Iraqi Journal of Science, 2023, Vol. 64, No. 2, pp: 922-943
DOI: 10.24996/ijs.2023.64.2.36

ISSN: 0067-2904

Benchmarking Framework for COVID-19 Classification Machine Learning Method Based on Fuzzy Decision by Opinion Score Method

Mahmood M. Salih[1]*, M. A. Ahmed[1], Baidaa Al-Bander[2], Kahlan F. Hasan[3], Moceheb Lazam Shuwandy[1], Z.T. Al-Qaysi[1]

[1] *Department of Computer Science, College of Computer Science and Mathematics, Tikrit University, Tikrit, Iraq*
[2] *Department of Computer Engineering, College of Engineering, University of Diyala, , Iraq*
[3] *Informatics institute, Istanbul Technical University, Istanbul, Turkey*

Received: 25/11/2021 Accepted: 10/3/2022 Published: 28/2/2023

Abstract

Coronavirus disease (COVID-19), which is caused by SARS-CoV-2, has been announced as a global pandemic by the World Health Organization (WHO), which results in the collapsing of the healthcare systems in several countries around the globe. Machine learning (ML) methods are one of the most utilized approaches in artificial intelligence (AI) to classify COVID-19 images. However, there are many machine-learning methods used to classify COVID-19. The question is: which machine learning method is best over multi-criteria evaluation? Therefore, this research presents benchmarking of COVID-19 machine learning methods, which is recognized as a multi-criteria decision-making (MCDM) problem. In the recent century, the trend of developing different MCDM approaches has been raised based on different perspectives; however, the latest one, namely, the fuzzy decision by opinion score method that was produced in 2020, has efficiently been able to solve some existing issues that other methods could not manage to solve. because of the multiple criteria decision-making problem and because some criteria have a conflict problem. The methodology of this research was divided into two main stages. The first stage related to identifying the decision matrix used eight different ML methods on chest X-ray (CXR) images and extracted a new decision matrix so as to assess the ML methods. The second stage related to FDOSM was utilized to solve the multiple criteria decision-making problems. The results of this research are as follows: (1) The individual benchmarking results of three decision makers are nearly identical; however, among all the used ML methods, neural networks (NN) achieved the best results. (2) The results of the benchmarking group are comparable, and the neural network machine learning method is the best among the used methods. (3) The final rank is more logical and closest to the decision-makers' opinion. (4) Significant differences among groups' scores are shown by our validation results, which indicate the authenticity of our results. Finally, this research presents many benefits, especially for hospitals and medical clinics, with a view to speeding up the diagnosis of patients suffering from COVID-19 using the best machine learning method.

Keywords: COVID-19, Evaluation and benchmarking, Machine learning, multi-criteria decision making, MCDM, Fuzzy Decision by Opinion Score Method.

*Email: mahmaher1989@gmail.com

اطار قياس طرق تعلم الماكنة لتصنيف كوفيد – 19 باستعمال القرار الضبب المبني على درجة الرأي

محمود ماهر صالح[1]*, محمد أكثم احمد[1], بيداء البندر[2], كهلان فائق حسن[3] , مصحب لزام شوندي[1], زيدون طارق عبدالوهاب[1]

[1]علوم الحاسوب, كلية علوم الحاسوب و الرياضيات, جامعة تكريت, صلاح الدين, العراق

[2]قسم هندسة الحاسوب, كلية الهندسة, جامعة ديالى, ديالى, العراق

[3]معهد المعلومات, اسطنبول التقنية, اسطنبول, تركيا

الخلاصة

اعلنت منظمة الصحة العالمية في العام 2020 عن الجائحة التي تسبب بها كوفيد – 19 و التي ادت الى توقف العالم و فرض منع التجوال في جميع مناطق المعمورة. و ادت هذه الجائحة الى شلل في المنظومة الصحية العالمية و في جميع دول العالم. مما دعا الباحثين في شتى المجالات العلمية الى تقديم خدماتهم من اجل الخلاص من هذا الوباء. و احد ابرز الطرق التي استعملت في تشخيص كوفيد – 19 هو التعلم الماكنة. علماً ان هناك العديد من الطرق التي استعملت من اجل تصنيف مرض كوفيد–19 و لكن السؤال ما هي الطريقة الافضل بأخذ النظر معايير متعددة للتقييم في الوقت الواحد؟ في هذا البحث قدم الباحثون اطار لقياس و تحديد اي نوع من طرق تعلم الماكنة هو الافضل بالاستناد الى معايير مختلفة. حيث تم استعمال احدث طرق اتخاذ القرار بالاعتماد على معايير متعددة و هي طريقة القرار المبني على الرأي الضبابية. ان منهجية هذا البحث تقسم الى قسمين: القسم الاول متعلق بكيفية تكوين مصفوفة القرار بالاعتماد على ثمان طرق من طرق تعلم الألة و تسعة معايير للتقييم. اما القسم الثاني فيرتبط بكيفية استعمال طريقة القرار المبني على الرأي الضبابية من اجل اتخاذ قرار اي من الطرق هي الافضل. و قد اظهرت النتائج ان طريقة الشبكة العصبونية هي الافضل على ضوء القرار متعدد المعايير. ان اهمية هذا البحث تكمن في اختيار افضل طريقه من طرق تعلم الماكنة و التوصية باستعمالها في المستشفيات و المراكز الصحية.

1. Introduction

Coronavirus disease (COVID-19), caused by SARS-CoV-2, has been announced as a global pandemic by the WHO, which results in the collapsing of the healthcare systems in several countries around the globe [1]. It is an RNA-type virus that causes a wide range of serious and harsh respiratory infections targeting both humans and animals [2]. Coronavirus is typically transmitted from animal to human, but nowadays it is transmitted among people by modifying its form. COVID-19 has emerged as a dangerous virus capable of causing a worldwide pandemic [3]. Thousands of people have lost their lives as a result of this virus, and its harmful effects and consequences on public health are still ongoing and unresolved [4].

As there are no particular remedies or vaccines for COVID-19, specialists, in order to develop a potential vaccine, are testing and evaluating various clinical trials. In spite of the lack of a vaccine, infection can be avoided by following certain precautionary procedures, including staying home, washing hands, quitting smoking, and covering the mouth and the nose when sneezing or coughing. The abovementioned precautions would not prevent the virus; they, however, can protect people from it (COVID-19) and slow down its spread [5]. Therefore, early detection of COVID-19 patients is essential for disease control and cure [6].

The shortage of diagnostic tools and the constraints on their development have slowed the identification of disease, and as a result, the number of patients and casualties has increased. The incidence of COVID-19 disease would be reduced if it were detected and diagnosed early [7]. Researchers work tirelessly to find potential solutions that will aid in the control of the pandemic in their respective areas. Analyzing lungs' images for COVID-19 taken by CT scans and X-rays is one of the most common and successful approaches used by researchers [8].

These imaging modalities involve specialists in radiology for manual inspection of each patient case, which takes time and effort and is therefore a difficult and challenging task [9]. Although the diagnosis based on radiological images is a fast process and also has some advantages over the PCR test in terms of recognition accuracy in the earlier phases of the COVID-19, the system's backbone is the need for experts to understand the images. Basically, diagnostic strategies based on artificial intelligence (AI) will allow experts to obtain a precise and straightforward description of the X-ray images to identify COVID-19 [10, 11]. The provision of healthcare includes the advancement of emerging technologies such as AI, machine learning (ML), big data, and the Internet of Things (IoT) to tackle new diseases [12]. With a view to monitoring the disease, AI can be utilized in tracking the spread of COVID-19 based on location and time.

It has been marked by persisting observations that COVID-19 has respiratory behaviors that differ from normal cold and seasonal influenza, showing extreme tachypnea (fast breathing) [13]. Machine and deep learning have become established and prestigious disciplines in deploying artificial intelligence to mine, analyze, identify, and recognize patterns in data. Increasing the size of clinical data, varying data sources, and the advances in those fields have enabled us to get the benefit of clinical decision making and computer-aided systems, which are increasingly vital [14]. Besides, as the growth rate of COVID-19 is non-stationary and non-linear, maintaining excellence in the healthcare process and accurately predicting COVID-19 play a significant role. Recently, various machine learning models have been used for COVID-19 prediction, such as ANN [15], the K-Nearest Neighbors (KNN) classifier [6], Support Vector Machines (SVM), Naive Bayes (NB), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Random Forest (RF), and Decision Trees (DT) [16].

On the other hand, two common criteria are used in the literature to evaluate ML algorithms that were applied for COVID-19 diagnosis, including (i) group reliability and (ii) time complexity. Furthermore, several sub-criteria belonging to the reliability group have been considered, including but not limited to F1-score, precision, average accuracy, error rate, recall, true negative (TN), true positive (TP), false negative (FN), and false positive (FP) [17] and AUC [5]. Table 1 reports the recent non-clinical techniques represented by machine learning algorithms along with the criteria used to evaluate COVID-19 pandemic diagnostic models.

Table 1: The existing ML models applied in the literature for COVID-19 pandemic diagnosis utilizing radiography images

Study	ML Models	Image Type	# Classes	F1-Score	Sp	Sn	Pr	Acc
SVM classifies ResNet50's deep features [1]	SVM	X-Ray	3	95.34	×	95.33	×	95.33
Infection Size Adaptive Random Forest method with decision tree with ML models (LR, SVM, NN) [2]	LR, SVM, NN	CT	2	×	83.30	90.70	×	87.90
MobileNetV2 and SqueezeNet with SVM [3]	SVM	X-Ray	2	×	98.58	99.63	98.33	98.89
Multi-Level Thresholding with SVM [4]	SVM	X-Ray	2	×	99.70	95.76	×	97.48
ResNet152 model with Random Forest and XGBoost classifiers [5]	RF, XGBoost	X-Ray	3	97.7	98.8	97.7	×	97.70
Traditional ML models (SVM, DT, MLP, kNN, RF) [6]	SVM, DT, MLP, kNN, RF	X-Ray	7	×	89.0	×	×	×
Residual Exempler Local Binary Pattern with ML models (DT, SVM, kNN) [7]	DT, SVM, kNN	X-Ray	2	×	100.0	98.85	×	99.69
ML-based classifier including Decision Tree, Ensemble, kNN, 3-naïve Bayes, and SVM [8]	Decision Tree, Ensemble, kNN, 3-naïve Bayes, SVM	CT	2	×	90.32	93.54	90.63	91.94
Fast Fourier Transform (FFT) with SVM model [9]	SVM	CT	2	×	94.76	95.99	×	95.37
Features fromVGG-16, GoogleNet and ResNet50 with SVM classifier [10]	SVM	CT	2	98.28	97.60	98.93	97.63	98.27
Statistical feature extraction techniques with SVM [11]	SVM	CT	2	98.58	99.68	97.56	99.68	97.71

However, for evaluating and benchmarking the ML methods, considering all the aforementioned criteria simultaneously led us to the multi-criteria problem [17]. The multi-criteria problem can be found when the criteria have a trade-off, i.e., between the accuracy and time criteria [18, 19]. And the conflict criteria are another issue in the evaluation process [20, 21]. As a result, multi-criteria decision making is the best scheme for evaluating and benchmarking ML methods.

The method proposed in [17] is a state-of-the-art research work presenting multi-criteria decision making (MCDM) as a solution for evaluating and benchmarking machine learning methods. The authors used existing MCDM methods to accomplish their target. They used the entropy method to extract the objective weight from the decision matrix, while the Technique

for Order of Preference by Similarity to Ideal Solution (TOPSIS) method was exploited for ranking the machine learning methods without making contributions to the MCDM theory or any machine learning method.

However, both the Entropy and TOPSIS methods could affect the final decision as follows: (1) the objective weight does not express the expert's point of view because it extracts the weight from the decision matrix depending on mathematical equations using the Entropy method, and (2) TOPSIS suffers from many drawbacks such as normalization, distance measurement, ideal solutions, and being time-consuming [19, 22-25]. Group decision-making context is the most common configuration used for MCDM, depending on multiple decision-makers' preferences [25-29].

The authors in [17] used the objective weight method (entropy method) for MCDM; however, they have not made a validation for the final rank, which raises questions on the validity of the performance of their presented method. In [22], the authors developed the Fuzzy Decision by Opinion Score Method (FDOSM), which was proven to be efficient and powerful compared to the existing methods. Other weighting MCDM methods and theoretical challenges were solved using the FDSOM technique, demonstrating that the [17] suffers from its [22, 23]. Multi-criteria decision-making approach based on the Fuzzy Decision by Opinion Score Method (FDOSM) has been exploited to address all the aforementioned issues [22, 30, 31]. Therefore, in this research, we proposed FDOSM to evaluate and benchmark COVID-19 ML methods.

2. Methodology

In this section, the proposed framework for evaluating and benchmarking the machine learning methods for classifying COVID-19 based on the FDOSM is presented in detail. Section 2.1 presents the first part of our methodology, in which we describe the decision matrix of machine learning methods. Section 2.2, which explains the second phase of the developed method and describes the steps of the FDOSM used in benchmarking the COVID-19 ML methods, is presented. Figure 1 shows the map of the developed method presented for benchmarking the COVID-19 ML methods based on FDOSM.

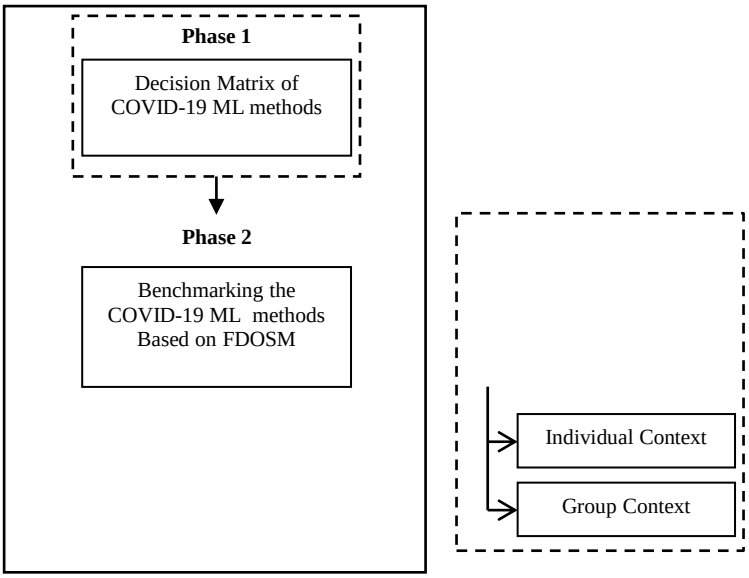

Figure 1: Block diagram of proposed benchmarking methodology for COVID-19 ML methods

2.1 Phase 1: Creating the COVID-19 Machine Learning Methods Decision Matrix

The proposed methodology to benchmark and evaluate the classifiers of COVID-19 comprises two phases: the identification phase and the benchmarking phase. The identification phase (i.e., the first phase) is intended to develop the DM on the basis of the intersection between performance criteria and models of COVID-19 diagnosis. The benchmarking process is the second phase of the proposed methodology, which is dedicated to COVID-19 diagnostic system benchmarking and ranking based on the FDOSM technique.

A. Identification Phase

The main purpose of this stage is to develop DM based on the intersection of multiple evaluation criteria in performance measurements and models. Significant terms, such as criteria, alternatives, and the decision matrix, should be specified in any MCDM case. In our proposed method, these terms are defined as follows:

1- Identifying the Alternatives: The alternatives are the different elements that are targeted to be ranked based on decision-makers, expert opinion, and MCDM techniques. In this study, the developed system uses a chest X-ray (CXR) image to conduct the COVID-19 diagnosis. Based on the literature, eight different ML algorithms, both linear and nonlinear, were frequently applied to diagnose COVID-19. Therefore, as alternatives in the DM, we consider K-Nearest Neighbors (K_NN), Gradient Boosting (GB), Support Vector Machines (SVM), Decision Tree (DT), Logistic Regression (LR), Artificial Neural Network (ANN), Random Forest (RF), and Naive Bayes (NB) as our selected machine learning models. For comparative analysis, we implement these eight models to rank models on the basis of performance.
2- Identifying the Evaluation Criteria: Evaluation criteria are the various measurements that can be used to evaluate and benchmark alternatives. Figure 2 demonstrates the definitive collection of criteria used in this research.

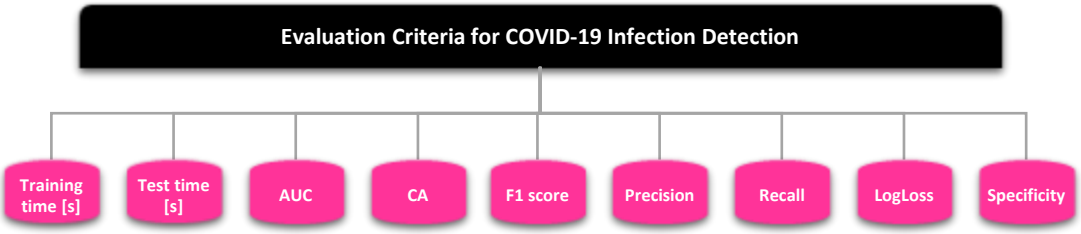

Figure 2: Evaluation Criteria for Diagnosing Systems

		Predictive Label		
Actual Label		COVID-19 (+ve)	COVID-19 (-ve)	
	COVID-19 (+ve)	True Positive (TP)	False Negatives (FN)	**Sensitivity**
	COVID-19 (-ve)	False Positive (FP)	True Negatives (TN)	**Specificity**
		Precision	**Negative Predictive Value**	**Accuracy**

Figure 3: Confusion Matrix Parameters

We utilized the criteria: classification accuracy (CA), F1 score, recall, precision, log loss, specificity, and area under the curve (AUC), which are the most prevalent measures [32, 33].

There are four important confusion matrix parameters used with the mathematical formulation for recall, precision, accuracy, and F1 score (see Figure 3). "True positive" (TP) refers to the number of correctly detected positive samples, "true negative" (TN) refers to the number of correctly detected negative samples, "false positive" (FP) refers to the number of negative samples assorted as positive, and "false negative" (FN) refers to the number of positive specimens predicted as unfavorable."

Classification accuracy (CA) is the most widely used metric for evaluating classification models, and it describes the degree of similarity to the true value. It is the ratio of the correct number of identifications to the total number of input samples and is calculated using the following formula:

$$CA = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

Sensitivity (True Positive Rate): Also known as Recall, this refers to the number of correctly predicted samples of all the positive input samples. It can be interpreted as the capability of a test to correctly distinguish diseased patients, for instance. A highly sensitive test is the most significant indicator, which means that there are few false negative results detected, and thus fewer samples of a certain disease are missed. The formula for the sensitivity is:

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Precision is the number of correctly predicted samples among all the predicted samples. It tests the classifier's ability to reject irrelevant subjects. Precision is calculated using the following formula:

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

The F1 score is described as the weighted average of recall and precision. When the F1 value is equal to one, this represents the best value, while 0 refers to the worst score. Both precision and recall contribute equally to the F1 score. A low F1 score is an indication of both poor recall and poor precision. The formula for the F1 score is:

$$F - score = \frac{2*TP}{2*TP+FP+FN} \quad (4)$$

Specificity (True Negative Rate): The ability of a model to identify the true negatives of each available class for COVID_19 detections, specificity refers to the ability of a test to correctly identify the control subjects. A highly specific test leads to few false-positive cases. The equations for calculating the specificity metric are below:

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

Log loss is a significant classification metric based on prediction probabilities. It can be properly applied to find the probabilities of every output (predict proba) of a classifier rather than its discrete predictions (labels). For a given problem, a lower log-loss value results in better predictions. The equations for calculating the log loss metric are below.

$$LogLoss = -\frac{1}{n}\sum_{i=1}^{n} [y_i \cdot \log_e(\hat{y}_i) + (1 - y_i) \cdot \log_e(1 - \hat{y}_i)] \quad (6)$$

Where n is the number of samples in a given population (examples), $\hat{y}_i$ is the predicted probability per class.

Area Under the Curve (AUC) is a related Receiver Operating Characteristics (ROC) curve that aims to evaluate the performance of the classification model at various threshold

settings. The AUC value reveals how the model performs by distinguishing between classes (i.e., degree of separability). The increase in the value of the AUC indicates better performance. For example, a higher AUC indicates that the model is better at distinguishing between disease (Covid-19) and normal (AUC).

3- Identifying the Decision Matrix: In this section, an intersection is designed between COVID-19 diagnostic alternatives (ML models) and the performance evaluation criteria of the diagnostic systems. Accordingly, the overlap between the eight diagnostic models and the nine evaluation criteria (i.e., CA, F1 score, precision, recall, log loss, specificity, train time, and test time) forms the COVID-19 Diagnostic Decision Matrix. The structure of our proposed decision matrix is illustrated in Table 2. The first column in DM represents the various alternatives and evaluation criteria represented in the top row. In this DM, the other rows represent the model outcome values in relation to the specified evaluation criteria.

Table 2 Structure of a decision matrix

Criteria / Alternative	Training time [s]	Testing time [s]	AUC	CA	F1 score	Precision	Recall	LogLoss	Specificity
$Model_1$									
$Model_2$									
$Model_3$									
$Model_4$									
.									
.									
$Model_n$									

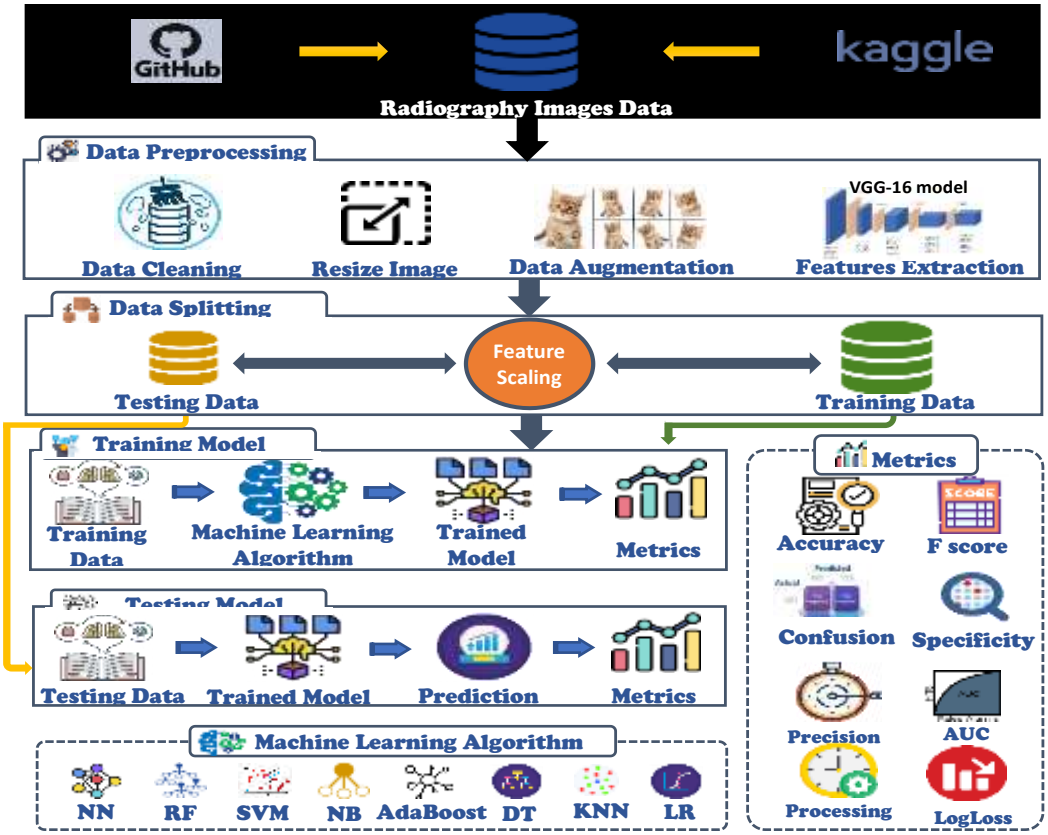

Figure 4: Flowchart for COVID-19 Detection System Modeling and Deployment

Experiment Setup: Figure 4 demonstrates the main processes that were followed to detect COVID-19 and normal cases using different machine learning algorithms.

- ***Dataset organization***

In this study, two datasets that are publicly available were chosen as the main source of chest X-ray (CXR) images. The chosen dataset contains CXR images of COVID-19 patients and healthy patients. From the GitHub repository, the first publicly accessible dataset created by Dr. Joseph Cohan is collected. This dataset includes CXR images of positive COVID-19 patients, Middle East respiratory syndrome (MARS), severe acute respiratory syndrome (SARS), and acute respiratory distress syndrome (ARDS) [34]. We obtained the second dataset, "Chest X-Ray Images (Pneumonia)," from the Kaggle repository, which includes CXR images of patients with pneumonitis and normal [35]. Figure 5 shows CXR image samples of a patient with COVID-19 and a normal person from the collected datasets.

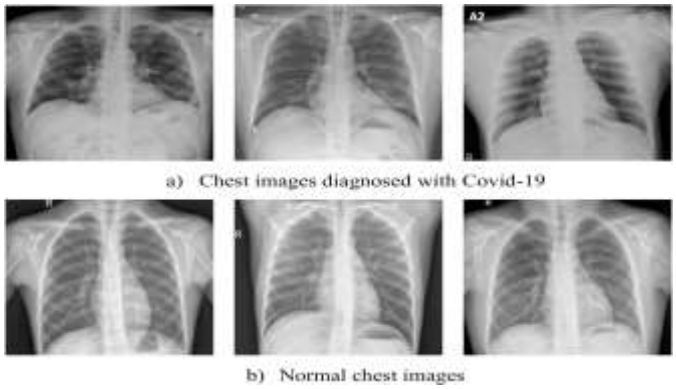

Figure -5 An example of COVID-19, and Normal CXR images

- ***Data Setting***

340 CXR (chest X-ray) images collected from GitHub were captured with a frontal (260 images) or sideways view (80 images). In our experiments, only 260 frontal CXR image samples were considered. For the second dataset, there are 5863 images in the Kaggle chest X-ray dataset categorized into two classes: normal and pneumonia. From the Normal category, we have selected 260 images randomly to construct a balanced dataset. The final set of the dataset comprises 520 images: 260 samples of COVID-19 and 260 normal examples.

- ***Data Pre-processing***

The COVID CXR images collected from the GitHub and Kaggle repositories vary in size from 508 × 500 to 4248 × 3480 pixels. Therefore, we resized the images to 224 × 224 pixels for the experimental setting. In accordance with the model requirement, the "Preprocess_input" function implemented in Keras, not to mention Keras is one of the Python language libraries is used to apply pre-processing and resize and transform the input image. Then, different techniques of data augmentation are added to the training samples so as to enhance the model's efficiency by doubling the size of the data. For this purpose, the Keras API, namely "Image_Data_Generator" is used. Methods, namely "in place" and "on the fly," were used in this experiment for data augmentation, where the images are transformed randomly during training. The main advantage of this approach is that at each epoch, the network considers new images that increase the generalizability of the model.

In the augmentation process, each image is rescaled, rotated to a range of 20 degrees of rotation, zoomed to a range of 20%, and eventually flipped horizontally and vertically. In this study, we utilized the pre-trained model, namely VGG-19, for image feature extraction based

on transfer learning. VGG is a convolutional neural network model consisting of 16 weighted layers introduced by the Visual Geometry Group for image identification. Figure 6 illustrates the architecture of VGG16 [11].
The feature space provided by 13 convolutional layers and 5 max-pooling layers was used. The network is fed with images of size (224 x 224 x 3) as an input to the input layer, and the feature is obtained from the last layer for max-pooling. We do not use any of the fully connected layers. To evaluate the detection system, the dataset is split into two individual subsets, namely, the training subset and the testing subset. The training subset is utilized to train ML algorithms, and the other portion of the dataset (the test subset) is utilized to validate ML algorithms that have been learned. Therefore, 80% of the data is used for training, and 20% of the data is to be randomly tested
.

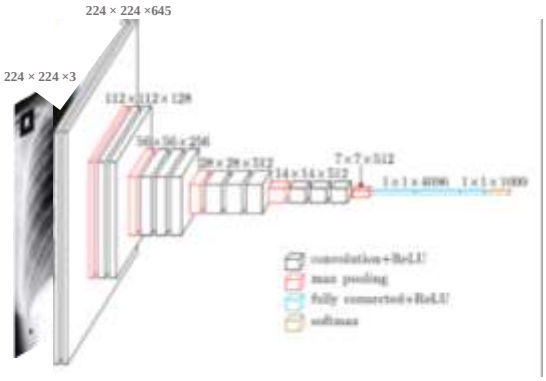

Figure 6: VGG16 Architecture

- ***COVID-19 Detection System Modelling and Evaluation***

The final step in the first phase is developing the COVID-19 detection system. A total of 8 detection systems are developed using the eight most commonly used ML algorithms that were applied in some of the preceding studies and showed adequate results when applied to the diagnosis of the COVID-19 dataset. These algorithms include neural networks, SVM, decision trees, K-NN, logistic regression, random forests, AdaBoost, and Naive Bayes. The developed system is preconfigured with the required machine learning libraries, such as Keras, Scikit-Learn, TensorFlow, Matplotlib, and NumPy. The developed program involves a three-step process.

The first step is related to the proper selection and preprocessing of datasets for machine learning training models. The learning process is implemented by training the classifier on the training dataset in the second step. In the third step, the pre-trained models were evaluated on unseen data known as a test dataset. The outcome of this stage determines the diagnostic efficacy of the model as it succeeds in classifying untrained examples and, thus, the model's feasibility to diagnose future cases. Ultimately, diagnostic models that produce an appropriate outcome can be deployed for diagnosis. The system was implemented and developed on a computer with an Intel Core i7 processor, 4 GB of RAM, and a 2 GB NVIDIA GPU, and it employs the Python programming language.

2.2 Phase 2: FDOSM to Benchmarking ML Methods

As shown in Figure 7, the second phase of the proposed system presents the stages of FDOSM used in the benchmarking evaluation of the COVID-19 ML methods. The first stage (data transformation unit) of FDOSM is described in Section 2.2.1. In Section 2.2.2, the second stage of FDOSM (data processing) is presented.

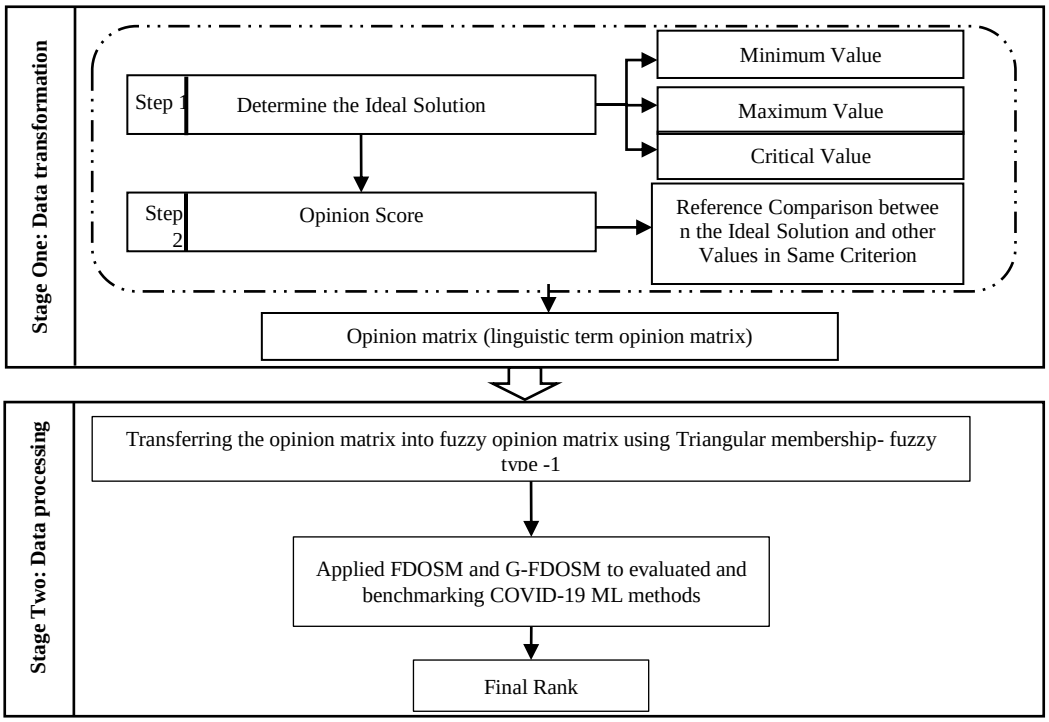

Figure 7: FDOSM Stages

2.2.1 Stage One: Data Transformation Unit

According to [22], the data transformation unit has two main steps to transform the decision matrix into an opinion matrix.

Step 1:

In this step, the selection of the ideal solution for each criterion used in the decision matrix of benchmarking ML methods (rain time [s], test time [s], AUC, CA, F1, Precision, Recall, Log Loss, and Specificity) is achieved using the following equation:

$$A^* = \{[(max_i\, v_{ij} \,|j \in J), (min_i\, v_{ij} \,|j \in J), (Op_{ij} \in I.J)|i = 1.2.3. \ldots . m]\}, \quad (7)$$

The max term refers to the typical value of the benefit ML criteria (AUC, CA, F1, precision, recall, log loss, and specificity), whereas the min term refers to the ideal solution of the cost ML criteria (train time [s] and test time [s]) and $Op_{ij}$is the critical value when the ideal intermediate value lies between the min and max. The decision-maker is responsible for determining this critical value. However, it is not required to set a critical value in the criteria of the utilized evaluation in the benchmarking of ML methods because all the criteria used in the decision matrix are either benefit or cost criteria.

Step 2:

Following the determination of the ideal solution, a reference comparison is conducted by the expert between the ideal solution and other alternative values that meet the same criterion, using five linguistic scale terms. The scales of linguistic terms are categorized as follows: huge difference, big difference, difference, slight difference, and no difference. This step can be represented by the following equation:

$$Op_{Lang} = \left\{\left((\tilde{v}_{ij} \otimes v_{ij} \,|j \in J).\left|i = 1.2.3. \ldots . m\right.\right)\right\} \quad (8)$$

Where $\otimes$ refers to the aforementioned reference comparison.

The outcome of the data transformation unit is the linguistic term "opinion matrix," identified as follows:

$$Op_{Lang} = \begin{matrix} A_1 \\ \vdots \\ A_m \end{matrix} \begin{bmatrix} op_{11} & \cdots & op_{1n} \\ \vdots & \ddots & \vdots \\ op_{m1} & \cdots & op_{mn} \end{bmatrix} \quad (9)$$

Once the opinion matrix is formulated, it is then converted into fuzzy numbers using an appropriate fuzzy membership.

2.2.2 Stage Two: Data-processing Unit

Two main configurations are applied in this stage; the first configuration is benchmarking ML methods based on individual FDOSM. The second configuration is benchmarking ML methods based on Group FDOSM. Both methods are described as follows:

1- Benchmarking ML Methods based on Individual FDOSM

Step 1: After establishing the opinion matrix, the fuzzification process that aims to convert the opinion matrix into a fuzzy opinion decision matrix using triangular fuzzy numbers (TFNs) is carried out [22]. This can be achieved by replacing the opinion terms with triangular fuzzy numbers, which are formulated by their membership function, which is defined as follows:

$$\mu A(x) = \begin{cases} 0 \; if x < a \\ \frac{x-a}{b-a} if a \le x \le b \\ \frac{c-x}{c-b} if b \le x \le x \\ 0 \; if x > c \end{cases}, \quad where \quad a \le b \le c \quad (10)$$

Remark: $\tilde{x} = (a_1, b_1, c_1)$ and $\tilde{y} = (a_2, b_2, c_2)$ are two non-negative TFNs, and $\alpha \in \mathbb{R}_+$. The arithmetic operations are defined according to the extension principle as follows:

1. $\tilde{x} + \tilde{y} = (a_1 + a_2, b_1 + b_2, c_1 + c_2)$,
2. $\tilde{x} - \tilde{y} = (a_1 - c_2, b_1 - b_2, c_1 - a_2)$,
3. $\alpha\tilde{x} = (\alpha a_1, \alpha b_1, \alpha c_1)$,
4. $\tilde{x}^{-1} \cong (1/c_1, 1/b_1, 1/a_1)$,
5. $\tilde{x} \times \tilde{y} \cong (a_1a_2, b_1b_2, c_1c_2)$,
6. $\tilde{x}/\tilde{y} \cong (a_1/c_2, b_1/b_2, c_1/a_2)$.

In Table 3, present the TFNs for each linguistic term.

Table 3: Conversion of opinion linguistic terms into triangular fuzzy numbers (TFNs)

Linguistic terms	TFNs
No difference	(0.00, 0.10, 0.30)
Slight difference	(0.10, 0.30, 0.50)
Difference	(0.30, 0.50, 0.75)
Big difference	(0.50, 0.75, 0.90)
Huge difference	(0.75, 0.90, 1.00)

Finally, the fuzzy decision matrix used in the evaluation and benchmarking of the ML methods was applied to the COVID-19 diagnosis task. We applied two contexts, including individual and group decision making on a fuzzy opinion matrix, to evaluate and benchmark the ML methods on the COVID-19 dataset.

Step 2: According to [22], direct aggregation is utilized on the fuzzy opinion decision matrix utilizing an aggregation operator (i.e., the arithmetic mean). The execution of the aggregation process is done by using the following equation to benchmark the ML methods:

$$Arithmetic\ mean\ A_{m(x)=}\frac{\sum_{i=1}^{n} x_i}{n}, \tag{11}$$

$$A_{m(x)=}\frac{\sum(a_f+a_m+a_l)(b_f+b_m+b_l)(c_f+c_m+c_l)}{n} \tag{12}$$

Step 3: The centroid method was used for the defuzzification process. The ranking of the ML method will be produced after the defuzzification process. The best option of the ML method is the one with the least value. The defuzzification process is applied using the following equation:

$$Diff = \frac{(a+b+c)}{3}. \tag{13}$$

2. Benchmarking ML methods based on Group Decision-making Context

The different aggregated decisions that were obtained from various assessors are essential to unifying the benchmarking output. This is due to the variance in the benchmarking COVID-19 ML methods among decision makers; thus, we consider the group decision-making context to incorporate all benchmarking by the decision makers to achieve the final benchmarking COVID-19 ML methods. The arithmetic mean is utilized so as to reach the final score of group decision making, where the lowest score value represents the best substitution. It should be noted that experts' opinions are integrated after the final ranking.

$$Group - FDOSM = \oplus R^* \tag{13}$$

$\oplus =$ Arithmetic mean.

$R^* =$ The Final result for each expert.

3. Result and Discussion

The results of the two main contexts (individual and group decision making) are reported as follows:

3.1 The Opinion Matrix and Fuzzy Opinion Matrix

The opinion matrix and fuzzy opinion matrix used in the evaluation and benchmarking of the ML techniques based on the COIVD-19 dataset are reported in this section. This process is realized by converting the original decision matrix presented in Table 2 into the opinion decision matrix depicted in Table 4 and judging the three decision-makers' preferences using the five Likert scales. The ideal solution was determined by the decision-maker as defined in Equation 7. Therefore, to establish the opinion matrix of the decision-maker, reference comparisons are conducted between the optimal solution and other values of alternatives under the same criteria. Table 4 presents the opinion decision matrix derived from the first decision maker's preference. In Table A of the Appendix, the other opinion matrices of other decision-makers are presented.

Table 4: The Opinion Decision Matrix of the First Decision Maker

Alternatives	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	H.D	B.D	NO.D	NO.D	NO.D	NO.D	NO.D	S.D	NO.D
SVM	B.D	H.D	NO.D	S.D	S.D	S.D	S.D	NO.D	S.D
Logistic Regression	S.D	B.D	S.D	S.D	DI	S.D	DI	S.D	DI
kNN	NO.D	H.D	S.D	DI	B.D	DI	B.D	DI	DI
Random Forest	DI	B.D	S.D	DI	B.D	DI	B.D	S.D	DI
Naive Bayes	S.D	B.D	DI	B.D	H.D	B.D	H.D	H.D	B.D
Tree	DI	NO.D	H.D	B.D	H.D	B.D	H.D	B.D	B.D
AdaBoost	DI	DI	H.D	H.D	H.D	H.D	H.D	H.D	H.D

*** NO.D: No Difference / S.D: Slight Difference / DI: Difference / B.D: Big Difference / H.D: Huge Difference**

After that, by replacing the linguistic terms with fuzzy numbers using the fuzzy membership of TFNs described in Equation 10, the opinion decision matrix is transformed into a fuzzy opinion decision matrix. (Refer to Table 3). Table 5 introduces the fuzzy opinion decision matrix of the first decision maker. In Table B in the Appendix, other fuzzy opinion matrices of the other decision-makers are listed.

Table 5: Fuzzy Opinion Decision Matrix of the First Decision Maker

Alternatives / Criteria	Neural Network	SVM	Logistic Regression	kNN	Random Forest	Naive Bayes	Tree	AdaBoost
Training time [s]	0.75	0.5	0.1	0	0.3	0.1	0.3	0.3
	0.9	0.75	0.3	0.1	0.5	0.3	0.5	0.5
	1	0.9	0.5	0.3	0.75	0.5	0.75	0.75
Testing time [s]	0.5	0.75	0.5	0.75	0.5	0.5	0	0.3
	0.75	0.9	0.75	0.9	0.75	0.75	0.1	0.5
	0.9	1	0.9	1	0.9	0.9	0.3	0.75
AUC	0	0	0.1	0.1	0.1	0.3	0.75	0.75
	0.1	0.1	0.3	0.3	0.3	0.5	0.9	0.9
	0.3	0.3	0.5	0.5	0.5	0.75	1	1
CA	0	0.1	0.1	0.3	0.3	0.5	0.5	0.75
	0.1	0.3	0.3	0.5	0.5	0.75	0.75	0.9
	0.3	0.5	0.5	0.75	0.75	0.9	0.9	1
F1	0	0.1	0.3	0.5	0.5	0.75	0.75	0.75
	0.1	0.3	0.5	0.75	0.75	0.9	0.9	0.9
	0.3	0.5	0.75	0.9	0.9	1	1	1
Precision	0	0.1	0.1	0.3	0.3	0.5	0.5	0.75
	0.1	0.3	0.3	0.5	0.5	0.75	0.75	0.9
	0.3	0.5	0.5	0.75	0.75	0.9	0.9	1
Recall	0	0.1	0.3	0.5	0.5	0.75	0.75	0.75
	0.1	0.3	0.5	0.75	0.75	0.9	0.9	0.9
	0.3	0.5	0.75	0.9	0.9	1	1	1
LogLoss	0.1	0	0.1	0.3	0.1	0.75	0.5	0.75
	0.3	0.1	0.3	0.5	0.3	0.9	0.75	0.9
	0.5	0.3	0.5	0.75	0.5	1	0.9	1
Specificity	0	0.1	0.3	0.3	0.3	0.5	0.5	0.75
	0.1	0.3	0.5	0.5	0.5	0.75	0.75	0.9
	0.3	0.5	0.75	0.75	0.75	0.9	0.9	1

In order to achieve the benchmarking ML techniques, three decision-making approaches were subsequently applied to the outcomes of fuzzy opinion matrices. The next sections clarify the outcome of each strategy.

3.2 Benchmarking Results according to the Individual FDOSM Context

This section shows the benchmarking results of the ML approaches using the individual FDOSM contexts of the three decision makers (see Table 6).

Table 6: Results of the Individual Decision-Making Context Used in Benchmarking the ML Methods

Alternatives	Decision Maker 1		Decision Maker 2		Decision Maker 3	
	Score	Rank	Score	Rank	Score	Rank
Neural Network	0.3	1	0.3	1	0.324074074	1
SVM	0.374074074	2	0.355555556	3	0.392592593	3
Logistic Regression	0.418518519	3	0.327777778	2	0.372222222	2
kNN	0.535185185	4	0.424074074	4	0.466666667	4
Random Forest	0.535185185	4	0.466666667	5	0.490740741	5
Naive Bayes	0.703703704	7	0.538888889	6	0.535185185	6
Tree	0.685185185	6	0.561111111	7	0.644444444	7
AdaBoost	0.801851852	8	0.7	8	0.787037037	8

The benchmarking results demonstrate the importance of the decision maker's opinion for each criterion in terms of the benchmark. As described in the previous section, the alternative that has the lowest score is the best, while the alternative that has the highest score is the least preferable option. Table 6 presents the results of FDOSM according to the decision makers' opinions. Three decision makers gave their thoughts on whether to use FDOSM and then produce the final results of benchmarking using the ML method. The results show that neural networks are the best ML model applied to the COVID-19 dataset for the three decision-makers, with scores of 0.3, 0.3, and 0.324074074, respectively.

Further, the worst alternative was AdaBoost for all decision makers, achieving scores of 0.801851852, 0.7, and 0.787037037, respectively. The variance in rank happened on the second and third alternatives. The rank was affected due to the opinion that was provided by the decision-makers. Comparing the final result of our method with the final results of [16], the variance in the rank of ML methods that were applied to the COVID-19 dataset is clearly noticeable. This variation is due to the preferences of the decision-makers. So, the final results of our proposed system are closer to decision makers' preferences than the final result of FDOSM with the opinion matrix of the decision maker. Due to the variation in the final rank for COIVD-19 ML methods, we also considered the group decision-making contexts to evaluate and benchmark COVID-19 ML methods.

3.3 Benchmarking Results according to Group Decision Making Context

As mentioned in the previous sections, group decision making is the most important configuration, which is widely used in the literature. In this section, the outcomes of group decision making (GDM) are presented. According to Equation 14, the final outcomes of the three decision-makers are combined using the "arithmetic mean operator" to report the final GDM raking for benchmarking COVID-19 ML methods. The conclusive outcome of GDM is shown in Table 7.

Table 7: Group FDOSM Context

Alternatives	Score	Rank
Neural Network	0.308025	1
SVM	0.374074	3
Logistic Regression	0.37284	2
kNN	0.475309	4
Random Forest	0.497531	5
Naive Bayes	0.592593	6
Tree	0.630247	7
AdaBoost	0.762963	8

In Table 7, it can be clearly seen that the best ML method for COVID-19 diagnosis is the neural network, which achieved the top outcome and the lowest score value of 0.308025. The worst technique, however, is AdaBoost, which has achieved the lowest rank and the highest score value of 0.762963. The rank of COVID-19 ML methods is in line when comparing the group decision-making result with the opinions of the three decision-makers. Therefore, the rank of the group decision-making context can be counted as the conclusive ranking outcome that can be used as the basis of the objective validation processes. In the next section, the objective validation results are described in detail.

4. Results Objective Validation

In this study, objective validation is used to demonstrate the COVID-19 ML methods for benchmarking group decision-making outcomes obtained by the FDOSM. The objective validation process is introduced by dividing the benchmarking COVID-19 ML methods into equal groups. This process is described in several MCDM studies [21, 25, 36, 37]. The number of COVID-19 ML methods within each group and the number of groups do not affect the objective validation output [38-40]. To validate the group benchmarking COVID-19 ML method results, various steps should be performed as follows: (1) The COVID-19 ML methods are sorted according to GDM results; (2) following sorting, the COVID-19 ML methods are separated into two equal groups; and (3) the mean ($\bar{x}$) for each group in the GDM result is calculated afterwards as defined in Equation 15.

$$\bar{\boldsymbol{x}} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{x_i} \tag{15}$$

The process of comparison is achieved on the basis of the findings of each group's mean. The method of the comparison is based on the average outcome in each and every group. The minimum values of the means of each group contribute to relevant outcomes since the lowest linguistic terms are assigned to the optimal solution of each criterion by the decision makers, which is the concept of FDOSM. The first group is thus considered to have the minimum mean for testing the validity of the result and is therefore compared with the second group. The mean outcome of the second group must be greater than or equal to that of the first group. If the findings of the evaluation are consistent with the assumption, then the outcomes are correct. In Table 8, the results of objective validation for benchmarking COVID-19 ML methods based on FDOSM are presented. For the first group, the obtained mean is 0.382562, which is lower than the mean of the second group with a value of 0.620833. This shows that the findings of benchmarking COVID-19 ML methods based on FDOSM are valid, closest to decision makers' opinions, logical, and have undergone systematic ranking.

Table 8: The Objective Validation of Group Benchmarking Results of COVID-19 ML Methods

Group	COVID-19 ML methods	Mean
1st Group	Neural Network	0.382562
	Logistic Regression	
	SVM	
	kNN	
2nd Group	Random Forest	0.620833
	Naive Bayes	
	Tree	
	AdaBoost	

As shown in Table 8, based on the effectiveness of the outcome of the group of benchmarking COVID-19 ML techniques obtained by the FDOSM, the mean of the first group (i.e., 0.382562) is lower than that of the second group (i.e., 0.620833). As a result, the group FDOSM results for the benchmarking COVID-19 ML methods are valid and underwent systematic ranking.

5. Conclusion

This research achieved the evaluation and benchmarking of the COVID-19 machine learning methods based on the new MCDM method, namely FDOSM. The methodology of this research consists of two stages, as shown in Figure 1. The first phase is regarding the construction of the COVID-19 machine learning methods decision matrix, whereas the second phase is regarding the FDOSM standards and procedures. The main contributions resulting from this research are creating a new decision matrix by applying eight machine learning methods as a set of alternatives and extracting nine evaluation criteria, along with applying the latest MCDM method (i.e., FDOSM) to evaluate and benchmark COVID-19 machine learning methods.

The objective validation of the final ranking results is applied by using a statistical method (i.e., mean), which achieved the benchmarking COVID-19 machine learning method results. Finally, this research presents many benefits, especially for hospitals and medical clinics, in order to speed up the diagnosis of patients suffering from COVID-19 by utilizing the best machine learning method. The following are the recommended future research directions:(1) Extending FDOSM into a z-number of environments (2) Alternative membership functions can be applied, such as intuitionistic interval-valued fuzzy numbers and intuitionistic trapezoidal fuzzy numbers, and their results compared. (3) Defuzzification approaches have the potential to achieve alternative rankings. (4) Last but not least, we intend to extend the aforementioned proposition with different environments (i.e., rough and gray) to be set as one of the main future directions.

References

[1] T. Mahmud, M. A. Rahman, and S. A. Fattah, "CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization," *Computers in biology and medicine*, vol. 122, p. 103869, 2020.

[2] T. Tuncer, S. Dogan, and F. Ozyurt, "An automated Residual Exemplar Local Binary Pattern and iterative ReliefF based COVID-19 detection method using chest X-ray image," *Chemometrics and Intelligent Laboratory Systems*, vol. 203, p. 104054, 2020.

[3] M. Nour, Z. Cömert, and K. Polat, "A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization," *Applied Soft Computing*, vol. 97, p. 106580, 2020.

[4] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons & Fractals,* vol. 140, p. 110120, 2020.

[5] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. Abo-Elsoud, "A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier," *Knowledge-Based Systems,* vol. 205, p. 106270, 2020.

[6] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Informatics in medicine unlocked,* vol. 19, p. 100360, 2020.

[7] S. Garfan, A. Alamoodi, B. Zaidan, M. Al-Zobbi, R. A. Hamid, J. K. Alwan, *et al.*, "Telehealth utilization during the Covid-19 pandemic: A systematic review," *Computers in biology and medicine,* p. 104878, 2021.

[8] H. Panwar, P. Gupta, M. K. Siddiqui, R. Morales-Menendez, and V. Singh, "Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet," *Chaos, Solitons & Fractals,* vol. 138, p. 109944, 2020.

[9] F. Ucar and D. Korkmaz, "COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images," *Medical hypotheses,* vol. 140, p. 109761, 2020.

[10] M. Ahmed, Z. Al-qaysi, M. L. Shuwandy, M. M. Salih, and M. H. Ali, "Automatic COVID-19 pneumonia diagnosis from x-ray lung image: A Deep Feature and Machine Learning Solution," in *Journal of Physics: Conference Series,* 2021, p. 012099.

[11] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial Intelligence (AI) applications for COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews,* vol. 14, pp. 337-339, 2020.

[12] A. Kumar, P. K. Gupta, and A. Srivastava, "A review of modern technologies for tackling COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews,* vol. 14, pp. 569-573, 2020.

[13] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and engineering sciences in medicine,* vol. 43, pp. 635-640, 2020.

[14] N. Hasan, "A methodological approach for predicting COVID-19 epidemic using EEMD-ANN hybrid model," *Internet of Things,* vol. 11, p. 100228, 2020.

[15] Y. Mohamadou, A. Halidou, and P. T. Kapen, "A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19," *Applied Intelligence,* vol. 50, pp. 3913-3925, 2020.

[16] M. A. Mohammed, K. H. Abdulkareem, A. S. Al-Waisy, S. A. Mostafa, S. Al-Fahdawi, A. M. Dinar, *et al.*, "Benchmarking Methodology for Selection of Optimal COVID-19 Diagnostic Model Based on Entropy and TOPSIS Methods," *IEEE Access,* 2020.

[17] O. Zughoul, A. Zaidan, B. Zaidan, O. Albahri, M. Alazab, U. Amomeni, *et al.*, "Novel triplex procedure for ranking the ability of software engineering students based on two levels of AHP and group TOPSIS techniques," *International Journal of Information Technology and Decision Making,* 2021.

[18] M. Alaa, I. S. M. A. Albakri, C. K. S. Singh, H. Hammed, A. Zaidan, B. Zaidan, *et al.*, "Assessment and ranking framework for the English skills of pre-service teachers based on fuzzy Delphi and TOPSIS methods," *IEEE Access,* vol. 7, pp. 126201-126223, 2019.

[19] O. Albahri, A. Zaidan, A. Albahri, H. Alsattar, R. Mohammed, U. Aickelin, *et al.*, "Novel dynamic fuzzy decision-making framework for COVID-19 vaccine dose recipients," *Journal of Advanced Research,* 2021.

[20] R. Malik, A. Zaidan, B. Zaidan, K. Ramli, O. Albahri, Z. Kareem, *et al.*, "Novel Roadside Unit Positioning Framework in the Context of the Vehicle-to-Infrastructure Communication System Based on AHP—Entropy for Weighting and Borda—VIKOR for Uniform Ranking," *International Journal of Information Technology & Decision Making,* pp. 1-34, 2021.

[21] O. S. Albahri, A. A. Zaidan, M. M. Salih, B. B. Zaidan, M. A. Khatari, M. A. Ahmed, *et al.*, "Multidimensional benchmarking of the active queue management methods of network congestion control based on extension of fuzzy decision by opinion score method," *International Journal of Intelligent Systems,* vol. n/a.

[22] M. M. Salih, B. Zaidan, and A. Zaidan, "Fuzzy decision by opinion score method," *Applied Soft Computing*, p. 106595, 2020.
[23] M. M. Salih, B. Zaidan, A. Zaidan, and M. A. Ahmed, "Survey on fuzzy TOPSIS state-of-the-art between 2007 and 2017," *Computers & Operations Research*, vol. 104, pp. 207-227, 2019.
[24] K. H. Abdulkareem, N. Arbaiy, A. Zaidan, B. Zaidan, O. S. Albahri, M. Alsalem, *et al.*, "A new standardisation and selection framework for real-time image dehazing algorithms from multi-foggy scenes based on fuzzy Delphi and hybrid multi-criteria decision analysis methods," *Neural Computing and Applications*, vol. 33, pp. 1029-1054, 2021.
[25] K. H. Abdulkareem, N. Arbaiy, A. Zaidan, B. Zaidan, O. Albahri, M. Alsalem, *et al.*, "A Novel Multi-Perspective Benchmarking Framework for Selecting Image Dehazing Intelligent Algorithms Based on BWM and Group VIKOR Techniques," *International Journal of Information Technology & Decision Making*, pp. 1-49, 2020.
[26] R. Bai, F. Li, and J. Yang, "A dynamic fuzzy multi-attribute group decision making method for supplier evaluation and selection," in *Control and Decision Conference (2014 CCDC), The 26th Chinese*, 2014, pp. 3249-3256.
[27] A. Hatami-Marbini and F. Kangi, "An extension of fuzzy TOPSIS for a group decision making with an application to tehran stock exchange," *Applied Soft Computing*, vol. 52, pp. 1084-1097, 2017.
[28] O. Albahri, J. R. Al-Obaidi, A. Zaidan, A. Albahri, B. Zaidan, M. M. Salih, *et al.*, "Helping doctors hasten COVID-19 treatment: Towards a rescue framework for the transfusion of best convalescent plasma to the most critical patients based on biological requirements via ml and novel MCDM methods," *Computer methods and programs in biomedicine*, vol. 196, p. 105617, 2020.
[29] M. M. Salih, O. Albahri, A. Zaidan, B. Zaidan, F. Jumaah, and A. Albahri, "Benchmarking of AQM methods of network congestion control based on extension of interval type-2 trapezoidal fuzzy decision by opinion score method," *Telecommunication Systems*, pp. 1-30, 2021.
[30] R. M. Maher, M. M. Salih, H. A. Hussein, and M. A. Ahmed, "A New Development of FDOSM Based on a 2-Tuple Fuzzy Environment: Evaluation and Benchmark of Network Protocols as a Case Study," *Computers*, vol. 11, p. 109, 2022.
[31] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, "Beyond classification: structured regression for robust cell detection using convolutional neural network," in *International conference on medical image computing and computer-assisted intervention*, 2015, pp. 358-365.
[32] P. Sukumar and R. Gnanamurthy, "Computer aided detection of cervical cancer using pap smear images based on adaptive neuro fuzzy inference system classifier," *Journal of Medical Imaging and Health Informatics*, vol. 6, pp. 312-319, 2016.
[33] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988*, 2020.
[34] P. Mooney, "Chest X-Ray Images (Pneumonia)," vol. 2020, ed, 2018.
[35] N. Kalid, A. Zaidan, B. Zaidan, O. H. Salman, M. Hashim, O. S. Albahri, *et al.*, "Based on real time remote health monitoring systems: a new approach for prioritization "large scales data" patients with chronic heart diseases using body sensors and communication technology," *Journal of medical systems*, vol. 42, pp. 1-37, 2018.
[36] N. Kalid, A. Zaidan, B. Zaidan, O. H. Salman, M. Hashim, O. Albahri, *et al.*, "Based on real time remote health monitoring systems: a new approach for prioritization "large scales data" patients with chronic heart diseases using body sensors and communication technology," vol. 42, p. 69, 2018.
[37] M. Qader, B. Zaidan, A. Zaidan, S. Ali, M. Kamaluddin, and W. Radzi, "A methodology for football players selection problem based on multi-measurements criteria analysis," *Measurement*, vol. 111, pp. 38-50, 2017.
[38] K. Mohammed, A. Zaidan, B. Zaidan, O. Albahri, A. Albahri, M. Alsalem, *et al.*, "Novel technique for reorganisation of opinion order to interval levels for solving several instances representing prioritisation in patients with multiple chronic diseases," vol. 185, p. 105151, 2020.
[39] K. Mohammed, J. Jaafar, A. Zaidan, O. Albahri, B. Zaidan, K. H. Abdulkareem, *et al.*, "A Uniform Intelligent Prioritisation for Solving Diverse and Big Data Generated From Multiple

Chronic Diseases Patients Based on Hybrid Decision-Making and Voting Method," vol. 8, pp. 91521-91530, 2020.

[40] K. H. Abdulkareem, "A Novel Multi-Perspective Benchmarking Framework for Selecting Image Dehazing Intelligent Algorithms Based on BWM and Group VIKOR Techniques," *International Journal of Information Technology & Decision Making*, 2020.

Appendix

Table A: The opinion matrix of the other decision makers

The opinion matrix of the second decision maker									
Alternatives	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	H.D	B.D	NO.D	NO.D	NO.D	NO.D	NO.D	S.D	NO.D
SVM	B.D	H.D	NO.D	S.D	S.D	S.D	S.D	NO.D	NO.D
Logistic Regression	S.D	B.D	NO.D	S.D	S.D	S.D	S.D	S.D	S.D
kNN	NO.D	H.D	NO.D	DI	DI	S.D	DI	S.D	DI
Random Forest	DI	B.D	S.D	DI	DI	S.D	DI	S.D	DI
Naive Bayes	NO.D	B.D	S.D	DI	DI	DI	B.D	B.D	B.D
Tree	S.D	NO.D	DI	B.D	B.D	DI	B.D	B.D	B.D
AdaBoost	S.D	B.D	DI	H.D	H.D	DI	H.D	H.D	B.D
The opinion matrix of the third decision maker									
Alternatives	Train time [s]	Test time [s]	AUC	CA	F1	Precision	Recall	LogLoss	Specificity
Neural Network	H.D	B.D	NO.D	NO.D	NO.D	NO.D	NO.D	DI	NO.D
SVM	H.D	H.D	S.D	S.D	S.D	S.D	S.D	NO.D	NO.D
Logistic Regression	DI	DI	S.D	S.D	S.D	S.D	S.D	DI	S.D
kNN	NO.D	H.D	S.D	DI	DI	DI	DI	DI	S.D
Random Forest	B.D	DI	S.D	DI	DI	DI	DI	DI	S.D
Naive Bayes	NO.D	DI	S.D	B.D	DI	DI	B.D	H.D	DI
Tree	B.D	NO.D	DI	H.D	H.D	B.D	B.D	B.D	DI
AdaBoost	B.D	DI	H.D	H.D	H.D	B.D	H.D	H.D	B.D

Table B: The fuzzy opinion matrix of the other decision makers

The fuzzy opinion matrix of the second decision maker								
Alternatives **Criteria**	Neural Network	SVM	Logistic Regression	kNN	Random Forest	Naive Bayes	Tree	AdaBoost
Training time [s]	0.75	0.50	0.10	0.00	0.30	0.00	0.10	0.10
	0.90	0.75	0.30	0.10	0.50	0.10	0.30	0.30
	1.00	0.90	0.50	0.30	0.75	0.30	0.50	0.50
Testing time [s]	0.50	0.75	0.50	0.75	0.50	0.50	0.00	0.50
	0.75	0.90	0.75	0.90	0.75	0.75	0.10	0.75
	0.90	1.00	0.90	1.00	0.90	0.90	0.30	0.90

	0.00	0.00	0.00	0.00	0.10	0.10	0.30	0.30
AUC	0.10	0.10	0.10	0.10	0.30	0.30	0.50	0.50
	0.30	0.30	0.30	0.30	0.50	0.50	0.75	0.75
	0.00	0.10	0.10	0.30	0.30	0.30	0.50	0.75
CA	0.10	0.30	0.30	0.50	0.50	0.50	0.75	0.90
	0.30	0.50	0.50	0.75	0.75	0.75	0.90	1.00
	0.00	0.10	0.10	0.30	0.30	0.30	0.50	0.75
F1	0.10	0.30	0.30	0.50	0.50	0.50	0.75	0.90
	0.30	0.50	0.50	0.75	0.75	0.75	0.90	1.00
	0.00	0.10	0.10	0.10	0.10	0.30	0.30	0.30
Precision	0.10	0.30	0.30	0.30	0.30	0.50	0.50	0.50
	0.30	0.50	0.50	0.50	0.50	0.75	0.75	0.75
	0.00	0.10	0.10	0.30	0.30	0.50	0.50	0.75
Recall	0.10	0.30	0.30	0.50	0.50	0.75	0.75	0.90
	0.30	0.50	0.50	0.75	0.75	0.90	0.90	1.00
	0.10	0.00	0.10	0.10	0.10	0.50	0.50	0.75
LogLoss	0.30	0.10	0.30	0.30	0.30	0.75	0.75	0.90
	0.50	0.30	0.50	0.50	0.50	0.90	0.90	1.00
	0.00	0.00	0.10	0.30	0.30	0.50	0.50	0.50
Specificity	0.10	0.10	0.30	0.50	0.50	0.75	0.75	0.75
	0.30	0.30	0.50	0.75	0.75	0.90	0.90	0.90
The fuzzy opinion matrix of the third decision maker								
Alternatives / **Criteria**	Neural Network	SVM	Logistic Regression	kNN	Random Forest	Naive Bayes	Tree	AdaBoost
	0.75	0.75	0.30	0.00	0.50	0.00	0.50	0.50
Training time [s]	0.90	0.90	0.50	0.10	0.75	0.10	0.75	0.75
	1.00	1.00	0.75	0.30	0.90	0.30	0.90	0.90
	0.50	0.75	0.30	0.75	0.30	0.30	0.00	0.30
Testing time [s]	0.75	0.90	0.50	0.90	0.50	0.50	0.10	0.50
	0.90	1.00	0.75	1.00	0.75	0.75	0.30	0.75
AUC	0.00	0.10	0.10	0.10	0.10	0.10	0.30	0.75

CA	0.10	0.30	0.30	0.30	0.30	0.30	0.50	0.90
	0.30	0.50	0.50	0.50	0.50	0.50	0.75	1.00
	0.00	0.10	0.10	0.30	0.30	0.50	0.75	0.75
	0.10	0.30	0.30	0.50	0.50	0.75	0.90	0.90
	0.30	0.50	0.50	0.75	0.75	0.90	1.00	1.00
F1	0.00	0.10	0.10	0.30	0.30	0.30	0.75	0.75
	0.10	0.30	0.30	0.50	0.50	0.50	0.90	0.90
	0.30	0.50	0.50	0.75	0.75	0.75	1.00	1.00
Precision	0.00	0.10	0.10	0.30	0.30	0.30	0.50	0.50
	0.10	0.30	0.30	0.50	0.50	0.50	0.75	0.75
	0.30	0.50	0.50	0.75	0.75	0.75	0.90	0.90
Recall	0.00	0.10	0.10	0.30	0.30	0.50	0.50	0.75
	0.10	0.30	0.30	0.50	0.50	0.75	0.75	0.90
	0.30	0.50	0.50	0.75	0.75	0.90	0.90	1.00
LogLoss	0.30	0.00	0.30	0.30	0.30	0.75	0.50	0.75
	0.50	0.10	0.50	0.50	0.50	0.90	0.75	0.90
	0.75	0.30	0.75	0.75	0.75	1.00	0.90	1.00
Specificity	0.00	0.00	0.10	0.10	0.10	0.30	0.30	0.50
	0.10	0.10	0.30	0.30	0.30	0.50	0.50	0.75
	0.30	0.30	0.50	0.50	0.50	0.75	0.75	0.90