Neutrosophic Sets and Systems, {Special Issue: Advances in SuperHyperStructures and Applied Neutrosophic Theories)}, Vol. 74, 2024
University of New Mexico

Detection of Cardiovascular Diseases Using Predictive Models Based on Deep Learning Techniques: A Hybrid Neutrosophic AHP-TOPSIS Approach for Model Selection

Julio Barzola-Monteses [1,2,*], Rosangela Caicedo-Quiroz [3], Franklin Parrales-Bravo [1], Cristhian Medina-Suarez [1], Wendy Yanez-Pazmino [4,5], David Zabala-Blanco [6], and Maikel Y. Leyva-Vazquez [7,8]

[1] Artificial Intelligence Research Group, University of Guayaquil, Guayaquil, Ecuador; julio.barzolam@ug.edu.ec; franklin.parralesb@ug.edu.ec; cristhian.medinas@ug.edu.ec
[2] Center for Applied Technology Studies, Bolivarian University of Ecuador, Durán, Ecuador; jjbarzolam@ube.edu.ec
[3] Center for Integrated Care and Health Promotion, Bolivarian University of Ecuador, Durán, Ecuador; rcaicedoq@ube.edu.ec
[4] School of Computer Science, University of Birmingham, Birmingham, United Kingdom; w.yanez@bham.ac.uk
[5] Faculty of Electrical and Computer Engineering, Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, Ecuador; wenpayan@espol.edu.ec
[6] Faculty of Engineering Sciences, Universidad Católica del Maule, Talca, Chile; dzabala@ucm.cl
[7] Artificial Intelligence Research Group, University of Guayaquil, Guayaquil, Ecuador; maikel.leyvav@ug.edu.ec
[8] GIIAR Research Group Universidad Politécnica Salesiana, Guayaquil, Ecuador;

Abstract: In Ecuador and globally, cardiovascular diseases are the leading cause of mortality, accounting for a worrying 26.49% of deaths in 2019. An approach based on deep learning is applied to improve the capacity for early prediction and reduce its incidence. In this work, three different models were proposed and compared: deep neural networks (DNN), convolutional neural networks (CNN), and multilayer perceptron (MLP). Experiments were conducted in two scenarios: one using a dataset that included 12 variables, and another in which the variables were reduced to those most significantly correlated with cardiovascular disease, i.e., 4 variables; both scenarios with 918 clinical records per variable. Using the Neutrosophic AHP-TOPSIS method for model selection, the CNN model trained with the original dataset was identified as the best-performing model among the proposed options. In specific terms, the evaluation metrics of the CNN model were as follows: an accuracy of 92.17%, a sensitivity of 94.51%, a specificity of 90.78%, an F1-Score of 93.30%, and an area under the ROC curve of 90.03%.

Keywords: Heart Disease, Prediction, Convolutional Neural Network, Deep Neural Network, Multilayer Perceptron, Neutrosophic AHP-TOPSIS

1. Introduction

The concept of cardiovascular disease (CVD) refers to any condition that can affect the health of the heart and blood distribution in the human body. These conditions can result in lesions in the arterial vessels, which have the potential to cause difficulties in several essential organs [1]. According to the World Health Organisation (WHO), cardiovascular diseases constitute the most lethal set of conditions affecting the human population, being the leading cause of mortality globally. These pathologies account for more than 37% of all deaths worldwide, and this proportion is expected to continue to increase until 2030 [2]. In the context of Ecuador, statistics for 2019 revealed that these diseases alarmingly constituted 26.49% of all deaths registered in the country [3]. Data obtained from the 2018 STEPS survey sponsored by WHO provided an even more worrying perspective by showing that approximately one-quarter of the Ecuadorian population, i.e. around 25%, present three or more risk factors related to chronic non-communicable diseases [4].

Factors that increase the likelihood of developing cardiovascular disease include family history, smoking, elevated LDL cholesterol levels, hypertension, advanced age, and diabetes [5]. In addition, lifestyle behaviors and medical issues such as sedentary lifestyles, unhealthy diets, obesity, and excessive alcohol consumption also contribute significantly to risk [6]. Assessment of disease severity in patients is performed by a variety of methods, such as stress tests, chest X-rays, CT scans, cardiac MRI, coronary angiography, and electrocardiograms [7]. Accurate and early detection is essential in resource-limited settings to improve treatments and long-term outcomes [8]. Lack of specialists and errors in diagnosis and treatment pose risks, underlining the importance of early detection for effective care [1].

In this article, three different deep learning (DL) models were proposed and compared for assisting healthcare practitioners in the early detection of cardiovascular diseases: deep neural networks (DNN), convolutional neural networks (CNN), and multilayer perceptron (MLP). Datasets with patient medical history such as electrocardiogram information, and risk factors such as age, gender, blood pressure, and cholesterol level, among others, are analyzed. Through the process of training using historical patient data, the model can learn to recognize subtle patterns that may signal an individual's likelihood of developing cardiovascular disease in the future, which will support decision-making by doctors and specialists during the diagnosis and prescription of treatments of CVD.

The rest of the paper is organized as follows. Section 2 summarizes works related to this study. In section 3, a brief explanation of the machine learning models to be applied is presented. Section 4 shows the experimental evaluations and discussion of the results. Finally, Section 5 includes some final comments and directions for future work.

2. Related works

In recent years, there has been a growing interest in using predictive models based on deep learning techniques in the healthcare field [9]. These models offer a promising avenue for improving the accuracy of predictions, particularly in the context of cardiovascular diseases. Through their ability to decipher complex patterns in data, these models present a possible solution for improved diagnosis and prognosis.

Miao et al. [10] explored the application of DNNs in the detection and prediction of coronary heart disease (CHD). Their approach was to improve diagnostic accuracy, especially in regions with limited resources and limited cardiology expertise. By incorporating regularisation techniques and random elimination mechanisms into its DNN architecture, the study yielded a diagnostic accuracy of 83.67%, sensitivity of 93.51%, and specificity of 72.86%, highlighting the feasibility of DNN models in the accurate diagnosis of CHD.

A similar thread is presented in the work of Ibrahim et al. [11] where early identification of acute myocardial infarction (AMI) took center stage. Using a variety of DL models, including CNN, Recurrent Neural Networks (RNN), and XGBoost, the researchers achieved significant accuracy. Their results, with ROC curve values of 90.7%, 82.9%, and 96.5% respectively, underlined the efficacy of these models in the timely classification of AMI.

Joo et al. [12] ventured into the realm of machine learning (ML) and big data in the context of cardiovascular disease prediction. Their approach incorporated attributes such as age, gender, blood pressure, cholesterol levels, smoking habits, and medical history. Using a variety of ML models, including logistic regression, deep neural networks, random forest, and lightGBM, the authors demonstrated that their methodology outperformed conventional approaches, emphasizing the importance of considering medication data.

In a different approach, Zheng et al. [13] presented a pioneering approach based on DNN to predict major adverse cardiovascular events (MACE) in patients with non-ST-segment elevation myocardial infarction (NSTEMI). A comparative analysis with traditional ML algorithms supported the superiority of the DNN model in terms of accuracy, sensitivity, F1 score, and area under the ROC curve was carried out.

These trends were continued by Priya and Thilagamani [14]who used support vector machines (SVM), gated recurrent units (GRU), and CNN, along with pulse wave velocity measurements and features derived from optimization algorithms. Their approach, with area under the ROC curve values of 87.4%, 91.2%, and 88.6%, offered a compelling avenue to identify risk factors for arterial stiffness in diabetic patients.

Sarra et al. [15] proposed an innovative ANN-based diagnostic system for cardiovascular diseases. By achieving a diagnostic accuracy of 93.44%, their ANN model outperformed conventional ML methods, pointing to the potential of ANN as an efficient tool for accurate disease diagnosis.

Similarly, Kishor and Jeberson [16] proposed an LSTMS-DBN approach for predicting cardiovascular events with deep learning using information collected by IoT devices. In particular, Long Short-Term Memory was used to identify and prevent arterial events and Deep Belief Network (DBN) to represent and select more efficient features of the recorded dataset. The LSTM-DBN approach shows an 88.42% mean accuracy compared to other deep learning algorithms including simple RNN, GRU, CNN and Ensemble.

In another work, Khanna et al. [17] presented an IoTDL-HDD model based on IoT and deep learning for the detection of cardiovascular diseases using biomedical electrocardiogram signals. The model uses the BiLSTM feature extraction technique, Artificial Flora optimization (AFO) as a hyperparameter optimizer and Fuzzy Deep Neural Network (FDNN) for disease classification. The IoTDL-HDD model achieves an accuracy of 9.452% in comparison with other models including Deep Neural Network (DNN), Fuzzy Support Vector Machine (FSVM), XGBoost, and others.

Along the same line, Venkatesan et al. [18] designed a cardiac disease diagnosis approach that relies on fuzzy c-means neural network (FNN) and a deep convolutional neural network for improving the accuracy of prediction of the diseases. The results revealed an accuracy of 86.4% and an F1-score of 97 in comparison with convolutional neural network (CNN) and ensemble machine learning (EML).

Likewise, Pan et al. [19] proposed an Enhanced Deep Learning approach assisted by a Convolutional Neural Network (EDCNN) to improve the diagnostics of heart disease. The model is implemented in an Internet of Medical Things Platform (IoMT) and achieved a precision rate of up to 99.1% in comparison with conventional approaches such as Artificial Neural Networks (ANN), DNN, and Recurrent Neural Networks (RNN) among others.

In the same way, Shekhar [20] presented an IoT-centered Deep Learning Modified Neural Network (DLMNN) to assist healthcare practitioners in the effective diagnosis of heart diseases by relying on wearable IoT devices attached to the patient body. The outcome of the conducted experiments shows an improvement in terms of accuracy of 92.59% and security of 95.87% in comparison with existing algorithms.

Notably, Hussain et al. [21] proposed a unique 1D CNN architecture to improve patient classification, obtaining an impressive training accuracy of 98.9% and testing accuracy of 90.32%, while Sharma et al. [22] moved into the realm of risk prediction for patients, achieving an accuracy of 71.4% when classifying patients with diabetes and hypertension, two key risk factors for heart disease.

In summary, these studies follow a logical sequence of advances in the use of ML and DL models for the detection and prediction of cardiovascular disease. Each research builds on the previous one, showing how these advanced techniques are transforming medical care towards a more precise and personalized approach. While challenges remain to be addressed, these advances demonstrate a promising path toward a more effective medical future.

3. Method

This section breaks down the process into two parts. The first subsection details the systematic procedure used to process the clinical information used during the training and testing phase of the cardiovascular disease predictive model. The second subsection outlines the proposed architectures under the predictive models by leveraging the scikit-learn and Keras libraries and describes the methods used for the collection and thorough evaluation of the performance of each model.

3.1. Experimental setup

3.1.1. Data

A dataset hosted on the Kaggle platform was used to train the model. This dataset consists of a total of 918 records and 12 columns corresponding to parameters related to cardiovascular disease risk factors [23]. These factors are used to assess heart diseases in patients and determine their presence and risk. As mentioned above, the dataset consists of 12 attributes and each of them has a different data type. Table 1 describes in detail the attributes of the dataset used.

Given that there were some categorical variables in the dataset used, it was decided to use the mapping process to obtain descriptive statistics for the categorical columns. In this process, a unique numerical value is assigned to each category or label present in an object-type variable.

In summary the features present in the dataset were as follows: the average age of the individuals is around 53.51 years; information on the gender of the individuals is included; the attribute "Chest Pain Type" is categorized into levels (ASY:0, ATA:1, NAP:2, TA:3); the resting blood pressure averages around 132.40 mmHg; the average cholesterol level is around 198.80 mg/dL; the variable "Fasting Glucose" is binary (0 or 1) and denotes if fasting blood glucose exceeds 120 mg/dL; the "Resting Electrocardiogram" results are recorded in three types (LVH:0, Normal:1 and ST:2); the maximum heart rate reached during exercise averages about 136.81 beats per minute; the presence of "Exercise Induced Angina" is indicated; the attribute "ST Segment Depression" is measured relative to rest, with an average of about 0.89; the "ST Segment Slope" reflects its variation during exercise (Up:0, Flat:1 and Down:2); finally, the target variable "Heart Disease" indicates whether there is presence or absence of heart disease.

Table 1. The twelve attributes and descriptions used for the development of the Deep Learning-based models.

Attribute	Type	Description
Age	Numeric	Patient's age in years
Sex	Categorical	Sex of patient (M: Male, F: Female)
ChestPainType	Categorical	Type of chest pain (TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic)
RestingBP	Numeric	Resting blood pressure in mm Hg
Cholesterol	Numeric	Serum cholesterol level in mm/dl
FastingBS	Numeric	Fasting blood sugar level (1: if > 120 mg/dl, 0: otherwise)
RestingECG	Categorical	Resting electrocardiogram results (Normal, ST, LVH)
MaxHR	Numeric	Maximum heart rate reached (numerical value between 60 and 202)
ExerciseAngina	Categorical	Exercise-induced angina (Y: Yes, N: No)
Oldpeak	Numeric	ST-segment depression (numeric value)
ST_Slope	Categorical	ST-segment slope (Up: Ascending, Flat: Flat, Down: Descending)
HeartDisease	Numeric	Diagnosis of heart disease (1: Sick, 0: Healthy)

Based on the correlation matrix presented in Figure 1, it can be seen there are strongly correlated variables. On the one hand, variables such as ExerciseAngina, Oldpeak, and ST_Slope show moderate to strong positive correlations concerning the presence of heart disease. This suggests that when these values increase, there is a greater likelihood that the patient has heart disease. On the other hand, variables such as MaxHR, ChestPainType, and Sex show moderate negative correlations with heart disease, indicating that higher values for these variables are associated with a lower likelihood of heart disease.

Based on the analysis of the correlation matrix [24], it was recommended to create a new dataset using the original dataset, named "Dataset B", including only the variables ExerciseAngina, Oldpeak, ST_Slope and the target variable (HeartDisease). This action sought to reduce dimensionality and focus on the most influential characteristics for the analysis. Comparing the performance of the models on both datasets provided a clearer perspective on whether dimensionality reduction negatively affected predictive ability, which was a crucial step in ensuring the robustness and effectiveness of the model in line with the objectives of the analysis.

In terms of the training, validation, and testing process, it was considered due to the importance of splitting the dataset into separate sets to avoid over-fitting the model and to assess its performance in real-life situations. This was achieved through a function called "train_val_test_split", which partitioned the data into training (70%), validation (15%), and testing (15%) [25]. The training set allowed the model to learn patterns, while the validation set was used to adjust hyperparameters and avoid overfitting. The test set, which contained previously unseen data, evaluated the final effectiveness of the model.

Another crucial aspect was featuring scaling, which was carried out using the technique of scaling by standardization [26]. Using the StandardScaler method, numerical features were given a mean of zero and a standard deviation of one. This process is applied to each attribute of the dataset.

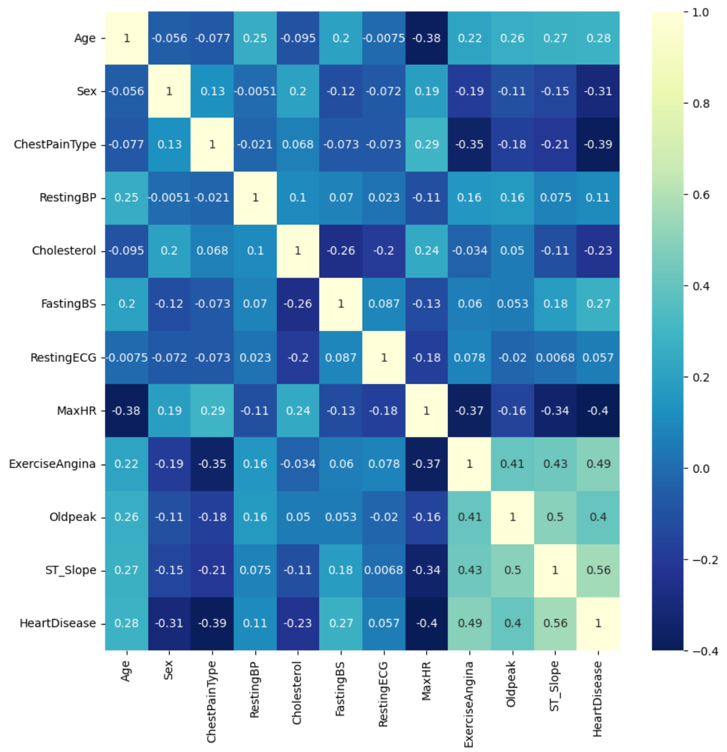

Figure 1. Correlation matrix of the variables of the dataset analyzed.

3.1.2. Model features and target

The experimental design addresses a multivariate issue in both cases using the original dataset and dataset B. The case of the original dataset is characterized by eleven input variables and one output variable. The case of dataset B is characterized by three input variables and one output variable. In both cases, the output variable takes binary predictive values to classify the diagnosis of heart disease.

3.1.3. Problem definition

In the current scientific paper, the problem definition is the following: based on age, sex, chestPainType, restingBP, cholesterol, fastingBS, restingECG, maxHR, exerciseAngina, oldpeak, ST_Slope; can it be determined whether he/she has heart disease?

3.1.4. Hyperparameters

Refers to parameters external to the model itself, and their values typically can't be derived from the training data. These values are set by the model designer to fine-tune the learning algorithms. In this study, a grid search algorithm has been employed to optimize the models. This involved an iterative process where hyperparameters were adjusted to identify configurations that resulted in the best generalization on the test dataset. The grid search procedure explored all possible combinations of hyperparameters and selected the most effective subset. This procedure will be applied to the best model selected during the validation.

3.1.5. Model evaluation.

Evaluation models are applied using the following procedure:

Training procedure is applied to estimate model performance quickly.

K-fold cross-validation is applied to get a more robust estimate of how the model generalizes. K=5 partitions are considered.

The selected model resulting from the above processes is applied hyperparameters tuning to improve performance metrics.

3.1.6. Performance evaluation

In literature reviews focusing on binary classification performance, the evaluation of models often involves the use of several standard metrics. These metrics include the false negative rate (FNR), false positive rate (FPR), true negative rate (TNR), true positive rate (TPR), positive predictive value (PPV), and accuracy (ACC). These metrics are commonly employed to assess and select models with the lowest predictive errors [27]. To compute these metrics, the following equations are used:

$$FNR = \frac{\text{FN}}{\text{TP} + \text{FN}} \tag{1}$$

$$FPR = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{2}$$

$$specificity = TNR = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{3}$$

$$recall = sensitivity = TPR = \frac{TP}{\text{TP} + \text{FN}} \tag{4}$$

$$precision = PPV = \frac{TP}{\text{TP} + \text{FP}} \tag{5}$$

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

$$acc = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{7}$$

Where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative.

3.2. Model design

The definition of model architecture was a fundamental step in this work. Three main models were selected to be used and compared to obtain the best performance in cardiovascular disease prediction. Based on state-of-the-art information and analysis, the following models were chosen: Deep Neural Network (DNN), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN). These models are widely recognized for their effectiveness in classification problems and have distinctive characteristics.

3.2.1. Deep Neural Network (DNN)

The DNN is a DL architecture that incorporates multiple hidden layers. Each layer contains neurons that process and transmit information through the network [28], ultimately arriving at the output layer as shown in Figure 2(a).

The model structure starts with a flattening layer to transform the input data into a one-dimensional shape. Dropout layers are applied at a rate of 20% to avoid overfitting by randomly deactivating neurons during training. The model incorporates dense layers, also called fully connected layers, which employ the ReLU activation function, and a "he_normal" weight initialization designed for ReLU. The model consists of three dense layers, with different numbers of units (300, 100, and 10), allowing the network to learn more complex representations. Between each dense layer, a dropout layer is used to regularize the model and prevent overfitting. The last dense layer has a single unit and a sigmoid activation function, suitable for binary classification.

3.2.2. Multilayer Perceptron (MLP)

It is a neural network architecture that includes multiple hidden layers [29]commonly used in regression and classification problems in ML [30]. This model, designed as a sequential network is specifically adapted to binary classification. Its structure comprises a flattening layer, followed by two dense layers with ReLU activation, each preceded by dropout layers (20%). The last layer, dense and with sigmoid activation, consists of a single output unit as shown in Figure 2(b).

2.2.3. Convolutional Neural Network (CNN)

An initial two-dimensional architecture was chosen for a CNN, using 3x3 cores. Performance improvements were then sought through a more parameter and computationally efficient structure. The final architecture is shown in Figure 2(c).

The proposed architecture for the CNN includes several key layers: the first, a reshape layer, transforms the input data from (11,) to (11, 1), adding an extra dimension common in sequential data; then, two 1D convolutional layers follow this reshape layer, the first with 40 filters and size 5 convolutions along with ELU activation function, and the second with 8 filters and also ELU activation, allowing the network to learn important features in the sequential data; a flattening layer converts the output data from the convolutional layers to one-dimensional format, ready to be processed by the dense layers; the last two layers are dense, the first with 80 units and ELU activation, and the last one is an output dense layer with a single unit and sigmoidal activation, thus setting up the model for binary classification and prediction of probability of belonging to the positive or negative classes.

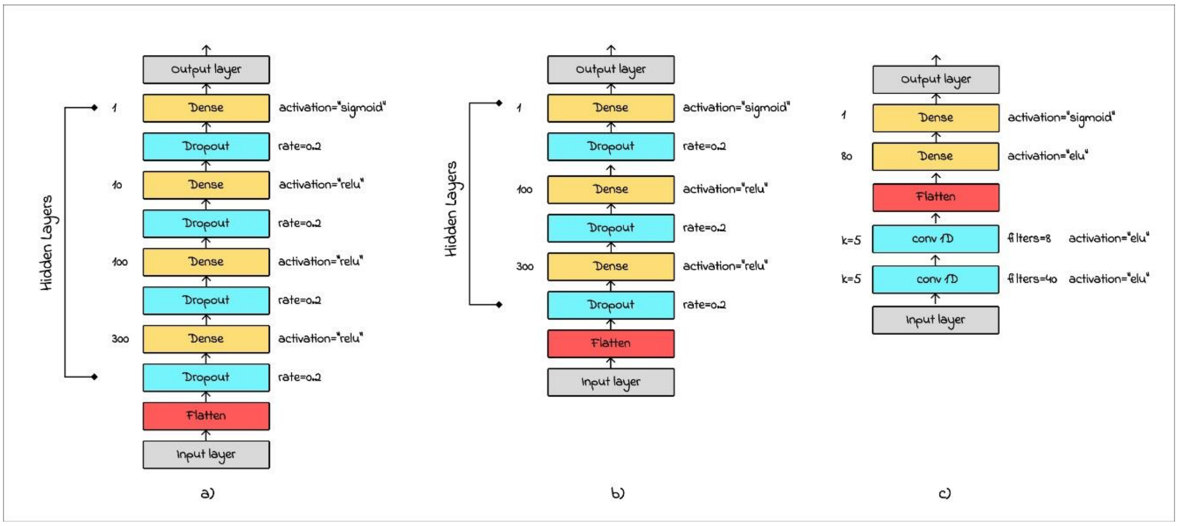

Figure 2. The architecture of the models considered: a) DNN model, b) MLP model, and c) CNN model.

4. Results and Discussion

This section presents a detailed analysis of the results obtained by validating and optimizing the proposed models for cardiovascular disease prediction. The model selection was based on a comprehensive comparison of the performance of three different models: DNN, MLP, and CNN. The models were evaluated on two different datasets, designated as the original dataset and B (reduced variable dataset).

4.1. DNN model

The compilation and training of the DNN models were carried out by configuring the Adam optimizer with a learning rate of 0.001. For the loss function "binary_crossentropy" was chosen, suitable for binary classification, a batch size of 32 (batch_size) and 50 epochs were determined to process the data.

The training results were evaluated using the confusion matrix, showing how the model classified the samples. The matrix reflected a higher number of correctly identified records in the original dataset compared to dataset B. A characteristic of the ROC curve is that if it is very close to the upper left corner, it implies a good performance of the model. Figure 3 shows the Classification DNN model's false positive and true positive rates using the ROC plot.

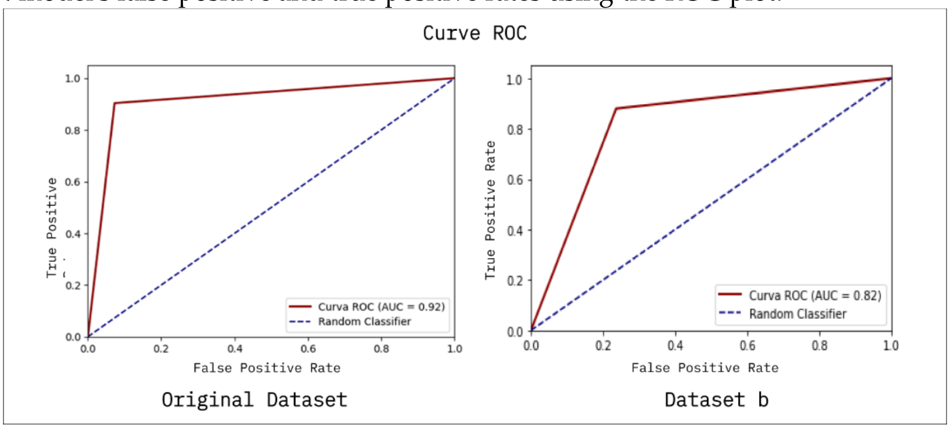

Figure 3. ROC of scores of DNN model

Once the metrics have been evaluated with the corresponding DNN model, the following Table 2 is obtained:

Table 2. DNN Evaluation Metrics

Metrics	Original Dataset [%]	Dataset B [%]
acc	91.30	83.33
precision	94.93	84.88
sensitivity	90.36	87.95
specificity	92.72	76.36
F1	92.59	86.39

The previous table shows that the DNN Classification model obtained an accuracy percentage of 91.30% by correctly classifying 75 data as negative cases, representing 92.72% specificity. As positive cases, 51 were obtained, representing 90.36% sensitivity. It is noteworthy that in this model analyzed, better results have been achieved using the original dataset.

4.2. MLP model

With a learning setting of 0.01, the model compiler was set up using the Adam optimizer. The loss function "binary_crossentropy" was chosen, which is suitable for the evaluation of predictions in this type of task. Following the same approach, a batch size of 32 samples and 50 epochs was set for the training process.

The training results of the MLP model were analyzed through the confusion matrix. In this MLP model, it is still maintained that better results are obtained with the original dataset. Figure 4 and Table 3 corroborate this observation, although the ROC curve and the performance metrics showed that the model had high accuracy.

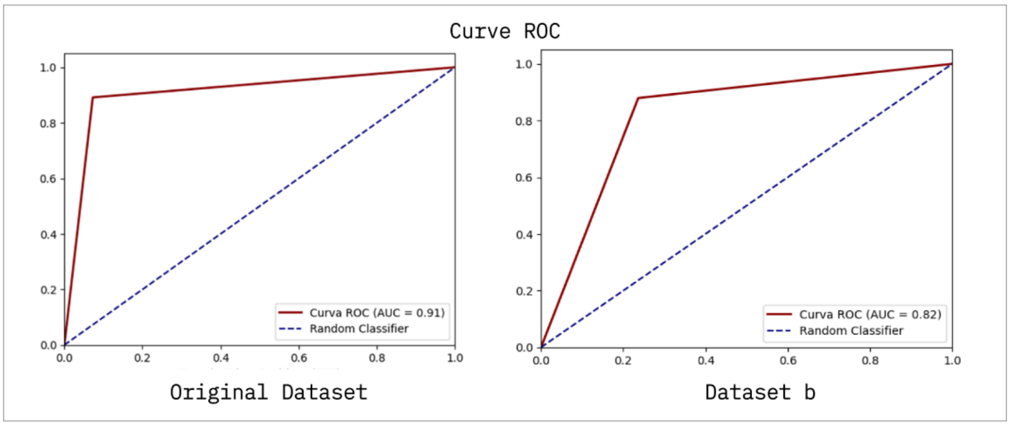

Figure 4. ROC of scores of MLP model

Table 3. MLP Evaluation Metrics

Metrics	Original Dataset [%]	Dataset B [%]
acc	90.5	83.33
precision	94.8	87.95
sensitivity	89.1	87.95
specificity	92.7	76.36
F1	91.92	86.39

The previous table shows that the MLP Classification model obtained an accuracy percentage of 90.50% by correctly classifying 74 data as negative cases, representing 92.70% specificity. As positive cases, 51 were obtained, representing 89.1% sensitivity. However, it can be noted that if we compare the DNN model and the MLP model, slightly better results are achieved with the DNN model. That is, preliminarily some neural network complexity contributes to slightly better prediction results.

4.3 CNN model

The optimizer is defined using the Adam algorithm with a learning rate of 0.0001. The model is then compiled using the loss function "binary_crossentropy", which is effective in minimizing loss during training. The model is trained for 50 epochs with the training data and evaluated with the validation set to get a complete picture of its performance.

The training results of the CNN model were analyzed through the confusion matrix. In this CNN model, it is still corroborated that better results are obtained with the original dataset. Figure 5 and Table 4 confirmed this observation, although the ROC curve and the performance metrics showed that the model had high accuracy.

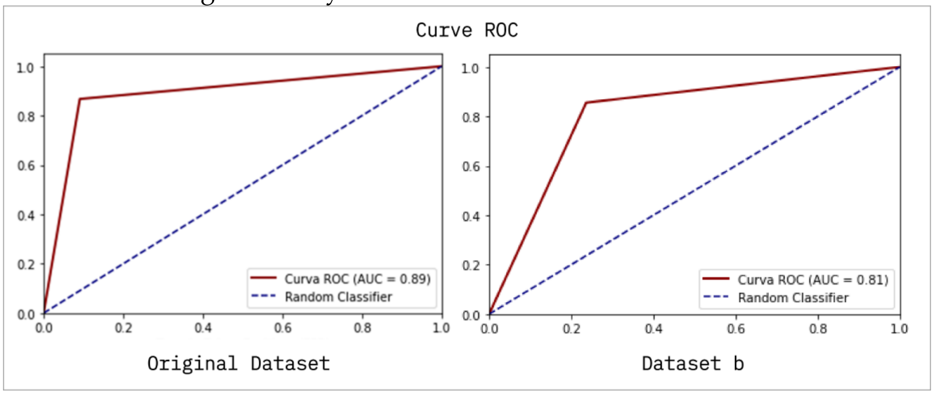

Figure 5. Confusion matrix of CNN training stage

Table 4. CNN Evaluation Metrics

Metrics	Original Dataset [%]	Dataset B [%]
acc	88.40	81.88
precision	93.50	84.52
sensitivity	86.74	85.54
specificity	90.90	76.36
F1	90.00	85.02

The previous table shows that the CNN classification model obtained an accuracy percentage of 88.40% by correctly classifying 72 data as negative cases, representing 90.90% specificity. As positive cases, 50 were obtained, representing 86.74% sensitivity. The evaluation metrics in Table 4 indicate good binary classification performance, with high precision, specificity, and F1 Score.

4.4. Selecting the best model

Table 5 provides a comprehensive overview of the performance of the three implemented models on two different datasets, designated as A and B, where A refers to the original dataset and B to the reduced dataset of variables. These results were obtained by applying K-fold cross-validation with K=5 partitions.

For each model and dataset, binary classification performance metrics are evaluated to measure its prediction ability with the heart disease variable. In terms of precision and accuracy, the best results with the CNN model are obtained using the original dataset (A), indicating its ability to predict positive cases with a low margin of error.

Table 5. Model selection using cross-validation

Model	Dataset	Accuracy	Precision	Recall	Specificity	F1-S	Roc-Auc
DNN	A	84.88%	84.09%	88.93%	80.21%	86.33%	**91.82%**
	B	85.17%	83.99%	90.69%	79.90%	87.17%	89.94%
MLP	A	83.80%	82.60%	88.35%	78.51%	85.32%	91.21%
	B	85.17%	83.66%	**91.27%**	79.24%	**87.26%**	90.07%
CNN	A	**85.20%**	**85.18%**	87.73%	**82.51%**	86.35%	88.82%
	B	83.32%	81.58%	89.14%	76.69%	85.10%	80.95%

The analysis highlights that dataset B benefits the sensitivity and F1-score of the MLP model, while the CNN model exhibits superior performance on most metrics compared to the DNN and MLP models. However, the choice of the optimal model must consider the context of the problem. A key point to note in Figure 6 is that the DNN and MLP models achieve a slightly higher value in true positives, with 51 positive values according to the confusion matrix, in contrast to the 50 positive values of the CNN model.

For selecting the model, the integration Neutrosophic Hierarchical Analytical Process and TOPSIS (NAHP-TOPSIS) [31,32] methods are used. In this case single valued neutrosophic trapezoidal numbers are uses the single-valued trapezoidal neutrosophic number,, is a neutrosophic set on, whose truth, indeterminacy and falsehood membership functions are defined as follows, respectively [33, 34]: $\tilde{a} = \langle(a_1, a_2. a_3, a_4); \alpha_{\tilde{a}}, \beta_{\tilde{a}}, \gamma_{\tilde{a}}\rangle\mathbb{R}$

$$T_{\tilde{a}}(x) = \begin{cases} \alpha_{\tilde{a}}\left(\frac{x-a_1}{a_2-a_1}\right), & a_1 \le x \le a_2 \\ \alpha_{\tilde{a}}, & a_2 \le x \le a_3 \\ \alpha_{\tilde{a}}\left(\frac{a_3-x}{a_3-a_2}\right), & a_3 \le x \le a_4 \\ 0, \text{ otherwise} \end{cases} \tag{8}$$

$$I_{\tilde{a}}(x) = \begin{cases} \frac{(a_2-x+\beta_{\tilde{a}}(x-a_1))}{a_2-a_1}, & a_1 \le x \le a_2 \\ \beta_{\tilde{a}}, & a_2 \le x \le a_3 \\ \frac{(x-a_2+\beta_{\tilde{a}}(a_3-x))}{a_3-a_2}, & a_3 \le x \le a_4 \\ 1, & \text{otherwise} \end{cases} \tag{9}$$

$$F_{\tilde{a}}(x) = \begin{cases} \frac{(a_2-x+\gamma_{\tilde{a}}(x-a_1))}{a_2-a_1}, & a_1 \le x \le a_2 \\ \gamma_{\tilde{a}}, & a_2 \le x \le a_3 \\ \frac{(x-a_2+\gamma_{\tilde{a}}(a_3-x))}{a_3-a_2}, & a_3 \le x \le a_4 \\ 1, & \text{otherwise} \end{cases} \tag{10}$$

Where, and. $\alpha_{\tilde{a}}, \beta_{\tilde{a}}, \gamma_{\tilde{a}} \in [0,1]$ $a_1, a_2, a_3, a_4 \in \mathbb{R} a_1 \le a_2 \le a_3 \le a_4$

For simplicity, we use the linguistic scale of trapezoidal neutrosophic numbers, see Table 1 and also compare with the scale defined in [35].

Table 6. Saaty's scale translated to a neutrosophic trapezoidal scale.

Saaty's scale	Definition	Neutrosophic Triangular Scale
1	Equally influential	$\tilde{1} = \langle(1,1,1); 0.50, 0.50, 0.50\rangle$
3	Slightly influential	$\tilde{3} = \langle(2,3,4); 0.30, 0.75, 0.70\rangle$
5	Strongly influential	$\tilde{5} = \langle(4,5,6); 0.80, 0.15, 0.20\rangle$

7	Very strongly influential	$\tilde{7} = \langle(6,7,8); 0.90, 0.10, 0.10\rangle$
9	Absolutely influential	$\tilde{9} = \langle(9,9,9); 1.00, 1.00, 1.00\rangle$
2, 4, 6, 8	Sporadic values between two close scales	$\tilde{2} = \langle(1,2,3); 0.40, 0.65, 0.60\rangle$ $\tilde{4} = \langle(3,4,5); 0.60, 0.35, 0.40\rangle$ $\tilde{6} = \langle(5,6,7); 0.70, 0.25, 0.30\rangle$ $\tilde{8} = \langle(7,8,9); 0.85, 0.10, 0.15\rangle$

Each metric with the others in terms of relative importance in the context of healthcare pair-wise comparison matrix is obtained.

Table 7. Pairwise Comparison Matrix of Criteria

	Accuracy	**Precision**	**Recall**	**Specificity**	**F1-Score**	**ROC-AUC**
Accuracy	$\tilde{1}$	$1/\tilde{2}$	$\tilde{3}$	$\tilde{2}$	$1/4$	$\tilde{1}$
Precision	$\tilde{2}$	$\tilde{1}$	$\tilde{2}$	$1/\tilde{3}$	$1/\tilde{4}$	2
Recall	$1/\tilde{3}$	$1/\tilde{2}$	$\tilde{1}$	$1/\tilde{2}$	$1/\tilde{2}$	$1/\tilde{3}$
Specificity	$1/\tilde{2}$	$\tilde{3}$	$\tilde{2}$	$\tilde{1}$	$1/\tilde{2}$	$1/\tilde{2}$
F1-Score	$\tilde{4}$	$\tilde{4}$	$\tilde{2}$	$\tilde{2}$	$\tilde{1}$	$\tilde{4}$
ROC-AUC	$\tilde{1}$	$1/\tilde{2}$	$\tilde{3}$	$\tilde{2}$	$1/\tilde{4}$	$\tilde{1}$

The weights derived from the AHP technique signify the significance of each parameter in the identification of cardiovascular disease. ROC-AUC (0.43) and F1-Score (0.35) possess the greatest significance, underscoring their essential functions in facilitating precise class differentiation and harmonizing precision with recall. Specificity (0.149) and Accuracy (0.143) underscore the necessity for dependable recognition of real negatives and overall precision. Meanwhile, Precision (0.141) and Recall (0.075) are of lesser significance yet are crucial for reducing diagnostic inaccuracies. These priorities correspond with the objective of attaining accurate and reliable forecasts in healthcare.

For selectin the model the original dataset result "A" is used then TOPSIS methods is applied. Based on the TOPSIS analysis using the updated AHP weights, the models are ranked as follows:

CNN (TOPSIS Score: 0.67),

DNN (TOPSIS Score: 0.62),

MLP (TOPSIS Score: 0.28).

The CNN model ranks highest due to its strong performance in key metrics such as ROC-AUC and F1-Score, which carry the most weight in this context. These metrics highlight the model's ability to balance precision and recall while maintaining robust class discrimination, crucial for accurate cardiovascular disease detection. The DNN model closely follows competitive performance, whereas the MLP model, with a lower score, demonstrates less alignment with the prioritized metrics.

Aditionaly in this work, hyperparameters were optimized to improve the forecasting metrics.

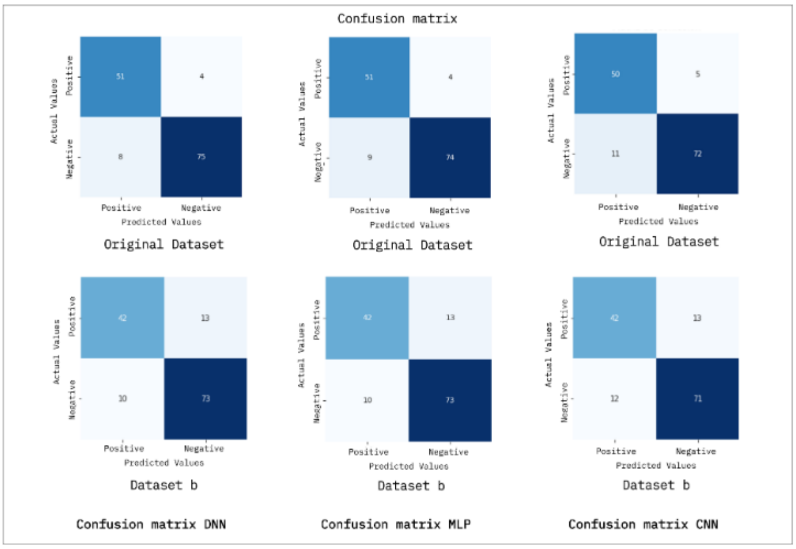

Figure 6. Confusion matrix DNN, MLP, and CNN models

4.5. Optimization of the best model

Hyperparameter optimization involves modifying different values in the model, such as the learning_rate, the number of hidden_layers, and the batch_size, among others. These additional settings ensure that the chosen model is the most suitable for achieving the work objectives and obtaining accurate and reliable results [36]. For this task, the Keras tuner library was used, recommended for its usefulness in the automated search of hyperparameters.

Table 6. Hyperparameters considered for the tuning of the CNN model

Hyperparameters	Values	Description
hp_units	min_value=8, max_value=256, step=8	Number of neurons
hp_layers	min_value=1, max_value=5, step=1	Number of hidden layers
hp_filters	min_value=2, max_value=40, step=2	Number of filters
hp_activation	['relu', 'sigmoid', 'tanh', 'softmax', 'elu'].	Activation function
hp_epochs	min_value=10, max_value=50, step=10	Number of epochs
hp_batch_size	min_value=4, max_value=40, step=4	Lot size used
hp_learning_rate	values=[$1e^{-1}$,$1e^{-2}$,$1e^{-3}$,$1e^{-4}$]	Learning rate

Table 6 shows the hyperparameters considered for the fitting and evaluation of the CNN model. These include the number of neurons, hidden layers, filters, activation function, training epochs, batch size, and learning rate. All these hyperparameters were implemented in a single function, called "model_builder", which defines the architecture, and compiles and trains the model.

During the configuration of the experimental design, CNN structures were executed, obtaining as the best topology the following model with hyperparameters: hp_units = 120, hp_layers = 2, hp_filters = 32, hp_activation = relu, hp_epochs = 50, hp_batch_size = 32 and hp_learning_rate = 0.0001.

Figure 7 and Table 7 show the contrast between the selected CNN model and the CNN model with its tuned hyperparameters. A considerable improvement is evident with the CNN model whose hyperparameters were adjusted to the original dataset. The confusion matrix shows an improvement in the prediction of negative cases of heart disease and a reduction of false negatives.

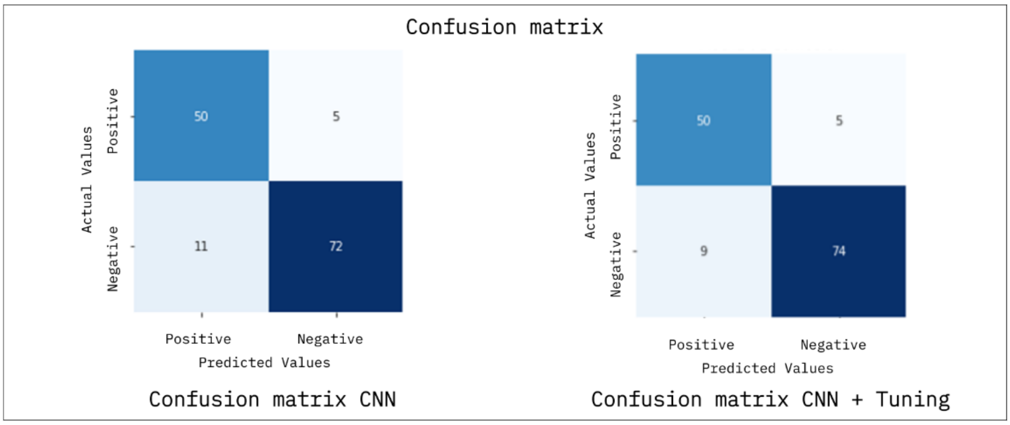

Figure 7. Comparison of the confusion matrix CNN and CNN Tuning

Table 7. Model selection by hyperparameter adjustment.

Model	Accuracy	Precision	Recall	Specificity	F1-Score	Roc-Auc
CNN	85.20%	85.18%	87.73%	82.51%	86.35%	88.82%
CNN + Tuning	92.85%	92.17%	94.51%	90.78%	93.30%	90.03%

The prediction of positive cases is maintained in both models. In the model with optimized hyperparameters all classification performance metrics exceed 90%, the highest of all with 94.51% being the sensitivity metric, i.e., the true positive rate.

5. Conclusions

This study applied deep learning models to predict whether a patient has heart disease. Experiments were conducted using two datasets with several cardiovascular disease risk factors. The original dataset with eleven risk factors and the reduced dataset with three risk factors considered statistical correlation were used. DNN, MLP, and CNN models were set up, trained, validated, and tested. Binary classification performance was used to select the best model for each type of architecture.

In all the models analyzed, the best results were obtained using the attributes and data from the original dataset, i.e., reducing the dataset by considering the statistical correlation of variables had no major influence on the results for this particular case study.

The best result in the general context of the binary forecasting metrics was achieved by the CNN model. This was ratified by applying an optimization with the hyperparameters of the model, achieving metrics that exceeded 90% performance.

The results of this study could serve as a baseline for other studies with machine learning models and recurrent neural networks, to further explore the binary forecasting performance of convolutional neural networks, which are generally used for image and video engineering problems. Another future study could consider the use of a dataset with a greater length and data density to test the incidence of the dataset size and number of variables considering reduction techniques applied in this area of health. Finally, future work would be to integrate the proposed model into a web or mobile app to support doctors in the diagnosis of cardiovascular. Moreover, future studies could explore the broader application of neutrosophic logic in the model selection process to handle uncertainty and indeterminacy more effectively, potentially enhancing decision-making in similar healthcare applications.

Funding: This work has been funded by University of Guayaquil through grant number FCI-039-2023.

Acknowledgments: The authors kindly acknowledge the support from University of Guayaquil. Computational and physical resources were provided by Universidad Bolivariana del Ecuador.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ruiz-Mori, E. *Riesgo y Prevención Cardiovascular*; 1st ed.; Lima, 2014; ISBN 978-612-00-1509-4.
2. OMS Enfermedades Cardiovasculares Available online: https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed on 24 August 2023).
3. MSP-Ecuador MSP Previene Enfermedades Cardiovasculares Con Estrategias Para Disminuir Los Factores de Riesgo Available online: https://www.salud.gob.ec/msp-previene-enfermedades-cardiovasculares-con-estrategias-para-disminuir-los-factores-de-riesgo/ (accessed on 24 August 2023).
4. WHO STEPS 2018 - Ecuador Available online: https://extranet.who.int/ncdsmicrodata/index.php/catalog/774 (accessed on 23 January 2023).
5. Nabel, E.G.; Braunwald, E. A Tale of Coronary Artery Disease and Myocardial Infarction. *New England Journal of Medicine* **2012**, *366*, 54–63, doi:10.1056/nejmra1112570.
6. Doughty, K.N.; Del Pilar, N.X.; Audette, A.; Katz, D.L. Lifestyle Medicine and the Management of Cardiovascular Disease. *Curr Cardiol Rep* **2017**, *19*, 1–10, doi:10.1007/S11886-017-0925-Z/METRICS.
7. Feldman, D.I.; Latina, J.; Lovell, J.; Blumenthal, R.S.; Arbab-Zadeh, A. Coronary Computed Tomography Angiography in Patients with Stable Coronary Artery Disease. *Trends Cardiovasc Med* **2022**, *32*, 421–428, doi:10.1016/J.TCM.2021.08.009.
8. Thomford, N.E.; Bope, C.D.; Agamah, F.E.; Dzobo, K.; Owusu Ateko, R.; Chimusa, E.; Mazandu, G.K.; Ntumba, S.B.; Dandara, C.; Wonkam, A. Implementing Artificial Intelligence and Digital Health in Resource-Limited Settings? Top 10 Lessons We Learned in Congenital Heart Defects and Cardiology. *https://home.liebertpub.com/omi* **2020**, *24*, 264–277, doi:10.1089/OMI.2019.0142.
9. Saber, H.; Somai, M.; Rajah, G.B.; Scalzo, F.; Liebeskind, D.S. Predictive Analytics and Machine Learning in Stroke and Neurovascular Medicine. *https://doi.org/10.1080/01616412.2019.1609159* **2019**, *41*, 681–690, doi:10.1080/01616412.2019.1609159.
10. Miao, K.H.; Miao, J.H. *Coronary Heart Disease Diagnosis Using Deep Neural Networks*; 2018; Vol. 9;.
11. Ibrahim, L.; Mesinovic, M.; Yang, K.-W.; Eid, M.A. Explainable Prediction of Acute Myocardial Infarction Using Machine Learning and Shapley Values. *IEEE Access* **2020**, *8*, 210410–210417, doi:10.1109/ACCESS.2020.3040166.
12. Joo, G.; Song, Y.; Im, H.; Park, J. Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea). *IEEE Access* **2020**, *8*, doi:10.1109/ACCESS.2020.3015757.
13. Zheng, H.; Sherazi, S.W.A.; Son, S.H.; Lee, J.Y. A Deep Neural Network Model for the Prediction of Major Adverse Cardiovascular Event Occurrences in Patients with Non-ST-Elevation Myocardial Infarction. In *Frontiers in Artificial Intelligence and Applications*; 2021; Vol. 340.
14. Priya, A.M.; Thilagamani, S.T. Prediction of Arterial Stiffness Risk in Diabetes Patients through Pulse Wave Velocity and Deep Learning Techniques. *Information Technology and Control* **2022**, *51*, 678–691, doi:10.5755/j01.itc.51.4.31641.

15. Sarra, R.R.; Dinar, A.M.; Mohammed, M.A. Enhanced Accuracy for Heart Disease Prediction Using Artificial Neural Network. *Indonesian Journal of Electrical Engineering and Computer Science* **2023**, *29*, 375–383, doi:10.11591/IJEECS.V29.I1.PP375-383.
16. Dami, S.; Yahaghizadeh, M. Predicting Cardiovascular Events with Deep Learning Approach in the Context of the Internet of Things. *Neural Comput Appl* **2021**, *33*, 7979–7996, doi:10.1007/s00521-020-05542-x.
17. Khanna, A.; Selvaraj, P.; Gupta, D.; Sheikh, T.H.; Pareek, P.K.; Shankar, V. Internet of Things and Deep Learning Enabled Healthcare Disease Diagnosis Using Biomedical Electrocardiogram Signals. *Expert Syst* **2023**, *40*, doi:10.1111/exsy.12864.
18. Venkatesan, M.; Lakshmipathy, P.; Vijayan, V.; Sundar, R. Cardiac Disease Diagnosis Using Feature Extraction and Machine Learning Based Classification with <scp>Internet of Things</Scp> (IoT). *Concurr Comput* **2022**, *34*, doi:10.1002/cpe.6622.
19. Pan, Y.; Fu, M.; Cheng, B.; Tao, X.; Guo, J. Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform. *IEEE Access* **2020**, *8*, 189503–189512, doi:10.1109/ACCESS.2020.3026214.
20. Sarmah, S.S. An Efficient IoT-Based Patient Monitoring and Heart Disease Prediction System Using Deep Learning Modified Neural Network. *IEEE Access* **2020**, *8*, 135784–135797, doi:10.1109/ACCESS.2020.3007561.
21. Hussain, S.; Nanda, S.K.; Barigidad, S.; Akhtar, S.; Suaib, M.; Ray, N.K. Novel Deep Learning Architecture for Predicting Heart Disease Using CNN. In Proceedings of the 2021 19th OITS International Conference on Information Technology (OCIT); IEEE, December 2021; pp. 353–357.
22. Sharma, R.; Gupta, S.; Garg, P. Model for Predicting Cardiac Health Using Deep Learning Classifier. In Proceedings of the 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT); IEEE, July 2022; pp. 25–30.
23. fedesoriano Heart Failure Prediction Dataset Available online: https://www.kaggle.com/fedesoriano/heart-failure-prediction (accessed on 24 July 2023).
24. Melo, E.; Peñafiel, J.; Barzola-Monteses, J.; Espinoza-Andaluz, M. An Initial Approach About Data Preprocessing Techniques Applied to Polymer Electrolyte Fuel Cells: A Case Study. In Proceedings of the International Conference on Science, Technology and Innovation for Society - CITIS2021; Rocha, Á., López-López, P.C., Salgado-Guerrero, J.P., Eds.; Springer Singapore: Singapore, 2022; Vol. 252, pp. 53–64.
25. Ahmed, F.; Lee, J.W.; Samantasinghar, A.; Kim, Y.S.; Kim, K.H.; Kang, I.S.; Memon, F.H.; Lim, J.H.; Choi, K.H. SperoPredictor: An Integrated Machine Learning and Molecular Docking-Based Drug Repurposing Framework With Use Case of COVID-19. *Front Public Health* **2022**, *10*, doi:10.3389/fpubh.2022.902123.
26. Carvalho, A.J.S.; Carr, C.N.; Silva, E.L. da S.; Pereira, L.V.M.; Ferreira Junior, U. da L. Patterns in Data Modeling: Addressing Common Patterns and Best Practices. In *INNOVATION IN HEALTH RESEARCH ADVANCING THE BOUNDARIES OF KNOWLEDGE*; Seven Editora, 2024.
27. Barzola-Monteses, J.; Caicedo-Quiroz, R.; Espinoza-Andaluz, M.; Bajaña-Díaz, E.; Loja-Yagual, R. Prognostic Precision for Crohn ′ s Disease Patients through Machine Learning Predictive Models. In Proceedings of the 42nd IEEE International Conference of Chilean Computer Science Society; IEEE: Concepciòn, Chile, 2023; pp. 5–10.
28. Cichy, R.M.; Kaiser, D. Deep Neural Networks as Scientific Models. *Trends Cogn Sci* **2019**, *23*, 305–317, doi:10.1016/j.tics.2019.01.009.
29. Taud, H.; Mas, J.F. Multilayer Perceptron (MLP). **2018**, 451–455, doi:10.1007/978-3-319-60801-3_27.

30. Ramchoun, H.; Amine, M.; Idrissi, J.; Ghanou, Y.; Ettaouil, M. Multilayer Perceptron: Architecture Optimization and Training. *International Journal of Interactive Multimedia and Artificial Intelligence* **2016**, *4*, 26, doi:10.9781/IJIMAI.2016.415.
31. Ortega, R. G., Vázquez, M. L., Sganderla Figueiredo, J. A., & Guijarro-Rodriguez, A. (2018). Sinos river basin social-environmental prospective assessment of water quality management using fuzzy cognitive maps and neutrosophic AHP-TOPSIS. Neutrosophic Sets and Systems, 23(1), 13.
32 Crespo Berti, L. A., Haro Terán, L. F., Esparza Pijal, S. B., & Benavides Morillo, R. A. (2024). Métodos AHP y Topsis para la estimación del ordenamiento jurídico positivo penal ecuatoriano vigente desde el foco de la imputación subjetiva. Neutrosophic Computing and Machine Learning, 34, 213–222.
33 Deli, I. (2018). Operators on Single Valued Trapezoidal Neutrosophic Numbers and SVTN-Group Decision Making. Neutrosophic Sets and Systems, 22(1), 11.
34 Seikh, M. R., & Dutta, S. (2022). Solution of matrix games with payoffs of single-valued trapezoidal neutrosophic numbers. Soft Computing, 26(3), 921-936.
35 Bhat, S. A. (2023). An enhanced AHP group decision-making model employing neutrosophic trapezoidal numbers. J. Oper. Strateg Anal, 1(2), 81-89.
36. Vinicius Queiroz Using Keras Tuner to Find the Best Hyperparameters for Your Neural Network Model Available online: https://medium.com/@viniciusqroz/using-keras-tuner-to-find-the-best-parameters-for-your-neural-network-model-2dc02e0a1203 (accessed on 24 July 2023).

Received: July 14, 2024. Accepted: September 19, 2024