Nutr. Santé, 2022, Vol. 11, N°01:34-47. DOI:10.30952/ns.11.1.4

Santé publique

Un modèle de réseaux de neurones artificiels pour prédire la mortalité des patients due au COVID-19"CoviSurv2021"

An Artificial Neural Networks model for predicting patients' mortality due to COVID-19 "CoviSurv2021"

Fatima Z. HADDAD[1], Saliha OUADAH[1], Lina LEFILEF[2], Mohammed A. BENBOURAS[3]

[1]Ecole Nationale Supérieure de Statistique et d'Economie Appliquée Pôle Universitaire de Koléa, Tipaza 42003, Algeria. [2]Université Mohammed SeddikBenyahia, BP 98 OuledAïssa, Jijel 18000, Algeria. [3]École Normale Supérieure d'Enseignement Technologique de Skikda (ENSET), Frères Boucetta city, Azzaba 21000, Skikda, Algeria

Auteur correspondant : mouhamed_amine.benbouras@g.enp.edu.dz

Reçu le 21 février2022, Révisé le 03 mai 2022, Accepté le 24 mai 2022

Abstract ***Introduction.*** COVID-19 is the pandemic of the century with the unusual circumstances it generated. Subsequently, there has been medical and human scarcity of resources leading to the health system collapse, especially in third world countries. ***Objective.*** To support the white army in grasping the pandemic behavior, several studies have pointed to the existence of patient-related factors affecting COVID-19 patients' mortality-risk. In the current study, Artificial Neural Network (ANN) has been employed to predict COVID-19 mortality. ***Material and methods.*** In particular, the modeling phase was done using a database of 684 samples collected from Mohamed Seddik Ben Yahia Hospital of Jijel, with antecedent diseases and blood biomarkers data of patients. Firstly, 18 parameters were selected in the input layer based on the literature recommendation and expert medical team consultation. Furthermore, the optimal inputs have been modeled using the ANN, and their performance was assessed through four performance measures (sensitivity, specificity, precision, and accuracy). ***Results***. The comparative study proved the effectiveness of (18-12-2) model trained by Tansig transfer function, which displayed a higher performance in predicting COVID-19 mortality, compared to other models proposed in the literature. Afterward, the proposed optimal model was utilized to develop a GUI public interface by Matlab software. ***Conclusion.*** Finally, a reliable and easy-to-use graphical interface is generated in the current study dubbed "CoviSurv2021". This latter will be very helpful for the medical staff to select priority patients who have upper urgency to be hospitalized, prioritize patients when the hospital is overcrowded, and gain time to provide the care needed.

Key words: COVID-19, SARS-Cov-2, Corona Virus, Artificial Neural Networks, CoviSurv2021

Résumé ***Introduction***. La COVID-19 est la pandémie du siècle avec les circonstances inhabituelles qu'elle a engendré. Par la suite, il y a eu une pénurie de ressources médicales et humaines qui a conduit à l'effondrement du système de santé, en particulier dans les pays du tiers monde. ***Objectif***. Pour soutenir le personnel de santé à appréhender les risques de la pandémie, plusieurs études ont mis en évidence l'existence de facteurs liés aux patients affectant le risque de mortalité des patients COVID-19. Dans la présente étude, les réseaux de neurones artificiels ont été utilisés pour prédire la mortalité liée au COVID-19. ***Matériel et méthodes***. En particulier, la phase de modélisation a été réalisée à partir d'une base de données de 684 échantillons prélevés à l'hôpital Mohamed Seddik Ben Yahia de Jijel, avec les données des antécédents médicaux et des biomarqueurs sanguins des patients. Premièrement, 18 paramètres ont été sélectionnés dans la couche d'entrée sur la base des recommandations de la littérature et de la consultation d'une équipe médicale d'experts. De plus, les entrées optimales ont été modélisées à l'aide des réseaux de neurones artificiels et leur performance a été évaluée à l'aide de quatre mesures de performance (sensibilité, spécificité, précision et exactitude). ***Résultats.*** L'étude comparative a prouvé l'efficacité du modèle (18-12-2) formé par la fonction de transfert de Tansig, qui a affiché une performance plus élevée dans la prédiction de la mortalité par COVID-19 par rapport aux autres modèles proposés dans la littérature. Par la suite, le modèle optimal proposé a été utilisé pour développer une interface publique GUI par le logiciel Matlab. ***Conclusion.*** Enfin, une interface graphique fiable et simple d'utilisation est générée dans la présente étude baptisée « CoviSurv2021 ». Cette dernière sera très utile au personnel médical pour sélectionner les patients prioritaires qui ont une urgence supérieure à être hospitalisés, prioriser les patients lorsque l'hôpital est surchargé et gagner du temps pour fournir les soins nécessaires.

Mots clés : COVID-19, SRAS-CoV-2, Corona virus, Réseaux de neurones artificiels, CoviSurv2021

Introduction

Human history has witnessed the spread of deathly diseases that took the lives of millions. To begin with, an infectious respiratory disease spread in China 18 years ago, it was known as SARS, this was the first shape of Coronavirus [1,2]. Later on, Saudi Arabia in 2012 witnessed a mortal contagious known as Middle East Respiratory Syndrome [3,4]. This latter spread again severely in the late of 2019 with a new sophisticated form, this new Coronavirus formed was named SARS-Cov-2 or COVID-19 (Severe Acute Respiratory Syndrome Coronavirus 2).

Coronavirus is a zoonotic virus *i.e.* It transmits from animals to humans. The reasonable ground standing behind Coronavirus name is the crown-like spikes on their surface [5]. Furthermore, four main subfamilies of corona-viruses have been detected, known as Alpha, Beta, Gamma, and Delta [6].

The current pandemic caused by the newly emerged Coronavirus SARS-CoV-2 has created a uniquely dangerous situation around the world [7], which forced national governments to impose drastic measures to contain it, especially since the international economy was disastrously affected. SARS-CoV-2 caused a disease called COVID-19, which is marked by some unique symptoms such as coughing, fever, chills, and a range of respiratory symptoms [8]. By the end of June 2021, the total number of the confirmed COVID-19 cases had exceeded 187 million worldwide, while the total number of deaths approached 4 millions. Unfortunately, recent studies have proven that this pandemic will live with us for many years to come, making the development of new methods to mitigate the dire consequences of COVID 19 by saving lives, an inescapable solution [9]. In the pandemic case, the most important thing is how to manage patients diagnosed with COVID-19 by selecting those at a higher risk of mortality at the very beginning of the symptoms, for the sake of providing them with the appropriate treatment, taking into consideration that the condition of high-risk patients can deteriorate rapidly. Previously published studies reported that

patients who died from COVID-19 initially had mild symptoms, but suddenly they turned into a critical stage causing their death [10]. For instance, in Italy, 75% of the deceased patients showed mild symptoms, such as fever, dyspnea, and cough, upon their admission to the hospital [10]. Thus, the development of an Artificial Intelligence (AI) interface to predict patients mortality is very essential.

On the other hand, machine learning and AI are gaining popularity in the field of computer sciences. They have, undoubtedly, enhanced human life in many areas, providing radical solutions to many problems that were difficult to solve by classical methods. The Application of AIhas been robustly adopted in various sectors like science, engineering, business, weather forecasting, and medicine due to the huge impact it had on them. These domains helped caregivers with medical decision-making through its multiple functions [11]. The well-known examples published in the literature are the prediction of the infection severity with the COVID-19 and death possibility. The latter is what has been covered in this research paper with ANN modeling assistant.

Since the emergence of the COVID-19 pandemic, several researchers have proposed machine learning (ML) models to predict the mortality risk of Coronavirus patients [12]. The significant benefit of proposing machine learning models is to aid the white army in selecting patients-priorities *i.e.* who urgently needs attention first or to be hospitalized. Prioritizing patients when the hospital is overcrowded, and gaining time to provide the care needed. Machine learning methods have proven to be effective in many applications [13]. The review of machine learning published studies used to predict COVID-19 mortality is provided in **Table 1**. Blood biomarkers used in the majority of them, have turned their proposed models to be very helpful. This is due to the fact that biomarkers could help in categorizing COVID-19 patients that have a high risk of dying, by giving vital information concerning their health status as provided in previous studies [14,15]. However, the aforementioned suggested models are limited to the use of fewer input parameters in modeling the mortality risk in patients with COVID-19. The important shortcoming of these models is that they disregard the antecedent diseases parameters. This has probably been conducted to oversimplify the complicated mechanism of the understudy phenol-mena when several factors have been deemed to enhance the model predictability. Although, Zhang *et al.*, (2021) have carried out a clinical study on 82 COVID-19 patients, which proved that respiratory, cardiac, hemorrhage, hepatic, and renal antecedent diseases had caused the death of 100%, 89%, 80.5%, 78.0%, and 31.7% patients, respectively. Most of the patients had augmented CRP (100%) and D-dimer (97.1%) [16]. Accordingly, the use of these parameters in relation to blood biomarkers could enhance the predictive capability of the provided model. Additionally, the majority of the available studies have suggested mathematical models in the form of equations, which make the proposed model hard to use in the future. Certainly, this idea has useless significance for other researchers and medical staff. To overcome this limitation, we have presented our model as a simple, reliable, and easy-to-use interface (dubbed “CoviSurv2021”), which can easily predict the COVID-19 mortality in the future cases by providing the input parameters.In such a case, the current proposed model can be readily used, and thus available to anyone interested in the problem of modelling, whatever his level.

Material and Methods

Dataset

In this paper, dataset contained infected and dead COVID-19 samples, collected from Mohamed Seddik Ben Yahia Hospital, Jijel, starting from August until December 2020. We had browsed patients files in the aforementioned hospital, which contained more than 1600 confirmed COVID-19 patients including 684 fully sufficient samples containing patients from both genders (male and female). It was worth mentioning that the disease was confirmed by a PCR test. We also removed the incomplete data samples and massing values to conduct a reliable study. In order to make an accurate modeling step, a considerable effort has been undertaken to make the database balanced, with a close number of samples for both recovered and deceased patients in training and validation data. The data samples in both the training and validation phase has been randomly chosen and completely detached. The **Fig. 1.** presents the metho-dology structure adopted in the present study. For a reliable study, we removed useless information, concentrating on both antecedent diseases and blood biomarkers parameters.

A medical team of experts was consulted to make sure that all of the relevant inputs were used.

Table 2 and **Table 3** present an overview of the variables adopted in this study.

Table 1. Review of machine learning studies used to predict COVID-19 mortality

Authors	Sample size	Methods	Inputs	Accuracy
Bhandari *et al*., 2020 [17]	70	Logistic Regression	Age, gender, symptomes, and blood biomarkers	70.0%
Pourhomayoun and Shakibi, 2020 [18]	117000	Neural Network, Random Forest, SVM, Decision Tree, and Logistic Regression	Age, chronic disease, fever, cough, dyspnea, gasp, shortness of breath, expectoration, hypertension, chest pain, fatigue, anorexia, dizziness, emesis, obnubilation, somnolence, diarrhea, mylagia, and sputum	93.75%
Ko *et al*., 2020 [19]	361	Random Forest, Deep Neural Network, and EDRnet	Blood biomarkers, age, and sex	92.2%
Aktar *et al*., 2021 [20]	545	Decision tree, random forest, variants of gradient boosting machine, support vector machine, K-nearest neighbor, and deep learning	Blood biomarkers	92.9%
Hu *et al*., 2020 [21]	183	Logistic regression, partial least squares regression, elastic net, random forest, and bagged flexible discriminant analysis	Epidemiological, clinical, and first laboratory findings	93.03%
Assaf *et al*., 2020 [22]	162	Artificial neural network, Random forest	Blood cell count, time from symptoms to admission, oxygen saturation, and blood lymphocytes count	83.0%
Chowdhury *et al*., 2021 [23]	375	Multi-tree XGBoost	Lactate dehydrogenase, neutrophils, lymphocyte, high-sensitivity C-reactive protein, age, and acquired at hospital admission	95.8%
Tezza *et al*., 2021 [10]	341	Recursive Partition Tree, Support Vector Machine, the Gradient Boosting Machine, and Random Forest	Age, oxygen saturation and lab parameters (creatinine, AST, lymphocytes, platelets, and hemoglobin)	93.0%
Li *et al*., 2021 [24]	2924	Gradient boosting decision tree, and logistic regression	Demographic, clinical, laboratory, and radiological characteristics	90.7%
Das *et al*., 2020 [25]	3,524	Logistic regression, support vector machine, K nearest neighbor, random forest, and gradient boosting	Sex, age, province, and exposure	90%
Gao *et al*., 2020 [26]	684	Logistic Regression, Support Vector Machine, Gradient Boosted Decision Tree, and Neural Network	Blood biomarkers	94.1%

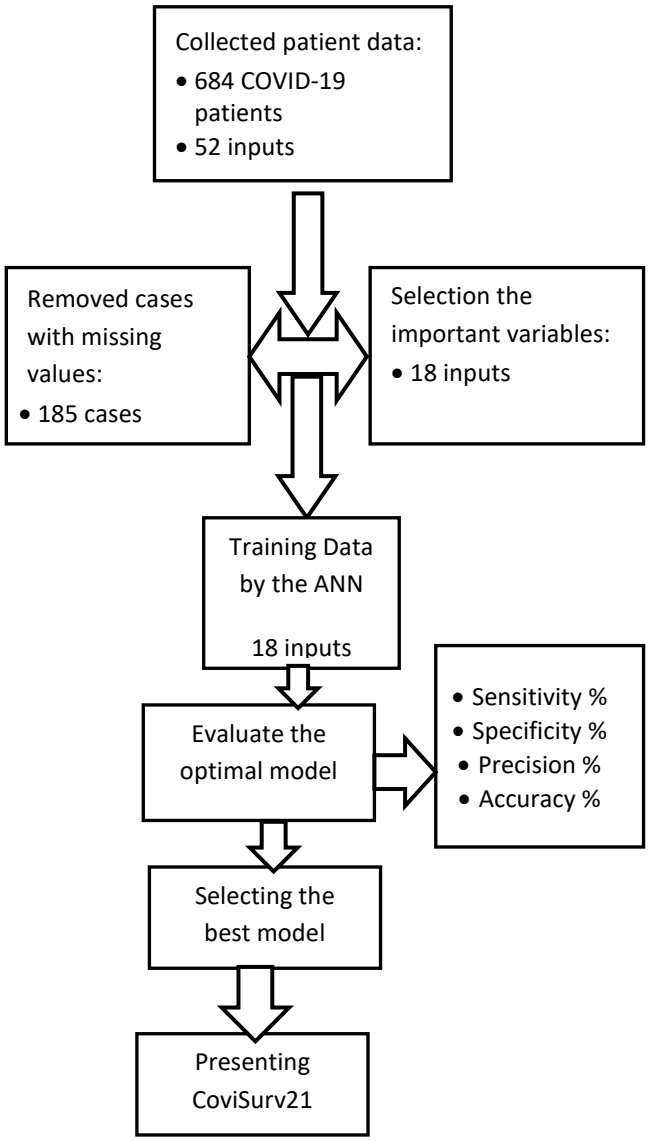

Fig. 1. Structure of the data adopted methodology

Table 2 provided a summary definition of the qualitative variables, while Table 3 provided a summary definition of the quantitative variables. The parameters collected in this part were as follows:

•Personal information of patients (gender and age of the patient),

•Diseases antecedents (Obesity, Insulin-Dependent Diabetes Mellitus, Non-Insulin-Dependent Diabetes Mellitus, High Blood Pressure, Lung Diseases, Cardio-vascular Diseases, and Kidney Diseases),

•Blood biomarkers (Glycemia, Glomerular Filtration Rate, Albumin to Creatinine Ratio, International Normalized Ratio, C-Reactive Protein, White Blood Cells, and Platelet Count)

• Computed Tomography.

Table 2. Database qualitative variables

Qualitative Variables	Abbreviations	Subdivision
Gender of Patient	G_P	1: Male 2: Female
Obesity	OBS	1: NO 2: YES
Insulin-Dependent Diabetes Mellitus	IDDM	
Non-Insulin-Dependent Diabetes Mellitus	NIDDM	
High Blood Pressure	HBP	
Lung diseases	L_Dis	
Cardiovascular Diseases	C_Dis	
Kidney Diseases	K_Dis	

Application of the ANN Model

Artificial Neural Networks (ANNs) are information processing systems that simulate biological neural systems. They consist of neurons that are arranged in various layers, as illustrated in **Fig. 2**.

Table 3. Definition of the database quantitative variables

Quantitatives Variables	Subdivision	Minimum	Maximum
Age of Patient	-----	27	93
Computed Tomography	1: 0% to 25% 2: 25% to 50% 3: 50% to 75% 4: 75% to 100%	0	100
Glycemia	-----	0.1	4.81
Glomerular Filtration Rate	-----	0.12	3.1
Albumin to Creatinine Ratio	-----	6	142
International Normalized Ratio	-----	1	7.8
C-Reactive Protein	-----	1	342
White Blood Cells	-----	1.59	35.01
Platelet Count	-----	31	789
Final element of blood clotting	-----	102	18160

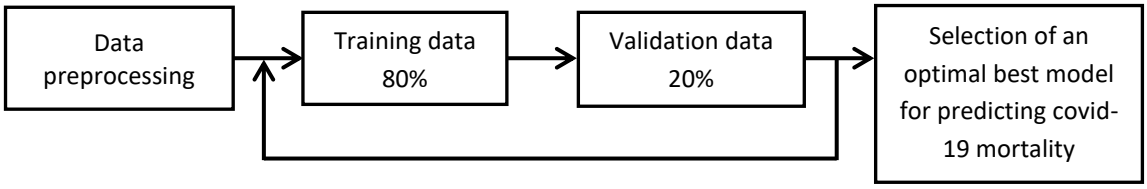

Fig. 2. Structure of the ANNs application

An input layer receives the external data, an output layer gives the problem solution, and a hidden layer is an intermediate layer that separates the other layers. The ANN rely on mathematical algorithms, which generally use the back-propagation algorithm to learn from real datasets, and provide a mathematical model, which could estimate the output parameter from the input ones in future studies [27]. **Fig. 3**. illustrates the structure of the ANN application. In addition, the ANN principal mechanism involves a method that usually follows three stages: training, testing, and validation. The objective was to decrease errors between the output and target values [28].

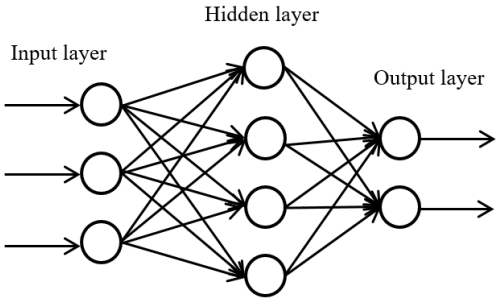

Fig. 3. ANNs architecture

The first stage is the training that aims to regulate iteratively the weight and bias values, frequently by using the popular training algorithm "Back Propagation Algorithm" for the sake of determining the selected stopping criteria. Therefore, the method could offer the optimal model, which is formed by weight and bias values, and transfer function. The latter can be selected based on the type of the understudied problem (*i.e.* linear, log-sigmoid, or tan-sigmoid function) [29,30]. The optimal model will be very handy for estimating effectively the target value when providing the input values, with the minimum error possible. Furthermore, the optimal model was composed by weight, bias, and transfer function suffers from the hard fitting used in future studies by the medical staff. Accordingly, in the current study, we presented the model in the form of easy-to-use and reliable interface.

Performance evaluation of ANN Model

ANNs have been used to train data with a number of varying nodes in the hidden layers, in order to find the optimal one (0–20 nodes in the hidden layer). The performance of each model is assessed according to several measures extracted from the confusion matrix. In the same vein, the confusion matrix is a technique for summarizing and evaluating the performance of a classification model. Hence, it is known as an error matrix, where N is the number of target classes and each row in the matrix represents the predicted class, and each column represents the actual class [31].

A confusion matrix is, generally, utilized to characterize and visualize the performance of the ANN model classifier and to offer an overview about the model misclassifies. For our study, the classification was binary *i.e.* the confusion matrix was a 2x2 matrix, because the target variable had two values: Positive or Negative as shown in **Table 4** [32]. The precision estimation of the suggested models was assessed *via* several performance measures and graphical presentation. The performance measures were Sensitivity, Specificity, Precision, and Accuracy. They are expressed in equations 1, 2, 3, and 4 respectively.

Table 4. Definition of the database qualitative variables

		Actual Values	
		Positive	**Negative**
Predicted Values	Positive	TP	FP
	Negative	FN	TN

TP: True Positive, for correctly predicted event values. FP: False Positive, for incorrectly predicted event values. TN: True Negative, for correctly predicted no-event values. FN: False Negative, for incorrectly predicted no-event values.

$$Sensitivity\ \% = \frac{TP}{TP + FN} * 100 \quad (1)$$

$$Specificity\ \% = \frac{TN}{TN+FP} * 100 \quad (2)$$

$$Precision\ \% = \frac{TP}{TP+FP} * 100 \quad (3)$$

$$Accuracy\ \% = \frac{TN + TP}{TN + TP + FP + FN} * 100 \quad (4)$$

Results

Descriptive data analysis

The database employed in this study consisted of a sample of 684 individuals (190 death and 494 recoveries), with 18 attributes (10 quantitative and 8 qualitative features. According to SPSS treatment, **Table 5** shows the descriptive statistics of the quantitative variables, which were very dispersed. Furthermore, a considerable effort has been done in order to gather as much patient information as possible for using a well-established database. The findings indicated that all variables were regularly distributed. Likewise, results confirmed that dataset included a wide range of data. Consequently, this dataset could be utilized to evolve new empirical models.

Evaluation of the COVID-19 mortality using ANN

To determine the most appropriate ANN model, the principal phase involved choosing the ideal input factors, which highly influenced on the target one, the next stage, was to define the optimal nodes number in the hidden layer. In this regard, to define the suitable input factors of the ANN method, both the antecedent diseases and blood biomarkers parameters have been utilized based on the literature recommendation and expert medical team consultation. Afterward, we have tried to determine the ideal number of nodes in the optimal ANN model for assessing COVID-19 mortality depending on four performance measures. The efficiency of each model during the training and validation phases is illustrated in **Table 6**. Four indicators measures were utilized in order to compare the suggested models to choice the optimal one in terms of sensitivity, specificity, precision, and accuracy.

The used data were separated into two sections, *i.e*. 80% for training and 20% for validation. As **Table 6** proved, the COVID-19 mortality was modeled using ANN method, where the network architecture characteristics have been changed, and compared using four indicator measurements to determine the optimal one. The diverse models produced sensitivity (0.92 to 1), specificity (0.639 to 1), precision (0.792 to 1), and accuracy (0.831 to 1) in the training phase. Similarly, in the validation phase, sensitivity (0.72 to 1), specificity (0.467 to 0.846), precision (0.692 to 0.909), and accuracy (0.676 to 0.865) were obtained. Results indicated that the best performance was obtained from the ANN model with 12 neurons in the hidden layer (18-12-1) trained by Tan-Sigmoid function. This model was deemed to be the optimal one with the purpose that it produced the highest precision in terms of sensitivity (1/0,955), specificity (1/0,733), precision (1/0,84), and accuracy (1/0,865) during the training/validation phase. Finally, the optimal ANN model revealed the higher results in both training and validation data.

Generation of the best appropriate ANN model

Fig. 4. presents the architecture of the optimal ANN model. The latter was effectively trained in 25 steps. The results proved that the optimal mean squared error was the seventeenth one (MSEvalidation =0.0118, MSEtrainng=0.0046, and MSEall= 0.0074), as presented in **Fig. 5**. The errors histogram found from the results of modeling the optimal ANN model is presented in **Fig. 6.** The green and blue bars denoted validation and training data, respectively. The findings proved that the majority of errors between the output and target values ranged between -0.05 and 0.05.

Table 6. Descriptive statistics for the quantitative data

Variables	Minimum	Maximum	Mean	SD	Coefficient of Variation %
Age	27	93	65.57	13.68	21
Glycemia	0.10	4.81	1.57	0.76	48
Glomerular Filtration Rate	0.12	3.10	0.60	0.48	79
Albumin to Creatinine Ratio	6	142	16.98	17.05	100
International Normalized Ratio	1	7.80	1.76	5.83	41
C-Reactive Protein	1	342	66.36	48.95	74
White Blood Cells	1.59	35.01	10.21	5.90	58
Platelet Count	31	789	302.35	131.24	43
Final element of blood clotting	102	18160	1867.32	2621.25	140

Table 6. Changes of the performance measures with the number of nodes in the hidden				
Number of nodes	**Sensitivity**	**Specificity**	**Precision**	**Accuracy**
Training data				
1	0.989	0.644	0.807	0.851
2	1.000	0.734	0.832	0.885
3	0.966	0.639	0.792	0.831
4	0.953	0.726	0.828	0.858
5	0.952	0.692	0.798	0.838
6	0.943	0.754	0.845	0.865
7	0.920	0.885	0.920	0.905
8	0.955	0.850	0.903	0.912
9	0.952	0.862	0.898	0.912
10	0.966	0.871	0.913	0.926
11	0.954	0.951	0.965	0.953
12	1.000	1.000	1.000	1.000
13	0.976	0.848	0.889	0.919
14	1.000	0.984	0.988	0.993
15	1.000	0.984	0.989	0.993
16	0.933	0.881	0.922	0.912
17	0.977	0.903	0.933	0.946
18	0.989	0.984	0.989	0.986
19	1.000	0.984	0.989	0.993
20	1.000	0.938	0.955	0.973
Validation data				
1	0.947	0.722	0.783	0.838
2	0.917	0.538	0.786	0.784
3	0.810	0.688	0.773	0.757
4	1.000	0.467	0.733	0.784
5	0.880	0.583	0.815	0.784
6	0.952	0.625	0.769	0.811
7	0.905	0.688	0.792	0.811
8	0.850	0.647	0.739	0.757
9	0.720	0.583	0.783	0.676
10	0.857	0.583	0.783	0.758
11	0.810	0.750	0.810	0.784
12	0.955	0.733	0.840	0.865
13	0.846	0.636	0.846	0.784
14	0.783	0.714	0.818	0.757
15	0.818	0.733	0.818	0.784
16	0.789	0.722	0.750	0.757
17	0.818	0.467	0.692	0.676
18	0.857	0.688	0.783	0.784
19	0.810	0.688	0.773	0.757
20	0.833	0.846	0.909	0.838

Fig. 7. displays the scatter plots of the target and output values in the optimal ANN model. Results demonstrated high correlation coefficient: (R training=0.999, R validation=0.747, and R all=95).
Furthermore, the evaluation of the classifier should, preferably, be performed by the confusion matrix. The latter was an error matrix described by a table layout in order to help investigators in assessing the effectiveness of the classifier. In the confusion matrix, the instances in the target class are generally represented in the columns, whereas the instances in the output class are represented in the rows, as illustrated in **Fig. 8**.

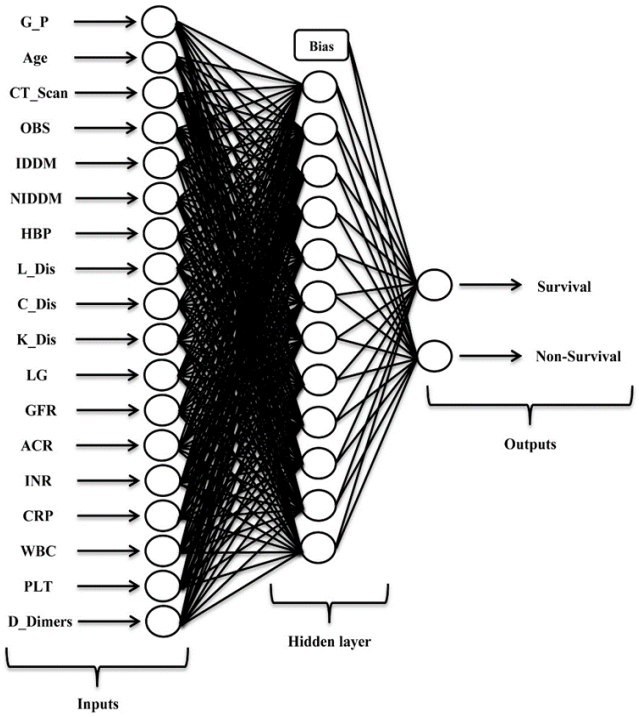

Fig. 4. The architecture of the most appropriate ANN model (18-12-2)

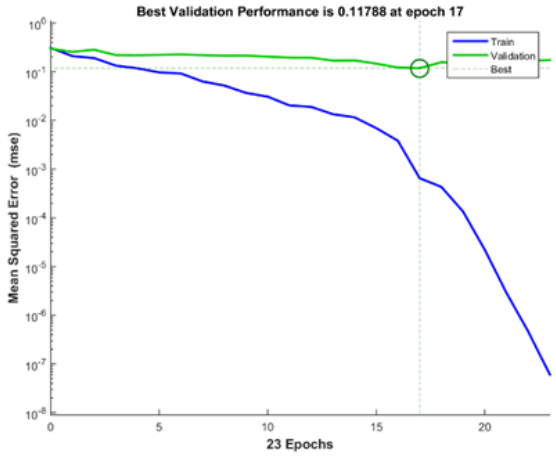

Fig. 5. Best performance epoch of the most appropriate ANN model (18-12-2)

Results indicated a good accuracy: 1 for training data, 0.87 for validation data, and 0.973 for all data. The latter means that the ratio of distinguishing the survival patient from the true survival patient was 96.4% in all data, the ability to distinguish the non-survival patients from the true non-survival ones was 98.6% and the ratio of a total number of the correct diagnoses of all data as survival and non-survival was 97.3%. On the other hand, **Fig. 9** illustrates the ROC analysis of training and validation data. It was worthy to mention that ROC analysis was utilized to define the true precision of the optimal model, determine the correlation among diagnosis sensitivity and specificity. The axes of the curves were true positive (calling the survival survival) and false positive (calling the non-survival survival). The curves were between the limits of 0 and 1, while the proximity to coordinate y and upper boundary indicated good results; the curves that had an inclination of 45°proved an ineffective learning of the model. The results of **Fig. 9** indicated clearly the presence of powerful good learning of the proposed ANN model. To test the effectiveness of the suggested ANN model, a comparative study was performed with 11 empirical models proposed in the literature for predicting COVID-19 mortality, as presented in Table 1. The comparison has been made depending on the accuracy of the optimal model. Published studies have proved that the accuracy was a significant indicator in evaluating the forecast accuracy, as the optimal model was characterized by an accuracy close to 100%. The findings proved that the proposed ANN model in our study was the best performing model, with the maximum accuracy (97.3 for all data).

Fig. 6. Plot-Error for classification by the ANN model

Graphical User Interface (GUI) design "CoviSurv 2021"

It is a popular usage detected in the majority of the studies using AI methods to present models in the

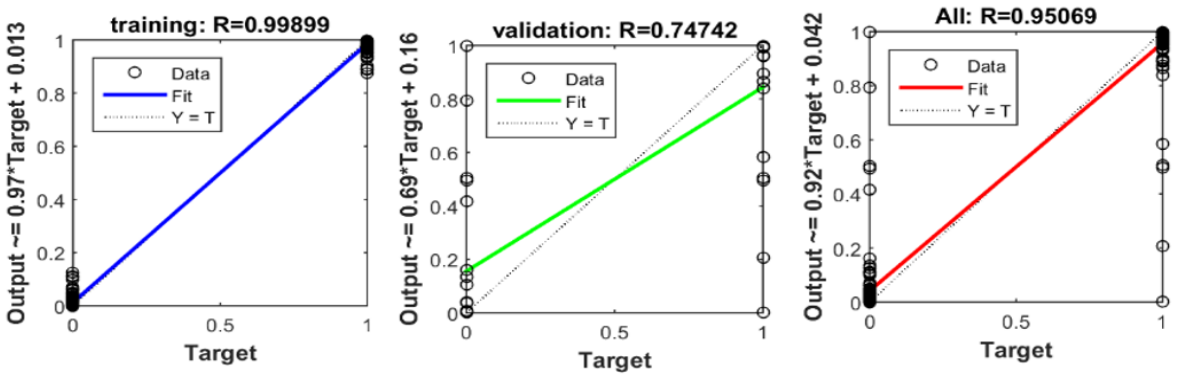

Fig. 7. Plot-Regression between the output and target value in the ANN model (18-12-1)

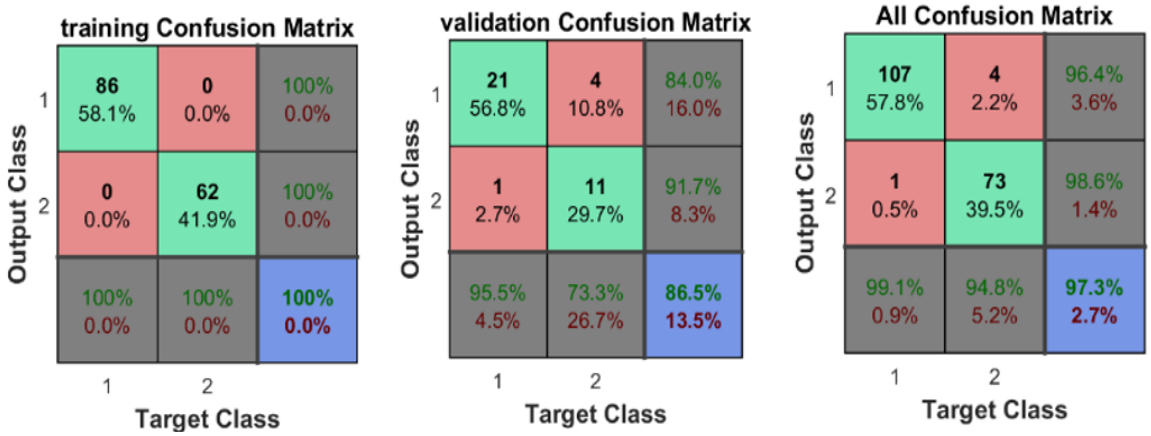

Fig. 8. Plot-Confusion between the output and target value in the ANN model (18-12-1)

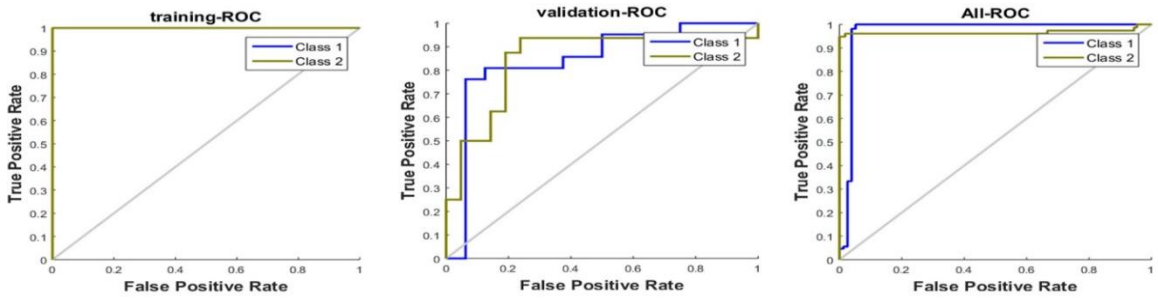

Fig. 9. Plot-Roc for classification by the ANN model

form of equations, which make the proposed model hard to use in the future. Certainly, this idea has useless significance for other researchers and medical staff. In the sake of doing a beneficial effort, the suggested ANN model should be presented in the form of a programmed Interface or a simple script by using a well-known programming language like Matlab and Python [33]. Accordingly, the ANN model can be easily used thus obtainable to anybody concerned in the modeling problematic.

In the current study, a graphical easy-to-use interface was presented depending on the best appropriate ANN model, as illustrated in **Fig. 10**.

The suggested model was subsequently utilized to design a GUI open-access interface. Matlab software has been used to program the interface so-called “CoviSurv2021”. The motive behind selecting this name was due to "Covi" relative to "COVID-19", “Surv” relative to survival to make as many people as possible survive, and 2021 was the year this interface was designed. CoviSurv2021 included patient personal information, computed tomography, antecedent diseases, and blood biomarkers. Initially, the user must define the patient personal information: choosing sex (male or female), enter age, and computed tomography pulmonary damage

ratio (0% to 25%, 25% to 50%, 50% to 75%, or 75% to 100%). Secondly, to answer yes or no whether the patient had antecedent diseases, such as obesity, Insulin-Dependent Diabetes Mellitus (type 1), Non-Insulin-Dependent Diabetes Mellitus (type 2), High Blood Pressure, Lung Diseases, Cardiovascular Diseases, and Kidney Diseases. Thirdly, the user needed to present some blood biomarkers analyses (Glycemia, Glomerular Filtration Rate, Albumin to Creatinine Rate, International Normalized Ratio, C-Reactive Protein, White Blood Cells, Platelet Count, and D-dimer). Finally, by clicking on Run, the classification result appeared in the outputs, either survival or non-survival that was, the death probability was high. The suggested CoviSurv2021 interface will be very useful for all medical staff, helping them to select patients who necessities to get attention firstly, who had the upper urgency to be hospitalized, prioritize patients when the hospital is overcrowded, and profit time to provide the needed care.

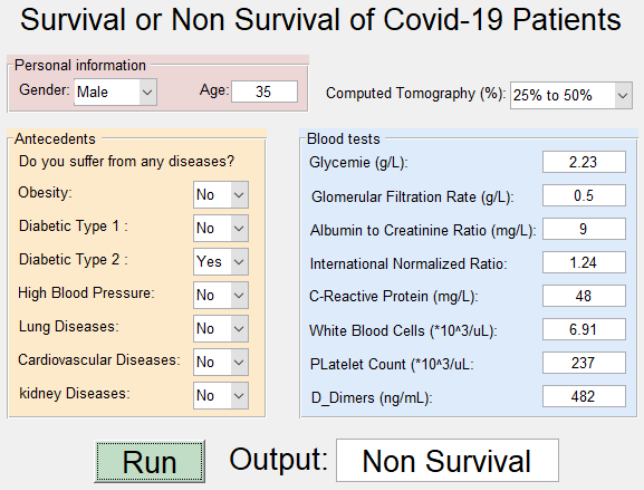

Fig. 10. CoviSurv2021 interface

Discussion

After the worldwide spread of COVID-19, the classification of patient-mortality risk was of great importance in both prevention, and treatment allocation, especially the last version “Omicron”, which spread faster, leading to over-crowding hospitals, and so the inability of the medical staff to treat large numbers of patients. For this exact reason,there was an urge to understand and define the pandemic behavior. Several studies have pointed to the existence of patient-related factors affecting COVID-19 patients’ mortality-risk. In the current study, ANN has been employed to predict COVID-19 mortality starting from collected antecedent diseases, and blood biomarkers data of patients. Only parameters that can be utilized as risk factors for the evaluating of likely mortality in COVID-19 patients have been adopted. This parameter was chosen based on the literature recommendation and expert medical team consultation. Afterward, a high accuracy model was developed depending on the ANN modeling which used a large number of data. The findings clearly indicated that the optimal model was the one containing 12 nodes in the hidden layer and trained by Tansig transfer function (18-12-2). The latter presented the optimal model, which provided the higher values in terms of sensitivity, specificity, precision, and accuracy, compared to the other models. Furthermore, the newly developed model was assessed by measuring its accuracy compared with other proposed models in the literature. The conclusion was that the optimal ANN model was more precise than the proposed empirical models. Furthermore, our model was closely followed by the SVM model which was proposed by Gao *et al.,* [26], it gave an acceptable accuracy and ranked second. Moreover, the results revealed the poor performance of the logistic regression model proposed by Bhandari *et al*., [17] in predicting COVID-19 mortality. With respect to the performance of machine learning models, the hierarchy followed the order of the presented model (in our study), models of Gao *et al.*, [26], Li *et al*., [24], Pourhomayoun and Shakibi, [18], Ko *et al*., [19], Aktar *et al*., [20], Assaf *et al*., [22], Hu *et al*., [21], Chowdhury *et al*., [23], Tezza *et al*., [10], and Das *et al*., [25]. Finally, the proposed optimal model was afterward used to develop a GUI public interface in order to facilitate its usage in future. A graphical reliable and easy-to-use interface was, afterward, presented in the current study, dubbed “CoviSurv2021”, which was programmed by Matlab software. The essential benefit of “CoviSurv2021” was to aid hospitals and medical services in selecting priority patients who have upper urgency to be hospitalized, prioritizing patients when the hospital is over-crowded, and gaining time to provide the care needed. Furthermore, both the antecedent diseases and blood biomarkers analysis were usually fast, affordable, and promptly accessible in hospitals. As a maximum, these parameters may last for twenty-four hours in the same health facility where patients are receiving the treatment. In other words, we can expect patients who suffer from the mortality-risk in the early stages of their hospitalization, giving them the appropriate care, while patients who are not at risk can be directedfor treatment at home to avoid the risk of overcrowding in hospitals. Our findings

represented a crucial contribution to the medical field. The elaborated model built in our study represented a reliable tool for predicting COVID-19 mortality at the early stage of hospital admission. The performance of the estimation has been highly developed compared to other models proposed in the literature.

In this pandemic crisis, the shortage of medical and personal resources is causing major problems for the health system, especially in third world countries. As a result, AI can play a significant role in COVID-19 patients management. In addition, CoviSurv2021 had several advantages. First, CoviSurv2021 could predict which patients were at high risk of death at the early stage of hospital admission (that was, within 24 hours of admission). Predicting mortality at the time of admission could be very informative for clinicians because the critical period that witnesses the disease propagation was from 10 to 14 days from the onset of symptoms, according to previous studies. Additionally, CoviSurv2021 could also provide advice on treatment priorities for people who needed to be treated intensively from day one of illness. While people who were not at risk of dying could be referred for home treatment to avoid the risk of overcrowding in hospitals. Second, the mortality results predicted by AI were explainable and easily understood by medical researchers. Although the interface to be offered will be easily usable by clinicians, nurses, and even patients at home. Finally, database collected from hospitals could help doctors and researchers in future studies to better understand the complex behavior of COVID-19.

Despite the amazing results offered in the current study, a number of significant shortcomings needed to be discussed. The important shortcomings was the small sample size used in the modelling, which could influence on the mortality identification accuracy. This might conduct to the suggested model incapability to predict from the new data input that were not utilized in the training phase. Additionally, the small database used in machine learning modelling could generate over-fitting and under-fitting difficulties. Therefore, investigators generally use large and diverse database, which was gathered from transferring knowledge between them. This was a significant concern to be taken in consideration for future research base on collecting data from several countries to improve its learning and, therefore, generating a well model. Further modeling utilizing meta-heuristic algorithms on the prediction of COVID-19 mortality, were powerfully suggested. It was worthy to note that meta-heuristic algorithms had demonstrated high-efficiency finding combined with machine learning methods, conducting to enhancing their training and rapidly converging to the optimal parameters [34-36]. Demographic, genetic, and sociological factors could have an influence on the mortality of COVID-19 patients, with that being said, we highly encourage further studies on these parameters. We seek to further develop CoviSurv 2021 interface for the purpose of turning it into an information source to help hospitals and medical services.

Conclusion

This study contributes to providing a simple, reliable, and easy-to-use interface for the prediction of risk mortality of patients suffering from COVID-19 dubbed “CoviSurv2021”. Firstly, a considerable number of the antecedent diseases and blood biomarkers data of patients have been collected. The selection of this parameter as risk factors for evaluating the likely mortality in COVID- 19 patients are done based on literature recommendation and expert medical team consultation. Eighteen parameters have been selected in the input layer. To achieve our aim, several ANN modelsare presented for a practical analysis aimed at modeling the COVID-19 mortality based on the antecedent disease and blood biomarker data of the patients. Afterward, the proposed models are evaluated via four performance measures (sensitivity, specificity, precision, and accuracy). The comparison of the results assessment between the different proposed models reveals the superiority of the (18-12-2) model, which gives the highest accuracy during both the training/validation phases. The accuracy values are used to compare the performances of the optimal ANN model with the proposed models in the literature. The findings indicate that the aforementioned ANN model is more effective than the empirical models. Finally, the proposed optimal model is afterward used to develop a GUI public interface by Matlab software. A reliable and easy-to-use graphical interface is afterward presented in the current study dubbed “Covi-Surv2021”. The fundamental benefit of “CoviSurv2021” is to facilitate for the medical staff the selection of first priority patients who has upper urgency to be hospitalized, prioritize patients when the hospital is over-crowded, and gain time to provide the care needed.

This work has opened up several questions that need further investigations to develop certain limitations. Firstly, there is a need to utilize more data from other countries to improve the training phase. Secondly, we

recommend the utilization of meta-heuristic algorithms combined with machine learning methods for predicting COVID-19 mortality in future studies. These algorithms have proved high-efficient findings, conducting to enhance their training and rapidly converging to the optimal solution.

Acknowledgments

The authors are greatly indebted to Dr Khelifa Azzouza, Dr Asma Haddad, and Mrs Nesrine Benachour for providing consultations and medical advice that helped us move forward with our work. We are extremely grateful to both the administrative and the nursing staff at Jijel Hospital, and at the top of them Mr Bilal Daas, Mr Nadir Haddad, Mr Salah Meriche, Mrs Samia Himeur, and Mrs Aldjia Karaoui for their valuable support. Also, we would like to express our thank to Mr Abdelkader Mousouni for the logistical and administrative support. Finally, we express our special thanks to Ms Asma Djaidja for the special proof reading to the current paper.

Conflict of interests

The authors declare no conflict of interests.

References

1. Cherry JD. The chronology of the 2002–2003 SARS mini pandemic. *Paediatr Respiratory Rev* 2004;5(4): 262-9.
2. Anderson RM.,Fraser C., Ghani AC., Donnelly CA., Riley S., Ferguson, NM. et al. Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. *Philosophical Transactions Royal Society London. Series B: Biological Sciences*. 2004;359(1): 1091-105.
3. Azhar E., Hui D., Memish Z., Drosten C., Zumla A. The middle east respiratory syndrome (MERS). *Infectious Dis Clinics* 2019;33(4): 891-905.
4. Zumla A., Hui D., Stanley Perlman P. Middle East Respiratory Syndrome HHS Public Access. *Lancet* 2015; 386(1): 995-1007.
5. Albaser A. Corona Virus (SARS CoV-2) that causes COVID-19. *J Pure Appl Sci* 2020;19(1): 120-6.
6. National Center for Immunization and Respiratory, D. of V.D February 15th 2020. Human Coronavirus Types. https://www.cdc.gov/coronavirus/types.html
7. Brunese L., Mercaldo F., Reginelli A., Santone A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. *Computer Methods Programs Biomedicine* 2020;196: 105608.
8. Burdick H., Lam C., Mataraso S., Siefkas A., Braden G., Dellinger R. et al. Prediction of respiratory decompensation in COVID-19 patients using machine learning: The READY trial. *Computers Biology Medicine* 2020;124(1): 103949.
9. Jung J. A long way to the recovery: COVID-19 will not disappear. *J Korean Medical Sci* 2021;36: 1-4.
10. Tezza F., Lorenzoni G., Azzolina D., Barbar S., Leone LA., Gregori D. Predicting in-Hospital Mortality of Patients with COVID-19 Using Machine Learning Techniques. *J Personalized Medicine* 2021;11(5): 343-66.
11. Pannu A. Artificial intelligence and its application in different areas. *Int J Engineering Innovative Technology* 2015;4(10): 79-84.
12. Yaşar Ş., Çolak C., Yoloğlu S. Artificial Intelligence-Based Prediction of COVID-19 Severity on the Results of Protein Profiling. *Computer Methods Programs Biomedicine* 2021; 202(1): 105996.
13. Bioud NI., Ouled Laid I., Benbouras MA. Estimating the fundamental period of infilled RC frame structures via deep learning. *Urbanism Architecture Constructions* 2022;14(1): 1-22.
14. Li M., Zhang Z., Cao W., Liu Y., Du B., Chen C. et al. Identifying novel factors associated with COVID-19 transmission and fatality using the machine learning approach. *Science Total Environment* 2021;764(1): 1-10.
15. Al Youha S., Doi SA., Jamal MH., Almazeedi S., Al Haddad M., Al Seaidan M. et al. Validation of the Kuwait Progression Indicator Score for predicting progression of severity in COVID-19. *MedRxiv* 2020;11(1): 1-15.
16. Zhang B., Zhou X., Qiu Y., Song Y., Feng F., Feng J. et al. Clinical characteristics of 82 cases of death from COVID-19. *PloS one* 2020;15(1): 1-8.
17. Bhandari S., Shaktawat AS., Tak A., Patel B., Shukla J., Singhal S. et al. Logistic regression analysis to predict mortality risk in COVID-19 patients from routine hematologic parameters. *Ibnosina J Medicine Biomedical Sci* 2020;12(2): 123-9.
18. Pourhomayoun M., Shakibi M. Predicting mortality risk in patients with COVID-19 using artificial intelligence to help medical decision-making. *MedRxiv* 2020; 11(1): 1-5.
19. Ko H., Chung H., Kang WS., Park C., Kim DW., Kim SE. et al. An artificial intelligence model to pre-

dict the mortality of COVID-19 patients at hospital admission time using routine blood samples: development and validation of an ensemble model. *J Med Internet Res* 2020; 22(12): e25442.
20. Aktar S., Ahamad MM., Rashed-Al-Mahfuz M., Azad A., Uddin S., Kamal AH. et al. Predicting Patient COVID-19 Disease Severity by means of Statistical and Machine Learning Analysis of Clinical Blood Testing Data. *JMIR Medical Informatics* 2021;9(2): 1-15.
21. Hu C., Liu Z., Jiang Y., Zhang X., Shi O, Xu K. et al. Early prediction of mortality risk among severe COVID-19 patients using machine learning. *MedRxiv* 2020;10(1): 1-7.
22. Assaf D., Gutman Y., Neuman Y., Segal G., Amit S., Gefen-Halevi S. et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Internal Emergency Medicine* 2020;15(1): 1435-43.
23. Chowdhury M., Rahman T., Khandakar A., Al-Madeed S., Zughaier SM., Doi SA. et al. An early warning tool for predicting mortality risk of COVID-19 patients using machine learning. *Cognitive Computation* 2021;14(1): 1-6.
24. Li S., Lin Y., Zhu T., Fan M., Xu S., Qiu W. et al. Development and external evaluation of predictions models for mortality of COVID-19 patients using machine learning method. *Neural Computing Applications* 2021;5(2): 1-10.
25. Das A., Mishra S., Gopalan S. Predicting COVID-19 community mortality risk using machine learning and development of an online prognostic tool. *Peer J* 2020;8(1): 1-7.
26. Gao Y., Cai G., Fang W., Li H., Wang S., Chen L. et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nature Communica-tions* 2020;11(1): 1-11.
27. Sairamya N., Susmitha L., George S., Subathra M. Hybrid Approach for Classification of Electroencephalographic Signals Using Time–Frequency Images With Wavelets and Texture Features. *Intelligent Data Analysis Biomedical Applications* 2019; 253-73.
28. Benbouras M. Predicting shear stress parameters in consolidated drained conditions using artificial intelligence methods. *Basic Appl Sci-Sci J King Faisal Univ* 2021;22(1): 1-7.
29. Benbouras M., Kettab Mitiche R., Zedira H., Petrisor A., Mezouar N., Debiche F. A new approach to predict the compression index using artificial intelligence methods. *Marine Georesources Geotechnology* 2018;37(3): 704-20.
30. Benbouras M., Kettab RM., Zedira H., Debiche F., Zaidi N. Comparing nonlinear regression analysis and artificial neural networks to predict geotechnical parameters from standard penetration test. *Urbanism. Architecture. Constructions* 2018;9(1): 273-86.
31. Ting K. Encyclopedia of Machine Learning and Data Mining, chap. Confusion Matrix, Springer US, Boston; 2017, 1-260.
32. Haddad F., Benbouras M. Application of Artificial Neural Networks Models in Diabetes Mellitus Classification. *Models Optim Math Anal J* 2020; 8(1): 14-21.
33. Tiwari A., Dadhania AV., Ragunathrao VA., Oliveira ER. Using machine learning to develop a novel COVID-19 Vulnerability Index (C19VI). *Sci Total Environment* 2021;773(1): 1-12.
34. Benbouras M., Petrisor A. Prediction of Swelling Index Using Advanced Machine Learning Techniques for Cohesive Soils. *Applied Sci* 2021; 11: 1-12.
35. Kuo C., Fu J. Evaluating the impact of mobility on COVID-19 pandemic with machine learning hybrid predictions. *Sci Total Environment* 2021; 758(1): 144-51.
36. Benbouras MA. Hybrid meta-heuristic machine learning methods applied to landslide susceptibility mapping in the Sahel-Algiers. *Int J Sediment Res* 2022 (in press).