Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 15 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i15.42653

PAPER

Optimizing Machine Learning Algorithms for Heart Disease Classification and Prediction

Abdeljalil El-Ibrahimi[1], Oumaima Terrada[1], Oussama El Gannour[1], Bouchaib Cherradi[1,2](✉), Ahmed El Abbassi[3], Omar Bouattane[1]

[1]EEIS Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco

[2]STIE Team, CRMEF Casablanca-Settat, Provincial Section of El Jadida, El Jadida, Morocco

[3]ERTTI Laboratory, FST of Errachidia, My Ismail University, Errachidia, Morocco

bouchaib.cherradi@enset-media.ac.ma

ABSTRACT

According to the World Health Organization (WHO), cardiovascular disease is one of the leading causes of death worldwide. Thus, the prevention of this kind of illness is considered as a huge human health challenge. Additionally, the diagnostic process often involves a combination of clinical examination, laboratory tests, and other diagnostic procedures, which can be complex and time-consuming. However, advances in medical technology and research have led to improved methods for diagnosing heart disease, which can help to improve patient outcomes. Furthermore, Machine Learning (ML) methods have shown promise in helping to improve the diagnosis of heart disease. Each method requires specific parameters to produce good results. In this paper, we propose a diagnosis support system based on optimized Machine Learning algorithms, which is Artificial Neural Network (ANN), Support Vector Machine (SVM), K_Nearest Neighbour (KNN), Naive Bayes (NB), and Decision Tree (DT) to analyze the major cardiovascular risk factors, such as age, gender, high blood pressure, etc. To train and validate the ML models, a medical dataset of 558 patients with atherosclerosis is used. In this work, we achieved a 96.67% as promising accuracy level for the atherosclerosis prediction with ANN.

KEYWORDS

machine learning, atherosclerosis, cardiovascular disease, optimization

1 INTRODUCTION

Atherosclerosis, which is a type of heart disease, is largely caused by the accumulation of cholesterol deposits, called plaques, on the walls of the arteries. These plaques can restrict or block the flow of blood, leading to a range of serious health problems [1]. When the plaques build up in the coronary arteries, which supply blood to the heart muscle, it can cause angina (chest pain) or a heart attack. When plaques build up in the arteries that supply blood to the brain, it can cause a stroke. When plaques build up in the peripheral arteries, it can cause Peripheral Arterial Disease (PAD), which can lead to pain, numbness, and even amputation of the

El-Ibrahimi, A., Terrada, O., Gannour, O.E., Cherradi, B., Abbassi, A.E., Bouattane, O. (2023). Optimizing Machine Learning Algorithms for Heart Disease Classification and Prediction. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(15), pp. 61–76. https://doi.org/10.3991/ijoe.v19i15.42653

Article submitted 2023-06-28. Revision uploaded 2023-08-06. Final acceptance 2023-08-23.

affected limb. Atherosclerosis is a chronic disease that develops over many years, and it is one of the leading causes of death globally [2]. The prevention of cardiovascular disease has become a major human health challenge in recent years. Early treatment of risk factors will reduce cardiovascular disease and serious long-term complications, as recommended by physicians [3].

Extracting useful information from patient records can be a time-consuming and challenging task, especially when dealing with large amounts of unstructured data, such as electronic health records. The manual process of extracting and analyzing this data can lead to errors and inconsistencies, which can affect the accuracy of the diagnosis. Artificial Intelligence (AI) techniques, particularly Machine Learning (ML) [4], have been widely used in medical diagnosis and treatment systems to automate the classification of patients and the prediction of cardiovascular diseases. These techniques have been applied to a wide range of domains, beyond healthcare [5]–[7], including: Medical image segmentation [8], [9], COVID-19 Disease [10], Natural Language Processing (NLP) [11], Image processing [12], Industrial automation, Predictive modelling and Diagnosis of Neurological disorders [13]. With the help of ML, we can process large datasets, identify patterns, and make predictions that would be difficult or impossible to do manually [14]. Early detection of heart disease can help patients make lifestyle changes that may reduce the likelihood of developing heart disease. Moreover, these changes allow for timely medical intervention, which can prevent or delay the onset of heart disease and its complications [15], [16].

In this work, we proposed an ML-based diagnostic system based on five optimized ML methods. The goal is to classify patients and predict heart diseases using Cleveland and Hungarian heart disease databases. These datasets are databases that use both controllable and uncontrollable risk factors, such as obesity, hypertension, smoking, diabetes, cholesterol, age, family history, and gender. The main contributions of this research can be summarized as follows:

- We developed and implemented an automatic medical diagnosis support system for classifying patients and predicting heart disease;
- We adjusted the parameters of the used ML algorithms to obtain optimal performance;
- The structure of the proposed system allows its use to be extended to the prediction of other diseases;
- In comparison with state of the art, the proposed system tested on two well-known datasets reach high performance.

The rest of the paper is organized as follows. In Section 2, we provide a general review of the related work, discussing previous studies and approaches in the field of ML-based diagnostic systems for heart disease. Section 3 presents the details of the proposed ML-based diagnostic system including its process and implementation, as well as the performance measures used for its evaluation. In Section 4, we afford a detailed analysis of the results obtained from the implementation of the proposed methods. Section 5 is dedicated to comparing the results obtained by this system with the results of related work. This section provides a comparison of the performance of the proposed system with other existing ML-based diagnostic systems for heart disease detection. Finally, in Section 6, we conclude this work by summarizing the main findings and discussing the potential future work.

2 RELATED WORK

In recent years, many researchers have proposed various methods for the diagnosis of atherosclerosis disease. In [17], authors used a set of fuzzy rules. The weighting of these rules used to give more importance to certain rules over others, and this method has been found to have an accuracy of 57.85%. In another work published in [18], the researchers used neural network ensemble method to build new models by combining the predicted values of several predecessor models. The obtained accuracy by this method was 89.01%. Furthermore, another proposed method in [19], called hybrid-genetic consists of increasing the performance of the neural network by improving its initial weights by genetic algorithm. The accuracy obtained by this method was 93%. In [20], the authors applied the Fast Decision Tree and c4.5 methods, and used an integration of the outcomes of the ML methods implemented on different datasets targeting CAD. The accuracy obtained by this approach was 78.06%. In [21], the authors proposed a new Intelligent Heart Disease Prediction System based on three ML Methods to predict patients affected by atherosclerosis. The highest accuracy achieved was 89%. Similarly, in [22], the authors proposed a decision support system based on the integration of the ANN and the Fuzzy-AHP methods. The accuracy achieved by this approach was 91.10%. Moreover in [23], the authors implemented five data mining algorithms in prediction heart diseases. The best performance was achieved by DT method, where the accuracy was 93.02%. Recently, the authors in [24] have proposed an MDSS for heart disease. This system is based on Fuzzy Analytic Hierarchy Process (AHP) and Fuzzy Inference System (FIS). The results show the efficiency and accuracy of this approach. In [25], the authors have proposed hybrid random forest with a linear model (HRFLM) for atherosclerosis early prediction. The accuracy reached was 88.7%.

Artificial intelligence techniques are widely used in the healthcare field, where they are used to diagnose many diseases other than heart diseases such as retinopathy [26] and Alzheimer's Disease [27]. ML techniques are effective in many fields, such as healthcare, for example, to recognize human images and their abnormal activity [28], [29].

The main objective of this work is to propose a new medical diagnosis aid system based on five optimized machine learning methods, to classify patients and predict cardiovascular diseases with an optimal accuracy rate. The methods we used are Artificial Neural Network (ANN), K-nearest neighbor (KNN), Naive Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT). The role of this system is to compute the performance of each algorithm and choose the model that gives the best performance by optimizing the main parameters.

3 THE PROPOSED METHODOLOGY AND IMPLEMENTATION

3.1 Overview of the proposed system

The proposed ML based diagnostic system is built using several steps, which are: (i) Understanding problems and objectives. (ii) Data mining and analysis. (ii) Preparation of data sets and choice of parameters. (iv) Execution of algorithms and search for optimal parameters. (v) Calculation of performance and selection of best algorithm. (vi) Presentation of results. The flowchart shown in Figure 1 provides a visual representation of the steps and how they are interconnected.

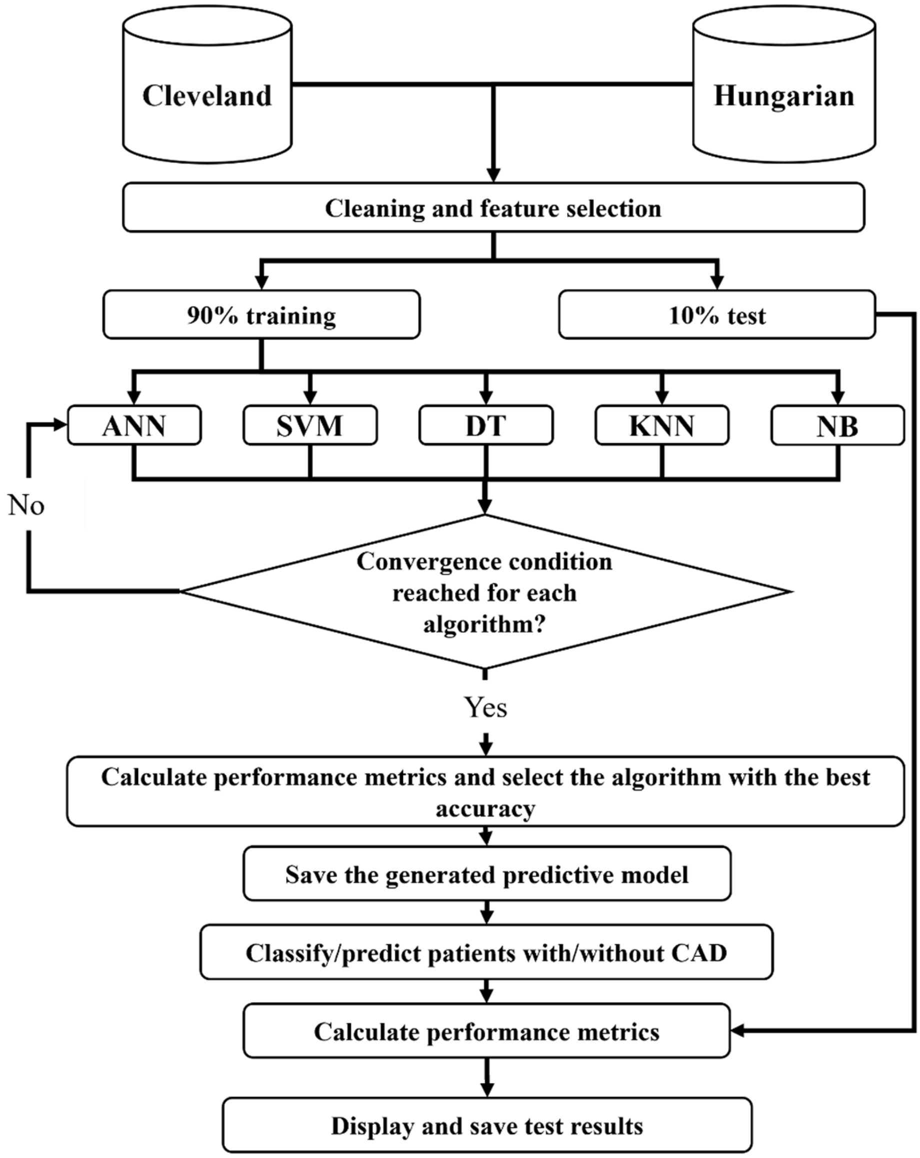

Fig. 1. Proposed system architecture

3.2 Machine learning algorithms

In this work, we have chosen to use five ML algorithms to classify patients and predict heart diseases. For our choice, we relied on the results of previous studies. The analysis of these results allowed us to see that the algorithms chosen to give good results when dealing with databases have a similar structure to the databases that we used. The chosen algorithms are ANN, KNN, NB, DT, and SVM. Then, we evaluate the performance of each algorithm using some evaluation metrics and choose the model that gives the best performance.

Artificial neural network (ANN). The artificial neural network is a biologically inspired ML method that attempts to simulate the tasks of neurons in the human brain [30]. Figure 2 presents an ANN sample. A neural network-based algorithm learns to make predictions by adjusting the weights and biases of its neurons based on the input and output data. Where x_i design the Input, w_i is the weights associated with x_i. After several training operations and using several optimization methods, we obtained improved results with the parameters shown in the Table 1.

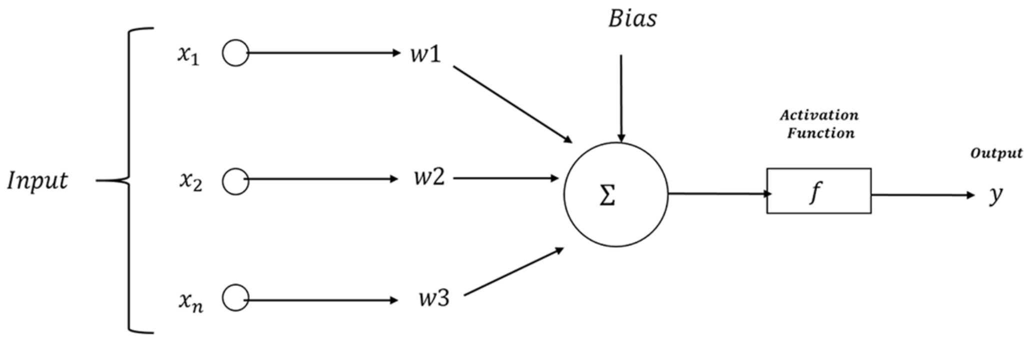

Fig. 2. Artificial neuron model

Table 1. Neural network hyperparameters

Parameter	Cleveland Dataset		Hungarian Dataset	
Learning Rate	0.0001		0.0001	
Layers number	5		7	
Neuron number	Input	64	Input	128
	Hidden 1	32	Hidden 1	68
	Hidden 2	13	Hidden 2	34
	Hidden 3	8	Hidden 3	17
	Output	1	Hidden 4	16
			Hidden 5	9
			Output	1
Activation functions	Input	Relu	Input	Relu
	Hidden 1	Relu	Hidden 1	Relu
	Hidden 2	Relu	Hidden 2	Relu
	Hidden 3	Relu	Hidden 3	Relu
	Output	Sigmoid	Hidden 4	Relu
			Hidden 5	Relu
			Output	Sigmoid
Training percentage	90%			
Testing percentage	10%			
Weights and bias	Random			
Loss function	Binary crossentropy			
Epochs	383		950	
Optimazer	Adam		RMSprop	

Support vector machine (SVM). Support Vector Machine (SVM) is a supervised ML algorithm originally introduced for the classification problems and later extended to several different situations. A linear or non-linear algorithm that finds the hyperplane maximally separates different classes in the feature space, called the optimal hyperplane. In the implementation step we use 90% for training and 10% for testing, the hyperparameters for this algorithm are shown in Table 2.

Table. 2. SVM Hyperparameters

Parameter	Cleveland and Hungarian Dataset
The C	1000
The gamma	10^{-5}
Kernel	Radial Basis Function

Decision tree (DT). The decision tree (DT) is a tree-based algorithm that recursively partitions the feature space into smaller regions, to make predictions based on the majority class in each region. This method was applied for classification and regression cases. A DT structure includes a node corresponding to a feature (attribute), a link (branch) corresponding to a decision (rule) and a leaf corresponding to a result. The algorithm works by dividing the information into two parts or subsets that support the values of the input variables. This algorithm is represented by a tree which contains nodes and leaves. This algorithm was trained with 90% of data, the features are split by the Gini impurity and the maximum depth chosen is 3.

The naïve bayes algorithm. A probabilistic algorithm that makes predictions based on the probability of each class given the input features. This is a supervised algorithm which is used for classification, using the equation:

$$\hat{p}\left(Y = i \,|\, X_1, X_2, \ldots, X_3\right) = \frac{\pi(Y = i)\prod_{j=1}^{q} P(X_i \,|\, Y = i)}{\sum_{j=1}^{K} \pi(Y = i)\prod_{j=1}^{q} P(X_j \,|\, Y = i)} \tag{1}$$

K-Nearest neighbors (KNN) algorithm. The KNN is a supervised learning algorithm that uses a distance metric to classify data points based on the proximity of the k nearest training examples [31]. The best results are obtained when we use the Euclidean distance for both datasets, and the odd k value $K = 23$ for the Cleveland dataset, while the k value for Hungarian dataset was $k = 18$.

3.3 Description of datasets

Cleveland dataset. David Aha collected this database, which contains 303 samplers and 76 attributes but only 298 samplers and 14 attributes are used due to missing data. This database contains 53.9% of healthy cases and 46.1% of coronary cases. The patients who are healthy are marked with a label of "0", while patients who are sick are marked with a different label, such as "1". This is a common approach in binary classification problems, where the goal is to predict one of two outcomes (in this case, healthy or sick). Figure 3 shows the distribution percentage of positive and negative cases for Cleveland dataset.

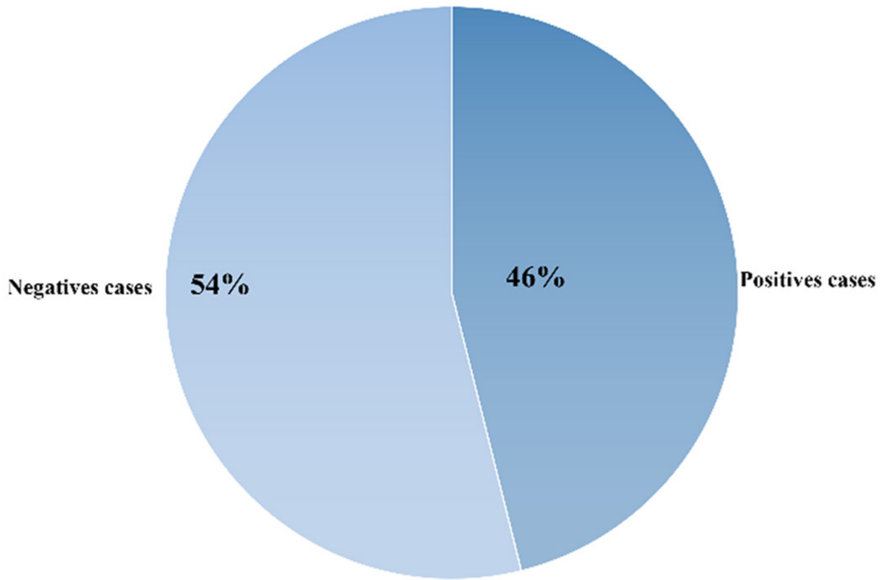

Fig. 3. Distribution of the Cleveland database

Hungarian dataset. This database is collected by Andras Janosi at the Institute of Cardiology, Budapest. It contains 294 patients and 10 attributes, only 262 members are used due to missing data. This database consists of 62.5% sick patients and 37.5% healthy patients. Figure 4 shows the distribution percentage of positive and negative cases for Hungarian dataset.

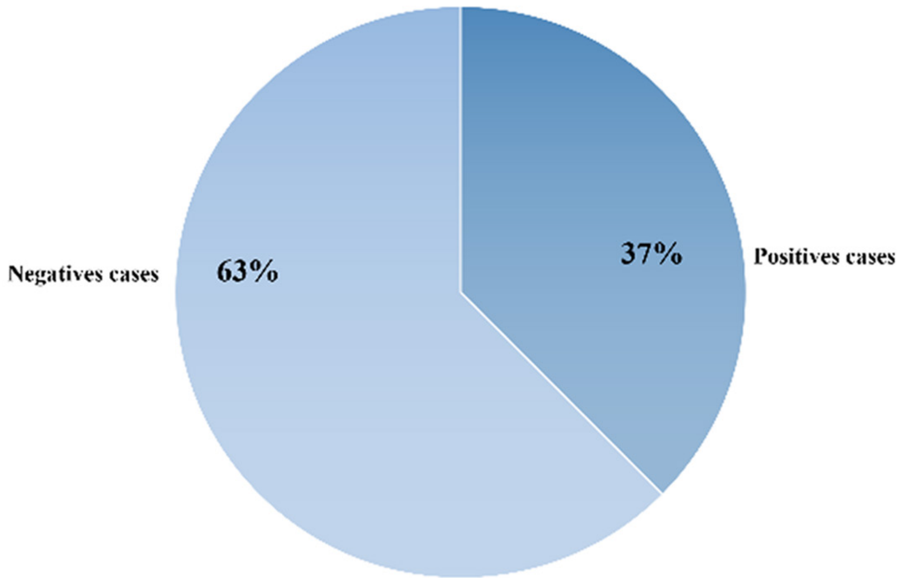

Fig. 4. Distribution of the Hungarian database

3.4 Selection of characteristics

After cleaning the databases by ignoring entries with missing data, the inputs data for each database are risk factors that have been identified by medical experts as having a clear influence on heart disease. In this work, we relied on the literature mentioned above that used the Cleveland or Hungary datasets for the selection of these features, and we use the correlation coefficient method to verify the importance of these features. Figures 5 and 6 show the correlation results with the target in the Cleveland and Hungary datasets.

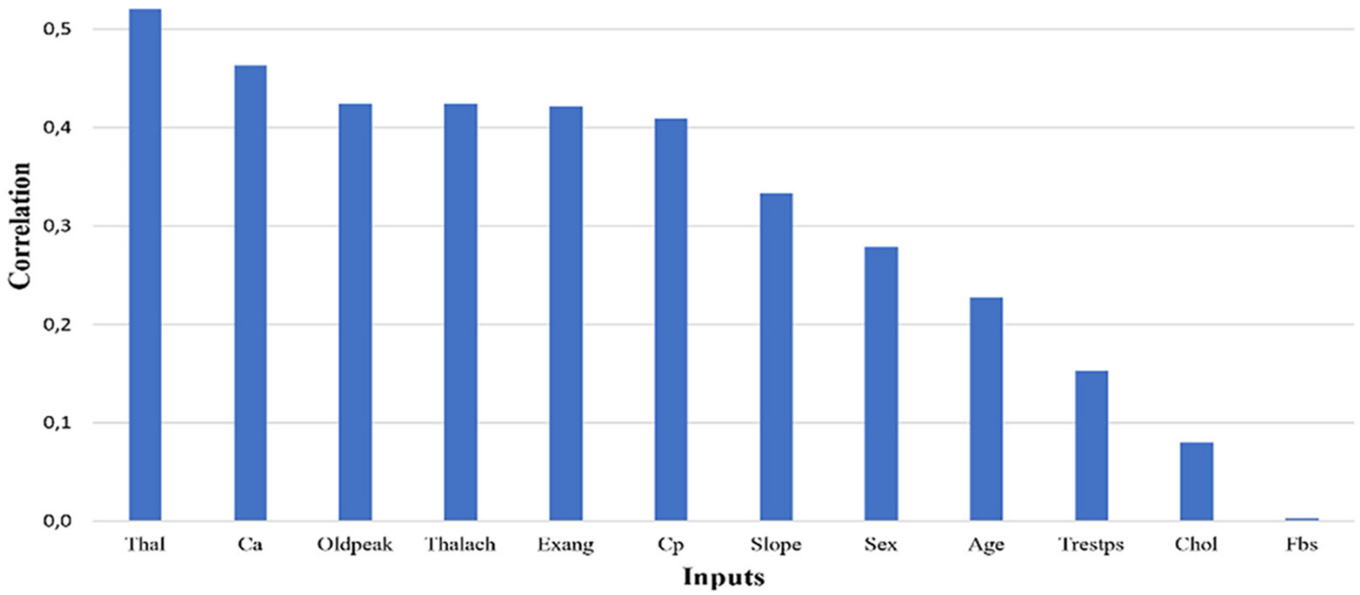

Fig. 5. Correlation with target in Cleveland dataset

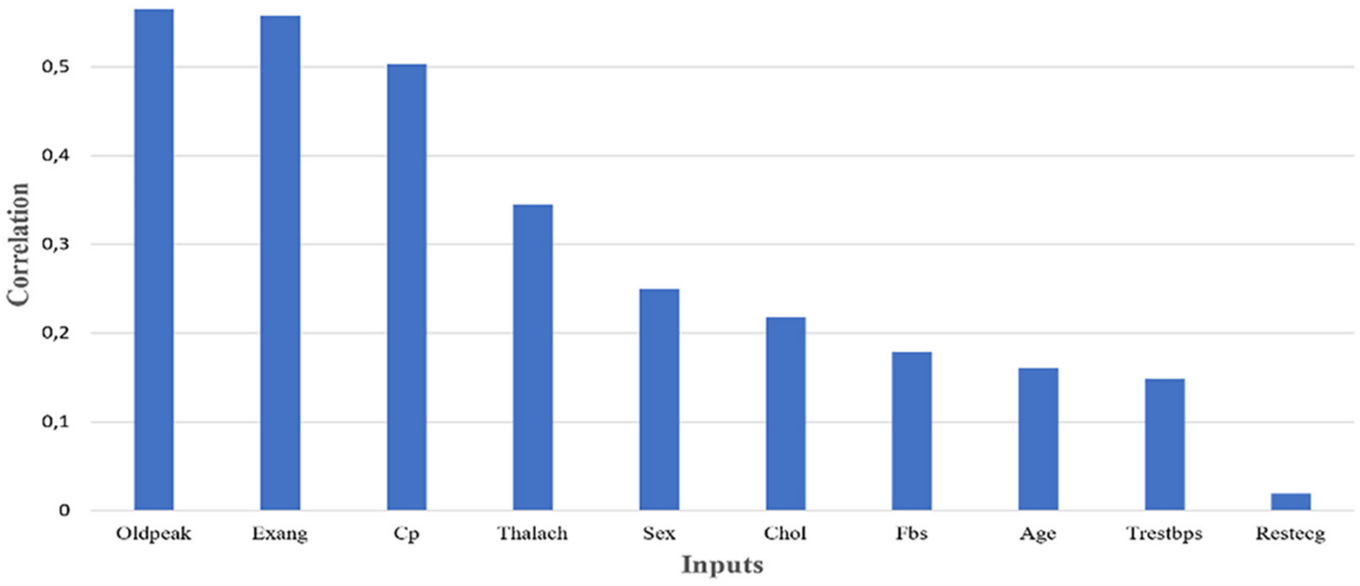

Fig. 6. Correlation with target in Hungarian dataset

The corresponding output used in these two databases for classification and prediction are binary labels indicating the real status of the patient. There are two classes: a patient with heart disease, which signifies a diameter narrowing of more than 50%, or a patient without heart disease, which signifies a diameter narrowing of less than 50%, based on the UCI Cleveland and Hungary databases. The goal is to classify patients as having or not having heart disease based on the input risk factors and the binary labels. Table 3 shows the characteristics selected for each database.

Table 3. The selected features

Database	No. of Features	Selected Features
Cleveland	13	Age, Gender (GD), cholesterol (chol), Chest pain type (CP), Resting blood pressure (Trestbps), Diabetes (Fbs), Electro cardiographic results (Restecg), Heart rate (Thalach), ST depression (Oldpeak), Angina (Exang), Heart status (Thal), ST segmen (slope), Number of major vessels (ca).
Hungarian	10	Age, Gender (GD), cholesterol (chol), Chest pain type (CP), Resting blood pressure (Trestbps), Diabetes (Fbs), Electro cardiographic results (Restecg), Heart rate (Thalach), ST depression (Oldpeak), Angina (Exang).

3.5 Performance evaluation metrics

Confusion matrix: A binary confusion matrix is an array used to assess the performance of a binary classifier model on a set of test data for which the true values are known. The matrix is of the dimension 2x2 that contains the following four elements: True positive (TP), False positive (FP), True negative (TN), False negative (FN).

The binary confusion matrix is widely used to evaluate the performance of binary classification models and it is used to calculate several performance metrics such as Specificity (SP), Sensibility (SS), Accuracy (ACC), F1-score (FS) and Matthews's correlation coefficient (MCC).

$$\mathrm{ACC} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$SS = \frac{TP}{TP+FN} \tag{3}$$

$$SP = \frac{TN}{TN+FP} \tag{4}$$

$$FS = \frac{2*TP}{2*TP+FP+FN} \tag{5}$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+GN)}} \tag{6}$$

4 RESULTS OF PERFORMANCE MEASURES

To determine which hyperparameter of each algorithm performs best, we use commonly used ML technologies such as Manual Search, Grid Search and Randomized Search. Then, we trained and evaluated the models with different combinations of hyperparameter values. The goal is to find the combination of hyperparameter values that result in the best performance of the model, as measured by the ACC, SS, SP, MCC, and F1-score.

Splitting data is a common technique used in ML to divide a dataset into two or more subsets, such as a training set and a test set. The goal of splitting the data is to use the training set to train the model and the test set to evaluate the model's performance. In the case we mentioned, the Cleveland and Hungarian datasets are each split into two parts, 90% for training and 10% for testing. The training set is used to train the model and optimize the hyperparameters of each algorithm, while the test set is used to evaluate the model's performance. After the training process, we evaluate each method using the performance metrics mentioned above by the test data. Table 4 shows the confusion matrix results. the remainder of the performance metrics are provided in Table 5. As we can see from the results presented in the Tables 4 and 5, the ANN model gives the best results compared to the other MLs techniques. Most importantly, we have achieved excellent results indicating the quality of the proposed system.

Table 4. The confusion matrix results in the test stage

Dataset	Method	TP	FP	FN	TN
Cleveland	ANN	18	0	2	10
	KNN	16	5	2	7
	DT	13	0	7	10
	SVM	10	0	4	16
	NB	7	2	2	19
Hungarian	ANN	16	0	2	9
	KNN	19	0	5	3
	DT	14	2	2	9
	SVM	14	2	3	8
	NB	13	3	1	10

Table 5. The performance metrics achieved in the test stage

Cleveland Dataset					
Method	SS (%)	SP (%)	ACC (%)	FS	MCC
ANN	83.33	100	**96.67**	0.91	0.87
KNN	77.78	76.19	76.67	0.67	0.50
DT	77.78	85.71	83.33	0.74	0.82
SVM	77.78	85.71	83.33	0.74	0.62
NB	77.78	90.48	86.67	0.78	0.68
Hungarian Dataset					
ANN	81.82	100	**92.59**	0.90	0.85
KNN	36.36	93.75	70.37	0.50	0.54
DT	81.82	87.50	85.19	0.82	0.85
SVM	72.73	87.50	81.48	0.76	0.61
NB	90.91	81.25	85.19	0.83	0.54

Furthermore, the results of the ROC curve metric are consistent with the above results, this metric uses threshold values for every classifier over the interval [0, 1] to the output field. Two major values are then calculated for each threshold, the TPR and the FPR. The result is shown in the Figures 7 and 8 for this parameter.

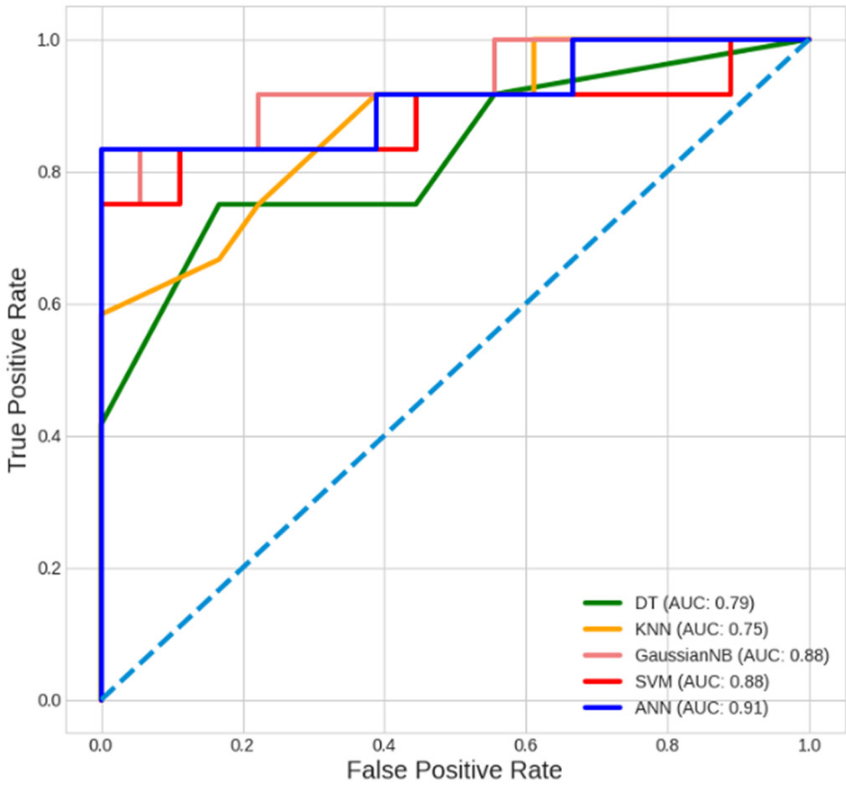

Fig. 7. ROC analysis of Cleveland dataset

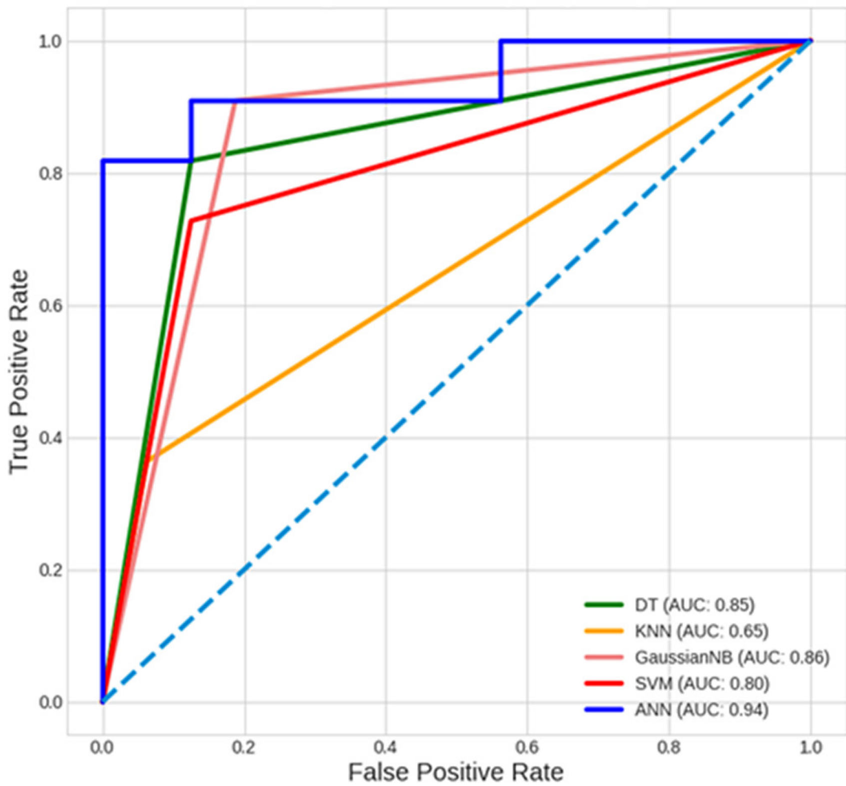

Fig. 8. ROC for Hungarian dataset

Additionally, the area under the ROC curve (AUC) is a commonly used scalar performance metric for binary classification models. A model with an AUC of 1.0 has a perfect prediction, while a model with an AUC of 0.5 has no discrimination ability. An analytical reading of Figures 7 and 8 shows clearly that the ANN model has a better sensitivity and specificity than the other models.

Presumably, the performance metrics (Tables 4 and 5) and ROC curve (Figures 7 and 8) can provide valuable insights into the performance of the different models used in the proposed system. Particularly, the ANN model appears to have the better SS, SP and ACC than the other models. A high sensitivity indicates that the model can correctly identify a large proportion of actual positive cases, while a high specificity indicates that the model is able to correctly identify a large proportion of actual negative cases. Therefore, a model with a high sensitivity and specificity is considered a good classifier.

5 COMPARISON WITH RELATED WORK

In literature several methods are implemented to classify patients and predict heart disease. The aim of each method is always to achieve the best results to the classify and predict patients with heart disease. Researchers have often compared the performance of different methods on the same dataset to determine which one gives the best results. The choice of method will also depend on the specific features of the dataset and the research question being addressed. Additionally, the interpretability of the model and the ability to use it for clinical decision-making are also important aspects to consider. Therefore, we compared our proposed system with some existing methods that deal with the same database, the results of the comparison are shown in the Table 6.

Table 6. The accuracy obtained in some approach for Cleveland and Hungarian datasets

Reference	Method	Dataset	Accuracy
[17]	Weighted fuzzy rules	Cleveland	62.35%
		Hungarian	46.93%
[18]	Neural Networks Ensemble	Cleveland	89.01%
[19]	Hybrid Neural Network	Cleveland	89.40%
		Hungarian	87.10%
[20]	C4.5	Cleveland	78.54%
		Hungarian	78.57%
	FDT	Hungarian	78.23%
This work	ANN	Cleveland	96.67%
		Hungarian	92.59%

As mentioned in the table, several methods have been used by researchers to classify and predict heart disease. This comparison shows that our proposed system gave the best performance results compared to other existing methods.

6 CONCLUSION

In this work, we proposed a new ML based diagnostic support system based on five algorithms known for their efficiency in data classification, which are ANN, KNN, NB, SVM and DT algorithms where we obtained an accuracy of 96.67% in the classification of heart diseases utilizing these famous datasets, Cleveland and Hungarian. The aim of this work was to find the hyperparameters of each algorithm that give the best results in terms of performance. The experimental results obtained in the test phase show that all the algorithms give better classification and prediction performance, especially the ANN algorithm which gives good results after being optimized. All this shows that these methods can produce better results if we continue to optimize them. Consequently, the challenge for future research will be to develop methods capable of optimizing these algorithms according to datasets to achieve the maximum possible performance for each algorithm. These results also show that the ANN method has a better ability to classify and predict diseases, which can be explained by the weighting of each input according to its impact on the patient's condition. All this confirms the ability of the ML methods mentioned above to predict heart disease. Finally, we compared the results obtained using this model with some previous methods, which shows that this system has added value in the classification and the prediction of patients with cardiovascular diseases compared with the methods available in the literature. In view of the promising results obtained, we recommend using this model to predict other diseases.

7 REFERENCES

[1] N. Aljefree, I. Shatwan, and N. Almoraie, "Association between nutrients intake and coronary heart disease among adults in Saudi Arabia: A case-control study," *Prog. Nutr.*, vol. 23, no. 3, p. e2021111, 2021. https://doi.org/10.23751/pn.v23i3.10694

[2] "The top 10 causes of death," https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [Accessed Jun. 11, 2023].

[3] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "Atherosclerosis disease prediction using supervised machine learning techniques," in *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco: IEEE, 2020, pp. 1–5. https://doi.org/10.1109/IRASET48871.2020.9092082

[4] S. Taha Ahmed and S. Malallah Kadhem, "Using machine learning via deep learning algorithms to diagnose the lung disease based on chest imaging: A Survey," *Int. J. Interact. Mob. Technol. IJIM*, vol. 15, no. 16, pp. 95–112, 2021. https://doi.org/10.3991/ijim.v15i16.24191

[5] J. Ali, A. Jusoh, N. Idris, N. A. Qureshi, S. A. Shah, and A. F. Abbas, "Combining mobile technologies, healthcare & service quality: A bibliometric analysis," *Int. J. Interact. Mob. Technol. IJIM*, vol. 16, no. 13, pp. 140–159, 2022. https://doi.org/10.3991/ijim.v16i13.30613

[6] I. Mustapha, N. Khan, M. I. Qureshi, A. A. Harasis, and N. T. Van, "Impact of industry 4.0 on healthcare: A Systematic Literature Review (SLR) from the last decade," *Int. J. Interact. Mob. Technol. IJIM*, vol. 15, no. 18, pp. 116–128, 2021. https://doi.org/10.3991/ijim.v15i18.25531

[7] O. Rahmaoui, K. Souali, and M. Ouzzif, "Improving software development process using data traceability management," *Int. J. Recent Contrib. Eng. Sci. IT IJES*, vol. 7, no. 1, pp. 52–58, 2019. https://doi.org/10.3991/ijes.v7i1.10113

[8] H. Moujahid, B. Cherradi, and L. Bahatti, "Convolutional neural networks for multimodal brain MRI images segmentation: A comparative study," in *Smart Applications and Data Analysis*, M. Hamlich, L. Bellatreche, A. Mondal, and C. Ordonez, Eds., Communications in Computer and Information Science, vol. 1207, Cham: Springer International Publishing, 2020, pp. 329–338. https://doi.org/10.1007/978-3-030-45183-7_25

[9] N. Ait Ali, B. Cherradi, A. El Abbassi, O. Bouattane, and M. Youssfi, "GPU fuzzy c-means algorithm implementations: Performance analysis on medical image segmentation," *Multimed. Tools Appl.*, vol. 77, no. 16, pp. 21221–21243, 2018. https://doi.org/10.1007/s11042-017-5589-6

[10] H. Moujahid, B. Cherradi, M. Al-Sarem, and L. Bahatti, "Diagnosis of COVID-19 disease using convolutional neural network models based transfer learning," in *Innovative Systems for Intelligent Health Informatics*, F. Saeed, F. Mohammed, and A. Al-Nahari, Eds., Lecture Notes on Data Engineering and Communications Technologies, vol. 72, Cham: Springer International Publishing, 2021, pp. 148–159. https://doi.org/10.1007/978-3-030-70713-2_16

[11] K. Jain and V. Prajapati, "NLP/Deep learning techniques in healthcare for decision making," *Prim. Health Care*, vol. 11, no. 3, p. 373, 2021.

[12] M. A. R. Ahad, S. Kobashi, and J. M. R. S. Tavares, "Advancements of image processing and vision in healthcare," *J. Healthc. Eng.*, vol. 2018, pp. 1–3, 2018. https://doi.org/10.1155/2018/8458024

[13] H. Kaur, A. K. Malhi, and H. S. Pannu, "Machine learning ensemble for neurological disorders," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12697–12714, 2020. https://doi.org/10.1007/s00521-020-04720-1

[14] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "A fuzzy medical diagnostic support system for cardiovascular diseases diagnosis using risk factors," in *2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, Kenitra: IEEE, 2018, pp. 1–6. https://doi.org/10.1109/ICECOCS.2018.8610649

[15] D. S. Celermajer, C. K. Chow, E. Marijon, N. M. Anstey, and K. S. Woo, "Cardiovascular disease in the developing world," *J. Am. Coll. Cardiol.*, vol. 60, no. 14, pp. 1207–1216, 2012. https://doi.org/10.1016/j.jacc.2012.03.074

[16] D. Farmakis, J. Parissis, J. Lekakis, and G. Filippatos, "Acute heart failure: Epidemiology, risk factors, and prevention," *Rev. Esp. Cardiol. Engl. Ed.*, vol. 68, no. 3, pp. 245–248, 2015. https://doi.org/10.1016/j.rec.2014.11.004

[17] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 24, no. 1, pp. 27–40, 2012. https://doi.org/10.1016/j.jksuci.2011.09.002

[18] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7675–7680, 2009. https://doi.org/10.1016/j.eswa.2008.09.013

[19] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm," *Comput. Methods Programs Biomed.*, vol. 141, pp. 19–26, 2017. https://doi.org/10.1016/j.cmpb.2017.01.004

[20] R. El-Bialy, M. A. Salamay, O. H. Karam, and M. E. Khalifa, "Feature analysis of coronary artery heart disease data sets," *Procedia Comput. Sci.*, vol. 65, pp. 459–468, 2015. https://doi.org/10.1016/j.procs.2015.09.132

[21] J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive data mining for medical diagnosis: An overview of heart disease prediction," *Int. J. Comput. Appl.*, vol. 17, no. 8, pp. 43–48, 2011. https://doi.org/10.5120/2237-2860

[22] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction," *Expert Syst. Appl.*, vol. 68, pp. 163–172, 2017. https://doi.org/10.1016/j.eswa.2016.10.020

[23] M. Abdar, S. R. Niakan Kalhori, T. Sutikno, I. Much Ibnu Subroto, and G. Arji, "Comparing performance of data mining algorithms in prediction heart diseases," *Int. J. Electr. Comput. Eng. IJECE*, vol. 5, no. 6, p. 1569, 2015. https://doi.org/10.11591/ijece.v5i6.pp1569-1576

[24] S. Nazari, M. Fallah, H. Kazemipoor, and A. Salehipour, "A fuzzy inference- fuzzy analytic hierarchy process-based clinical decision support system for diagnosis of heart diseases," *Expert Syst. Appl.*, vol. 95, pp. 261–271, 2018. https://doi.org/10.1016/j.eswa.2017.11.001

[25] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019. https://doi.org/10.1109/ACCESS.2019.2923707

[26] O. M. Al-hazaimeh, A. Abu-Ein, N. Tahat, M. Al-Smadi, and M. Al-Nawashi, "Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images," *Int. J. Online Biomed. Eng. IJOE*, vol. 18, no. 13, pp. 131–151, 2022. https://doi.org/10.3991/ijoe.v18i13.33985

[27] N. Gharaibeh, A. A. Abu-Ein, O. M. Al-hazaimeh, K. M. O. Nahar, W. A. Abu-Ain, and M. M. Al-Nawashi, "Swin Transformer-based segmentation and multi-scale feature pyramid fusion module for Alzheimer's disease with machine learning," *Int. J. Online Biomed. Eng. IJOE*, vol. 19, no. 4, pp. 22–50, 2023. https://doi.org/10.3991/ijoe.v19i04.37677

[28] M. Al-Nawashi, O. M. Al-Hazaimeh, and M. Saraee, "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments," *Neural Comput. Appl.*, vol. 28, no. S1, pp. 565–572, 2017. https://doi.org/10.1007/s00521-016-2363-z

[29] O. M. Al-Hazaimeh, M. Al-Nawashi, and M. Saraee, "Geometrical-based approach for robust human image detection," *Multimed. Tools Appl.*, vol. 78, no. 6, pp. 7029–7053, 2019. https://doi.org/10.1007/s11042-018-6401-y

[30] C. Bai and W. Huang, "Application of an artificial neural network to evaluate the integration effect of quality education and skill education," *Int. J. Emerg. Technol. Learn. IJET*, vol. 16, no. 7, pp. 218–232, 2021. https://doi.org/10.3991/ijet.v16i07.22121

[31] D. Cheng, S. Zhang, Z. Deng, Y. Zhu, and M. Zong, "KNN algorithm with data-driven k value," in *Advanced Data Mining and Applications*, X. Luo, J. X. Yu, and Z. Li, Eds., in Lecture Notes in Computer Science, vol. 8933. Cham: Springer International Publishing, 2014, pp. 499–512. https://doi.org/10.1007/978-3-319-14717-8_39

8 AUTHORS

Abdeljalil El-Ibrahimi obtained his master's degree in computer science and engineering of Complex Systems from the Faculty of Sciences and Technology of Errachidia (FSTE), Moulay Ismail University, Morocco in 2022. Currently, he is a Ph.D. student since 2022 in ENSET of Mohammedia, Hassan II University of Casablanca. His research interests include the development of a medical diagnosis support system based on machine learning, deep learning, fuzzy sets, and neutrosophic sets (E-mail: elibrahimiabdeljalil@gmail.com).

Oumaima Terrada got her Ph.D. degree in 2020 in computer science and artificial intelligence from the Faculty of Sciences and Technics, Mohammedia at Hassan II University of Casablanca. She received her bachelor's degree in mathematical sciences, Computer Science, and Applications (SMIA) from the Poly Disciplinary Faculty of Safi in 2012 at Cadi AYYAD University, Morocco. She holds a master's degree in computer science, Computer Graphics, and Imaging (M3I) from the Faculty

of Sciences Dhar Mahraz in 2014 at Sidi Mohamed Ben Abdellah University Fez, Morocco (E-mail: oumaima.terrada@gmail.com).

Oussama El Gannour obtained a Master's specialized degree in Internet of Things and Mobile Services from the National Higher School of Computer Science and Systems Analysis of Rabat, Mohammed V University. In 2023, he obtained his Ph.D. from EEIS laboratory, ENSET of Mohammedia, Hassan II University of Casablanca. His research interests include contributing to the integration of the Internet of Things and artificial intelligence in the medical field (E-mail: oussama.elgannour@gmail.com).

Bouchaib Cherradi received the B.S. degree in electronics in 1990 and the M.S. degree in Applied Electronics in 1994 from the ENSET Institute of Mohammedia, Morocco. He received the DESA diploma in Instrumentation of Measure and Control (IMC) from Chouaib Doukkali University at El Jadida in 2004. He received his Ph.D. in electronics and image processing from the Faculty of Science and Technology, Mohammedia. Dr. Cherradi works as an associate professor in CRMEF-Casablanca-Settat. In addition, he is associate researcher member of Electrical Engineering and Intelligent Systems (EEIS) Laboratory in ENSET Mohammedia, Hassan II University of Casablanca (E-mail: bouchaib.cherradi@enset-media.ac.ma).

Ahmed El Abbassi received the B.S. degree in applied sciences from the College of Ibn Khaldoun in El Jadida city, Morocco in 1988, the M.S. degree in high frequency and telecommunications from Faculty of Sciences of El Jadida, in 1997, and the Ph.D. on Transistors MOS: Modelling and characterization from Faculty of Sciences of El Jadida, Chouaib Doukkali University in 2002. His research interests are microelectronics and transmission systems. He is with EPIM laboratory, on leave from the Faculty of Science and Technology (FST), Moulay Ismail University (MIU), Errachidia, Morocco (E-mail: a.elabbassi@fste.umi.ac.ma).

Omar Bouattane obtained his Ph.D. degree in 2001 in Parallel Image Processing on Reconfigurable Mesh Computer from the Faculty of Science Ain Chock, Casablanca. He is the director of Electrical Engineering and Intelligent Systems (EEIS) Laboratory in ENSET of Mohammedia. His research interests include Massively Parallel Architectures, cluster analysis, pattern recognition, image processing, artificial intelligence, and electrical engineering (E-mail: o.bouattane@enset-media.ac.ma).