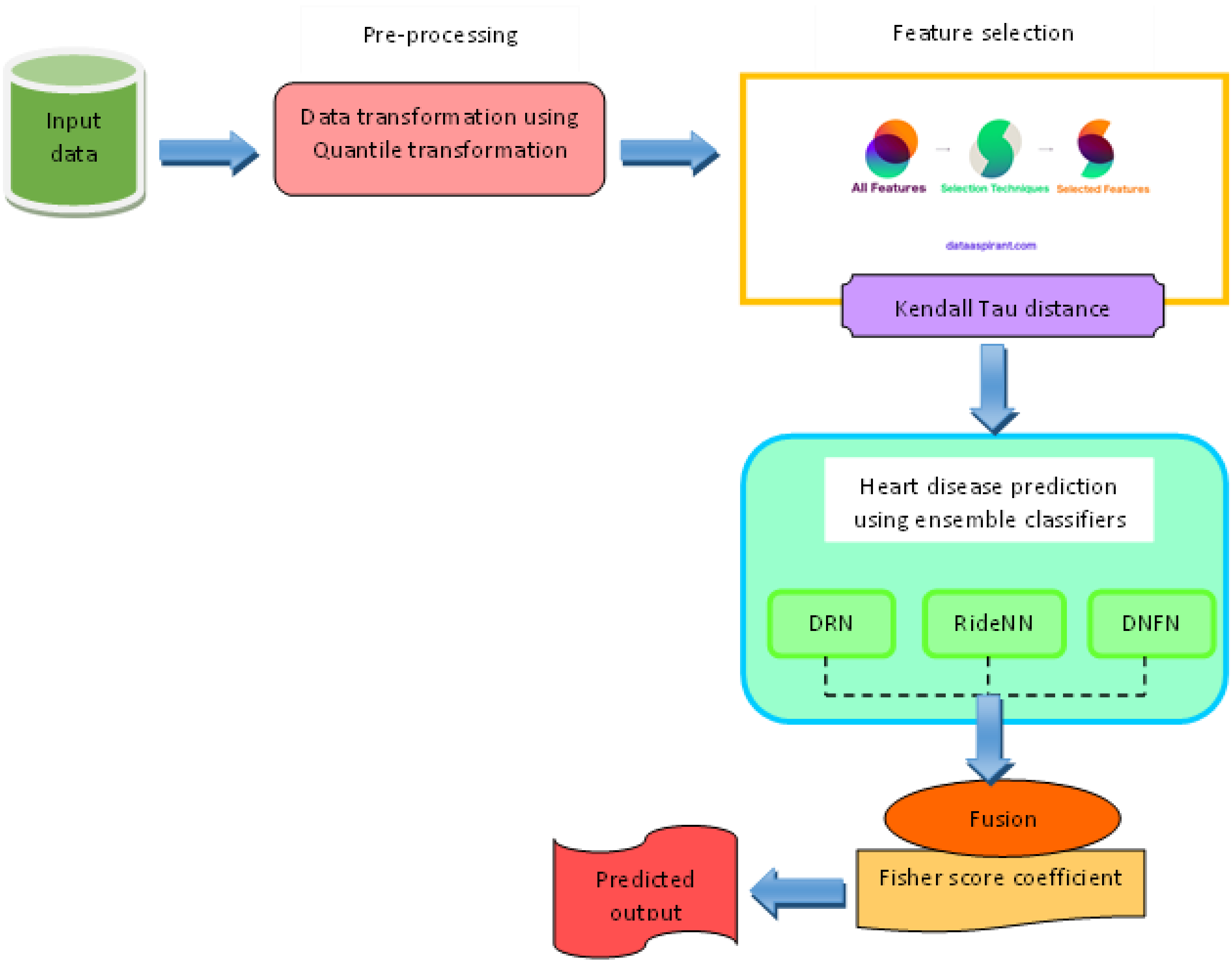

Fig 1. Block diagram of proposed ensemble learning-based CAD prediction model.

The different components of the proposed model have been discussed in the preceding sections, while the ensemble learning-based hybrid classifiers and ensemble learning technique are discussed in the subsequent sections.

3.5 Ensemble learning-based hybrid classifiers

DRN

DRN is mainly employed for different pattern recognition tasks. Unlike other deep learning techniques, DRN [25] has the benefits of high training speed, easier gradient transmission and so forth. DRN can be adopted to attain a higher performance for both classification and regression purposes. Thus, simple DRN is comprised with different layers, such as convolutional (Conv), pooling, activation function, batch normalization, and residual blocks.

The 2-D convolutional layer (Conv2d) was employed to minimize the parameters highly and enhanced the relevance of features with each other in the local receptive fields called kernels in the convolutional layer by performing cross-validation function using the following mathematical formulation:

$$Conv2d(F_i) = \sum_{w=0}^{l-1} \sum_{h=0}^{l-1} Z_{w,h} \bullet F_{i_{(u+w),(v+h)}} \tag{2}$$

$$Conv1d(F_i) = \sum_{k=0}^{c_{in}-1} Z_k * F_i \tag{3}$$

In Eq. (3), F_i : the input, u and v : coordinates; Z : $l \times l$ kernel matrix, which is referred as a learnable factor in training procedure. Moreover, w and h : location index of 2-D kernel matrix. In Eq. (4), Z_k : Kernel dimension for k^{th} input neuron and $*$ specifies cross-correlation factor without zero padding.

Once kernels with highly relevant features were selected; kernels located between two convolutional layers were condensed to reduce the spatial demission of features in the pooling layer by using the following Eq. 4 and Eq. 5:

$$w_{out} = \frac{w_{in} - k_w}{A} + 1$$
(4)

$$h_{out} = \frac{h_{in} - k_h}{A} + 1$$
(5)

Here, w_{in} and h_{in} denotes width and height of input 2D matrix, whereas w_{out} and h_{out} denotes result. Moreover, the width and height of the kernel dimension is represented as k_w and k_h, respectively.

Following this, convergence rate is increased while mitigating the vanishing gradient issue using the Rectified Linear Unit as a non-linear activation parameter in the activation function layer, as expressed in Equation 6.

$$\mathrm{Re}LU(F_i) = \begin{cases} 0, & F_i < 0 \\ F_i & F_i \geq 0 \end{cases}$$
(6)

The computational complexity was decreased while increasing the convergence rate by partitioning the kernels of features into mini-batches in the batch normalization layer which performs the batch normalization function for decreasing the internal covariate shift by maintaining input by means of scaling. This increases the learning rate and also vanishes the gradient problem. The size matching parameter is presented to correlate the size of input and output in a case when the dimension is dissimilar.

$$O_1 = M(F_i) + F_i$$
(7)

$$O_1 = M(F_i) + \delta.F_i$$
(8)

Here, the input and output of residual block is denoted as F_i and O_1, respectively. The function M signifies the mapping correlation between inputs and outputs and δ shows size matching parameter.

RideNN

The RideNN [26] classifier is devised by the combination of Rider Optimization Algorithm (ROA) with Neural Network (NN) classifier. By doing so, the output of the model can be categorized as presence and absence of heart disease. The RideNN classifier has three layers, namely input, hidden, and output layer. The input subjected to the RideNN classifier is expressed in Eq. 9 and Eq. 10 as follows,

$$F_i = \{F_i^1, F_i^2, \ldots F_i^a, \ldots F_i^B\};\ 1 \leq a \leq B$$
(9)

Here, B denotes the total number of features. The categorization task is accomplished based on the weights and the assigned weight of neurons in hidden layer is represented as follows,

$$N = \{N^1, N^2, \ldots N^B\}$$
(10)

The neuron in the hidden layer is considered to possess a bias, which is indicated as N^{B+1} and it outcomes the enough potential. Moreover, the weights and bias in output layer is expressed as N^{B+2} and N^{B+3}, respectively. Hence, the output of RideNN classifier is formulated utilizing a transfer parameter and it is represented as,

$$O_2 = N^{B+2} * \left[\log sig \left(\sum_{a=1}^{B} F_i^a * N^B + N^{B+1} \right) \right] + N^{B+3}$$
(11)

where, $\log sig$ illustrates the log-sigmoid transfer parameter that formulates result from overall input; F_i^a is input of a^{th} neuron; N^B is weight of the B^{th} neuron and N^{B+2} denotes the weight of output neuron.

DNFN

DNFN [27] is a hybrid classifier with deep neural network (DNN) and fuzzy neural network, which is for the first time employed for delivering effective optimization with minimized cost and peak reduction in predicting heart diseases. It consists of input layer, rule-based layer and normalization layer.

In the input layer, the input in the form of selected feature is fed to the DNFN as expressed by F_i. The degree to each input is allotted between 0 and 1 according to the mechanism described by the fuzzy model. Let us consider that there are two premises, like p and q, and one consequent W, which are simply described using equation as below,

$$L_{1,g} = \eta G_g(p)\ or\ L_{1,g} = \eta H_{g-2}(q), \forall g = 1,2,3,4$$
(12)

In the above equation, p and q specifies the inputs to individual g^{th} entity, ηG_g and ηH_{g-2} specifies the antecendent membership functions, while $L_{1,g}$ indicates the degree of membership.

$$\eta G_g(p) = \frac{1}{1 + \left| \frac{p - b_g}{d_g} \right| 2\omega_g}$$
(13)

where, ω_g, b_g, and d_g represents the membership operators of the premise parameters that are optimized by means of training.

In the rule base layer, set of rules is defined. Here, the membership values represent the firing strength of rule, which is given by,

$$L_{2,g} = W_g = \eta G_g(p) \eta H_{g-2}(q), \forall g = 1,2$$
(14)

In the normalization layer, an individual entity formulates proportion of firing strength of g^{th} rule with overall sum of firing strength of total rules. W_g denotes generic network function. The result of individual rule is normalized through firing strength, which is represented as follows,

$$L_{4,g} = \overline{W}_g I_g = \overline{W}_g \left(Aa_g p + Bb_g q + Cc_g \right), \forall g = 1,2$$
(15)

Here, Aa, Bb, and Cc denotes the set of consequent parameters. Thereafter, the final layer is assumed as the summation layer, which formulates summation of the preceding layer outputs. The final output is described as,

$$O_3 = L_{5,g} = \sum_g \overline{W}_g I_g = \frac{\sum_g W_g I_g}{\sum_g W_g} \tag{16}$$

3.6 Fusion based on Fisher score coefficient

The outputs from the ensemble classifier were fused together to obtain the final predicted result using Fisher score coefficient [28]. Fisher score is one of the commonly utilized feature selection techniques and it is used to combine the selected features. The basic principle of fisher score coefficient is to increase the lower bound of conventional fisher score. The predicted result after fusing the outcomes of ensemble classifier is denoted as FS .

$$FS = \min\left(E_d\left(FS_1, FS^*\right), E_d\left(FS_2, FS^*\right), E_d\left(FS_3, FS^*\right)\right) \tag{17}$$

$$FS = \frac{\sum Nn_\ell \left(\psi_{\partial\ell} - \psi_\partial\right)^2}{\sum Nn_\ell * \rho_{\partial\ell}^2} \tag{18}$$

Here, FS_1 denotes the fisher score of ∂^{th} feature with respect to O_1, FS_2 and FS_3 illustrates the fisher score of ∂^{th} feature with respect to O_2 and O_3, respectively. Moreover, $\psi_{\partial\ell}$ and $\rho_{\partial\ell}$ implies the mean and variance of ∂^{th} feature in ℓ^{th} class.

3.7 Performance evaluation metrics

In order to measure the performance of various classifiers utilized in our study will be analysed using different performance metrices in Python environment, which involve the accuracy, precision, specificity, and sensitivity.

Accuracy

The accuracy is a measure of the overall performance of the proposed solution in predicting the outcomes of the coronary artery heart disease, and is represented by the following formula [29]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100 \tag{19}$$

Specificity

It is used to measure the ability of the proposed solution in predicting the healthy individuals as healthy people, and unhealthy people as unhealthy subjects in the given dataset. This is given by the following formula [92]:

$$Specificity = \frac{TN}{TN + FP} * 100 \tag{20}$$

Sensitivity

This is measure of ratio of heathy people to the unhealthy people with coronary artery heart disease. This is measured through the following formula, which measures the ability of the proposed system in measuring the total people with heart diseases in the given dataset [92].

$$Specificity = \frac{TP}{TP + FP} * 100 \tag{21}$$

4. Results and Analysis

The aim of this research work to predict the CAD by developing a novel ensemble learning-based intelligent prediction model. The proposed approach utilized the combination of deep learning and neural networks and machine learning techniques. DRN is a variant of deep neural network, RideNN is a combination of Rider Optimization Algorithm (machine learning techniques) coupled with neural network, and DNFN which is a combination of deep neural network and fuzzy neural network. The ensemble learning method was applied on the classification outcomes of these classifiers. With this unique set of classifiers, experiments were conducted on UCI repository using Python tool with TenserFlow, Keras and other associated libraries to assess the performance of ensemble learning-based classifiers. Research was executed on 11th generation Intel Corei7 with NVIDIA GeForce RTX 3060 Laptop GPU, two GPU 6.0 GB and 7.9 GB RAM.

Each category of dataset was split into two sets: the training set which involved 90% of the dataset for training the models, and test set which involves 10% of the dataset to test the performance of the ensemble-based classification algorithms.

4.1 Performance evaluation of the proposed model

We compared the performance of the proposed ensemble learning model with four existing prediction models developed by other studies in order to evaluate the performance of our proposed solution in terms of prediction accuracy, sensitivity and specificity, which are majority vote ensemble classifier [1], randomized decision tree ensemble [9], LogiBoost [19], and Two-Tier classifier ensemble [22]. The outcomes obtained from the comparative analysis are shown in the Table 2.

Table 2: The comparison of the performance of proposed ensemble classifier using different performance metrics in context of four different datasets.

Dataset	Metrics	randomized decision tree ensemble	LogiBoost	Two-Tier classifier ensemble	majority vote ensemble classifier	Proposed ensemble classifier
Cleveland	Testing Accuracy	0.713	0.762	0.821	0.901	**0.946**
	Sensitivity	0.791	0.895	0.906	0.908	**0.914**
	Specificity	0.828	0.888	0.894	0.941	**0.949**
Hungarian	Testing Accuracy	0.887	0.902	0.903	0.936	**0.940**
	Sensitivity	0.711	0.831	0.833	0.840	**0.908**
	Specificity	0.820	0.826	0.895	0.941	**0.947**
Switzerland	Testing Accuracy	0.904	0.913	0.921	0.930	**0.947**
	Sensitivity	0.736	0.911	0.919	0.927	**0.935**
	Specificity	0.897	0.906	0.915	0.924	**0.942**
VA long beach	Testing Accuracy	0.904	0.919	0.925	0.930	**0.935**
	Sensitivity	0.910	0.925	0.930	0.935	**0.940**
	Specificity	0.897	0.913	0.918	0.924	**0.929**

The performance of the proposed model was evaluated against three performance metrics including prediction accuracy, sensitivity and specificity, and results are presented in in Table 2, it was observed that the proposed ensemble classifier-based prediction model outperformed the existing machine learning models by 94% accuracy, 91% sensitivity and 95% specificity in the context of prediction accuracy, sensitivity, specificity in the context of Cleveland dataset.

Similarly, when the proposed model was tested on Hungarian dataset, the prediction accuracy, sensitivity and specificity were recorded to be 94%, 91% and 95% compared to other machine learning models. In a similar vein, performance evaluation of the proposed model was done in the context of Switzerland dataset; it achieved 95% prediction accuracy, 94% sensitivity and 94% specificity compared to other models. Furthermore, prediction accuracy, sensitivity and specificity for the proposed model were 93.5%, 94% and 92.9%, respectively, compared to other models, when it was tested on VA Long Beach dataset.

5. Conclusion

In this paper, we have developed a novel ensemble classifier using the hybrid deep learning coupled with machine learning and neural network classifiers. Our proposed approach utilized Kendall tau coefficient as feature selection technique to ranks the features, followed by application of ensemble methods as a classifier for improving the classification performance and prediction accuracy of the classifiers. It was clear from the findings that learning accuracy can be increased by the proposed ensemble classifier on the Kendall Tau coefficient. The proposed ensemble classifier-based model will aid clinicians and health consultants to accelerate the diagnosis of coronary heart disease at eastly stage of disease by employing a s features subset utilized in this paper. Our future work intends to integrate the hybrid machine learning algorithms such as particle-swarm optimization-SVM, and PSO-DRN and Salp swam optimization combined with deep neural networks for developing the early diagnosis model for angina attack.

References

[1]. C. Beulah Christalin Latha, S. and Carolin Jeeva, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques”, Informatics in Medicine Unlocked, vol.16, pp.100203, January 2019.

[2]. K. Subhadra, and Vikas B, “Neural network based intelligent system for predicting heart disease”, International Journal of Innovative Technology and Exploring Engineering, vol.8, no.5, pp.484-7, March 2019.

[3]. K.Srinivas, Dr.G.Raghavendra Rao, and Dr. A.Govardhan, “Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques”, The 5th International Conference on Computer Science & Education, pp.1344-1349, August 2010.

[4]. L. N. Balai, G. K. J. A. K. S. (2022). Investigations on PAPR and SER Performance Analysis of OFDMA and SCFDMA under Different Channels. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 28–35. https://doi.org/10.17762/ijrmee.v9i5.371

[5]. Stephen F. Weng, Jenna Reps, Joe Kai, Jonathan M. Garibaldi, and Nadeem Qureshi, “Can machine-learning improve cardiovascular risk prediction using routine clinical data?”, vol.12, no.4, pp.e0174944, April 2017.

[6]. Resul Das, Ibrahim Turkoglu, and Abdulkadir Sengur, “Effective diagnosis of heart disease through neural

networks ensembles", Expert System with applications, vol.36, no. 4, pp.7675–7680, 2009.

[7]. Garg, K. . (2022). Compactness in General Category Theory. International Journal on Recent Trends in Life Science and Mathematics, 9(1), 19–27. https://doi.org/10.17762/ijlsm.v9i1.138

[8]. Afef Mdhaffar, Ismael Bouassida Rodriguez, Khalil Charfi, Leila Abid, and Bernd Freisleben, "CEP4HFP: Complex event processing for heart failure prediction", IEEE transactions on nanobioscience, vol.16, no.8, pp.708-17, November 2017.

[9]. Rahul Kumar Sevakula, and Nishchal Kumar Verma,"Assessing Generalization Ability of Majority Vote Point Classifiers", IEEE Transactions on Neural Networks and Learning Systems, vol.28, no. 12, pp.2985–2997, December 2017.

[10]. Sehirli, E., & Alesmaeil, A. (2022). Detecting Face-Touch Hand Moves Using Smartwatch Inertial Sensors and Convolutional Neural Networks. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 122–128. https://doi.org/10.18201/ijisae.2022.275

[11]. Hui Li, Yanying Cui, Wenze Li, Yue Shi, Chao Fing, Hongzhi Li, Ting Gao, Lihong Hu, and Yinghua Lu, "Ensemble learning for overall power conversion efficiency of the all-organic dye-sensitized solar cells", IEEE Access, vol.6, pp.34118-26, June 2018.

[12]. Ibomoiye Domor Mienye, Yanxia Sun, and Zenghui Wang, "An improved ensemble learning approach for the prediction of heart disease risk", Informatics in Medicine Unlocked, vol.20, pp.100402, January 2020.

[13]. Deekshith Shetty, Harshavardhan C A, M Jayanth Varma, Shrishail Navi, Mohammed Riyaz Ahmed, "Diving Deep into Deep Learning: History, Evolution, Types and Applications".

[14]. ZRui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, and Robert X. Gao,"Deep learning and its applications to machine health monitoring", Mechanical Systems and Signal Processing, vol.15, pp.213-37, January 2019.

[15]. U Rajendra Acharya, Oliver Faust, Vinitha Sree, Swapna G, Roshan Joy Martis,

[16]. Nahrizul Adib Kadri, and Jasjit S Suri, "Linear and nonlinear analysis of normal and CAD-affected heart rate signals", Computer methods and programs in biomedicine, vol.113, no.1, pp.55-68, January 2014.

[17]. Ahmad M. Karim, Mehmet S. Guzel, Mehmet R. Tolun, Hilal Kaya, and Fatih V. Çelebi, "A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing", Biocybernetics and Biomedical Engineering, vol.39, no.1, pp.148-159, 2019.

[18]. Rajamhoana SP, Devi CA, Umamaheswari K, Kiruba R, Karunya K, and Deepika R., "Analysis of neural networks based heart disease prediction system", In 2018 11th International Conference on Human System Interaction (HSI), pp.233-239, July 2018.

[19]. Chandrika, L., and K. Madhavi. "A Hybrid Framework for Heart Disease Prediction Using Machine Learning Algorithms." In E3S Web of Conferences, vol. 309. EDP Sciences, 2021.

[20]. Sally Fouad Shady. (2021). Approaches to Teaching a Biomaterials Laboratory Course Online. Journal of Online Engineering Education, 12(1), 01–05. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/43

[21]. Ghosh, Pronab, Sami Azam, Mirjam Jonkman, Asif Karim, FM Javed Mehedi Shamrat, Eva Ignatious, Shahana Shultana, Abhijith Reddy Beeravolu, and Friso De Boer. "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques." IEEE Access 9,19304-19326, 2021.

[22]. Reddy, N. Satish Chandra, Song Shue Nee, Lim Zhi Min, and Chew Xin Ying. "Classification and feature selection approaches by machine learning techniques: Heart disease prediction." International Journal of Innovative Computing, vol. 9, no. 1, 2019.

[23]. Mustafa Jan, Akber A Awan, Muhammad S Khalid, and Salman Nisar, "Ensemble approach for developing a smart heart disease prediction system using classification algorithms", Research Reports in Clinical Cardiology, vol.n9, no.33, 2018.

[24]. Farman Ali, Shaker El-Sappagh, S.M. Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, and Kyung-Sup Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion", Information Fusion, vol.63, pp.208-22, November 2020.

[25]. Qi Zhenya, and Zuoru Zhang, "A hybrid cost-sensitive ensemble for heart disease prediction", 2021.

[26]. Xiao-Yan Gao, Abdelmegeid Amin Ali, Hassan Shaban Hassan, and Eman M. Anwar, "Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method", Complexity, February 2021.

[27]. Bayu Adhi Tama, Sun Im, and Seungchul Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble", BioMed Research International, vol.2020, April 2020.

[28]. Garg, D. K. . (2022). Understanding the Purpose of Object Detection, Models to Detect Objects, Application Use and Benefits. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 01–04. https://doi.org/10.17762/ijfrcsce.v8i2.2066

[29]. Yaxing Zhao, Limsoon Wong, and Wilson Wen Bin Goh, "How to do quantile normalization correctly for gene expression data analyses", Scientific reports, vol.10, no.1, pp.1-11, 2020.

[30]. Patil, P. ., D. D. . Waghole, D. V. . Deshpande, and D. M. . Karykarte. "Sectoring Method for Improving Various QoS Parameters of Wireless Sensor Networks to Improve Lifespan of the Network". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 6, June 2022, pp. 37-43, doi:10.17762/ijritcc.v10i6.5622.

[31]. Safdar, Saima, Saad Zafar, Nadeem Zafar, and Naurin Farooq Khan. "Machine learning based decision support systems (DSS) for heart disease diagnosis: a review." Artificial Intelligence Review 50, no. 4 (2018): 597-623.

[32]. Zhicong Chen, Yixiang Chen, Lijun Wu, Shuying Cheng, and Peijie Lin, “Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions”, Energy Conversion and Management, vol.198, pp.111793, October 2019.

[33]. D. Binu and B. S Kariyappa, “RideNN: A new rider optimization algorithm-based neural network for fault diagnosis in analog circuits”, IEEE Transactions on Instrumentation and Measurement, vol.68, no.1, pp.2-6, May 2018.

[34]. Sakeena Javaid, Muhammad Abdullah, Nadeem Javaid, Tanzeela Sultana, Jawad Ahmed, and Norin Abdul Sattar, “Towards Buildings Energy Management: Using Seasonal Schedules Under Time of Use Pricing Tariff via Deep Neuro-Fuzzy Optimizer”, In2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp.1594-1599, June 2019.

[35]. K. Divya Vani, "Heart Disease Prediction using Long Short-Term Memory (LSTM) Deep Learning Methodology", vol.2, no.8, pp.1076-1081, 2021.

[36]. Muhammad, Yar, Muhammad Tahir, Maqsood Hayat, and Kil To Chong. "Early and accurate detection and diagnosis of heart disease using intelligent computational model." Scientific reports, vol. 10, no. 1, 1-17, 2020.