This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2995597, IEEE Access

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Benchmarking Methodology for Selection of Optimal COVID-19 Diagnostic Model Based on Entropy and TOPSIS Methods

Mazin Abed Mohammed[1]*, Karrar Hameed Abdulkareem[2], Alaa S. Al-Waisy[1], Salama A. Mostafa[3], Shumoos Al-Fahdawi[1], Ahmed M. Dinar[4], Wajdi Alhakami[5], Abdullah Baz[6], Mohammed Nasser Al-Mhiqani[7], Hosam Alhakami[5], Nureize Arbaiy[3], Mashael S. Maashi [8], Ammar Awad Mutlag[9], Begonya Garcia-Zapirain[10], and Isabel de la Torre Díez [11]

[1]College of Computer Science and Information Technology, University of Anbar, Anbar, 31001, Iraq
[2]College of Agriculture, Al-Muthanna University, Samawah 66001, Iraq
[3]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor, 86400, Malaysia
[4]Computer Engineering/University of Technology/Baghdad/Iraq
[5] Department of Information Technology, College of Computers and Information Technology, Taif University, 21944, Taif, Saudi Arabia
[6]Department of Computer Engineering, College of Computer and Information Systems, Umm Al-Qura University, 21955, Makkah, Saudi Arabia
[7]Information Security and Networking Research Group (InFORSNET), Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia
[8]Software Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia
[9] Ministry of education/general directorate of curricula, pure science department, Baghdad, Iraq
[10] eVIDA Lab. University of Deusto. Avda/Universidades 24. (48007) Bilbao. Spain.
[11] Department of Signal Theory and Communications, University of Valladolid, 47011 Valladolid, Spain;

Corresponding author: Mazin Abed Mohammed (mazinalshujeary@uoanbar.edu.iq).

ABSTRACT Nowadays, coronavirus (COVID-19) is getting international attention due it considered as a life-threatened epidemic disease that hard to control the spread of infection around the world. Machine learning (ML) is one of intelligent technique that able to automatically predict the event with reasonable accuracy based on the experience and learning process. In the meantime, a rapid number of ML models have been proposed for predicate the cases of COVID-19. Thus, there is need for an evaluation and benchmarking of COVID-19 ML models which considered the main challenge of this study. Furthermore, there is no single study have addressed the problem of evaluation and benchmarking of COVID diagnosis models. However, this study proposed an intelligent methodology is to help the health organisations in the selection COVID-19 diagnosis system. The benchmarking and evaluation of diagnostic models for COVID-19 is not a trivial process. There are multiple criteria requires to evaluate and some of the criteria are conflicting with each other. Our study is formulated as a decision matrix (DM) that embedded mix of ten evaluation criteria and twelve diagnostic models for COVID-19. The multi-criteria decision-making (MCDM) method is employed to evaluate and benchmarking the different diagnostic models for COVID19 with respect to the evaluation criteria. An integrated MCDM method are proposed where TOPSIS applied for the benchmarking and ranking purpose while Entropy used to calculate the weights of criteria. The study results revealed that the benchmarking and selection problems associated with COVID19 diagnosis models can be effectively solved using the integration of Entropy and TOPSIS. The SVM (linear) classifier is selected as the best diagnosis model for COVID19 with the closeness coefficient value of 0.9899 for our case study data. Furthermore, the proposed methodology has solved the significant variance for each criterion in terms of ideal best and worst best value, beside issue when specific diagnosis models have same ideal best value.

INDEX TERMS COVID19 Diagnostic, Machine learning, Benchmarking Methodology, Chest X-rays Images, Entropy, TOPSIS, Multi-criteria decision-making.

I. INTRODUCTION

Bioinformatics for the medical field is interdisciplinary sciences of computer science, information science and medicine. The Bioinformatics lays on various techniques, computer-aided tools, equipment that used to analyse and retrieve health data to process and produce meaningful information for the health care field [1]. Medical professionals facing a big challenge when making decisions of using the computer-decision support system [6,11] to provide a reliable solution which helps them investigate a particular sophisticated disease such as COVID-19 [2]. In the meantime, the COVID-19 pandemic is a life-threatening disease, its effect leads to death and it threatens the global health around the world. One of the critical investigation methods to detect the COVID-19 is chest radiography imaging that screens the chest of infected patients. Initial studies found that the chest radiography images can show the abnormalities of the chest of the people infected with COVID-19. Some current studies show that the artificial intelligence (AI) models such as traditional Machine Learning (ML) and Deep Learning (DL) methods have a high capability to detect the COVID-19 with a reasonable accuracy prediction using the chest radiography images of the infected patients with COVID-19. Unfortunately, to the best of the authors' knowledge, these ML approaches can be non-open sourced and not publicity available which prevent the research to access and investigate them for further research. Selection of an efficient automated tool to produces reliable solutions lays on how is it widely used [3].

AI methods able to produces rapid reliable solutions of COVID-19 with high diagnostic accuracy. Nowadays, many hospitals and medical research centres use a computer-aided system based on AI to automate the COVID-19 diagnosis instead of analysing the data manually. With a big number of the existing automated AI methods, selecting an appropriate method that produces an efficient, fast solution with error-free is a critical task [4]. As there is no ML classifiers for COVID-19 diagnose is superior [5, 60]. This is put the medical managers in a big challenge to find and evaluate different ML classifiers for COVID-19 diagnose in order to select the best method. It becomes more difficult when various ML classifiers and evaluation methods with different criteria are involved. Moreover, there are various ML classifiers to detect COVID-19. The decision-makers face, it is difficult to decide on the best method to be used. Subsequently, it is crucial to determine, evaluate and benchmark the classifiers in order to validate their results, especially, in handling medical cases. The invalid results lead to devastating consequences including the death of a patient, financial loss and legal accountability. An example of an invalid result is when the automated COVID-19 diagnoses tool shows positive COVID-19 while the patient is not infected with COVID-19. In this case, the patient is given inappropriate treatments which may affect the patient's health and recovery. In contrast, when the automated COVID-19 diagnose tool shows a negative COVID-19 while the patient is infected with COVID-19. The patient is not giving the right treatment which eventually affects the patient's health and recovery and the patient transmit the infection to other healthy people. Both cases seriously affect the reputation and credibility of medical and health organisations [5,7].

Thus, it is very important to select and use the right ML diagnosis method that provides reliable and accurate COVID-19 results. Acquiring such methods is costly. It also needs extensive evaluation and benchmarking for safety purpose. Conducting a comprehensive evaluation of the classifiers of COVID-19 is not an easy task. The evaluation process requires considering a huge number of measurements. The benchmarking and evaluation of COVID-19 diagnosis based on ML have two common criteria which are (i) time complexity and (ii) group reliability. Several sub-criteria belongs to the group reliability including precision, f1-score, recall, average-accuracy, error rate, true positive (TP), true negative (TN) , false positive (FP), and false negative (FN). In the study carried out by Cheng et al [8], many classifiers were evaluated based on the accuracy criterion. Others studies such as [5, 7] also only used accuracy criterion to evaluate and benchmark different diagnosis ML methods. However, in the COVID-19 diagnosis, it is important not to limit the use of only the common evaluation metric of accuracy [9]. In the context of medicine field, considering a variety of performance criteria for the diagnosis methods is important. Four instances, sensitivity criteria of True Positive, False Positive, True Negative and False Negative were used in [5, 7, 8, 9] for evaluation and benchmarking. However, these studies neglect some requirements that might affect the performance of diagnosis methods. Time complexity considers the execution time to provide a diagnosis result. Ultimately, the diagnosis method requires to maintain an accurate result in a shortest time. In reality, these requirements are conflicting in which achieving high accuracy requires time. One of those requirements must be more focused on when developing the classification model. The trade-off between accuracy and time has a strong impact on the benchmarking and the evaluation task [10]. Most review papers on comparison the state-of-art and traditional classifiers focus on the evaluation process itself regardless of the evaluation and benchmarking criteria used in the comparison. One of the reasons might be the difficulties when dealing with conflicting evaluation criteria [1, 61]. Besides the importance of dealing with the trade-off of the evaluation criteria, the importance of each criterion is another issue in the evaluation and benchmarking of classifiers.

In a COVID-19 classifiers system, several criteria are considered in the evaluation process. Each criterion has an important priority and it is used for a specific purpose that meets the system objectives. It is possible to reach a good combination between the low importance priority of a criterion with the high important priority of another criterion in the evaluation process [5, 9]. This is the reason for the existence of the trade-off among the evaluation criteria in different classifiers systems is a result of the differences in the importance of each criterion [61]. The simultaneity of many criteria and sub-criteria is another issue in the benchmarking processes. The conflict of the criteria and their distinct importance makes the benchmarking processes to be more difficult. A criterion set of the reliability lays on the confusion matrix that includes four arguments True Positive, True Negative, False Positive and False Negative. In experiments, some of arguments might miss out their values, this affects other values of the rest arguments in the reliability group. There is a disparagement of using these arguments but many evaluation studies still utilised these arguments for benchmarking processes to evaluate the classifiers systems. Although there exist several evaluation and benchmarking methods, most of them still considered insufficient.

The diagnosis model requirements are not fully covered using these evaluation methods. In addition, these methods have the reliability limitation as they require the calculation of all arguments. Also, they cannot compare two classifiers and match between them due to the inability of these methods to score the different classifiers based on their performances. The benchmarking and evaluation process in the COVID-19 classifiers systems considered as multi-objective/ criteria problem. This study aims to present an integrated methodology for evaluation and benchmarking various classifiers for COVID-19 diagnosis. This is a motivation to develop combined classifiers under one framework including all the performance aspects of the evaluation of COVID-19 classifiers models. The developed methodology is used as an assist tool to help the decision-makers in the medical and health organisation to decide which the best classifiers system should be used for COVID-19 diagnosis by evaluating different classifiers models.

The novelty and contributions of this work can be summarized as follow:

- Construct a multi-criteria decision matrix (DM) on the basis of determined evaluation criteria for COVID-19 diagnostic model.
- Proposes a new evaluation and benchmarking methodology for selecting the optimal COVID-19 diagnostic model based on Entropy and TOPSIS methods.
- Evaluate the proposed Benchmarking Methodology using 50 samples of Chest X-ray Dataset of COVID-19.

The remaining parts of this study are composed of six sections: “II” includes the related work of initial studies showing the abnormalities found in the COVID-19 diagnostic. “III” presents the Multicriteria Decision Making. “IV” provides the proposed methodology providing the benchmarking methodology for COVID-19 diagnosis models. “V” illustrates the results of Benchmarking Methodology for the selection of the optimal COVID-19 diagnostic model based on entropy and TOPSIS methods. “VI” addresses limitation and future studies .Finally, “VII” conclusion have been presented.

II. Related Works

Coronavirus disease known in the medical field as (COVID-19) has a disastrous impact on global health around the world. The severe cases of infected people are suffering from acute coronavirus respiratory syndrome known as (SARS-CoV-2). The major action has been taken to beat the COVID-19, is based on delivering the intensive care to patient’s that got a positive screening diagnosis of COVID-19. Providing them with the right treatments and mitigating the side effect as much as possible. Keeping the potentially infected people away from other people to prevent spread and transmit the virus. Polymerase Chain Reaction (PCR) [12] is one of the common examination methods that use to diagnosis the COVID-19. PCR mainly identify RNA of SARS-CoV-2 from respiratory samples usually collected from the potentially infected people through various testing methods whether oropharyngeal or nasopharyngeal specimens. Although the PCR is an efficient examination method, it requires resources and very long-time to conduct such complex process manually. Another examination method that wieldy used in radiography examination including Computed Tomography CT imaging and X-ray screening. COVID-19 is detected by screening the chest radiography imaging. Radiologists mainly interpret and analyses any abnormalities that might found in the chest imaging.

Initial studies show the abnormalities found in the chest imaging are strongly related to the presence of COVID-19 [13, 14]. A recommendation promotes the medical professional to use radiography screening as the major method for diagnosis of some epidemic diseases such as COVID-19 and SARS-CoV-2 [15]. The radiography screening method is rapid, efficient and available in most medical centers and hospitals. These features make the chest radiology imaging testing is a good tool especially it can supplement to PCR examination for higher sensitivity cases. [16] However, radiography examination relies on the radiologist’s experience and knowledge for interpretation and analysis as some visual metrics are fuzzy. A solution to overcome this dilemma is the utilization of a rapid computer-assist based AI techniques tool to help the medical practitioners in interpreting the chest radiography imaging of COVID-19 and gain an accurate diagnosis in a short time.

According to Wang et al (2020), A COVID-Net based on deep convolutional neural network (DCNN) is proposed for COVID-19 diagnosis using radiography images of chest screening. The COVID-Net train on radiography dataset for more 16,756 chest radiography images obtains from 13,645 patients. The COVID-Net is open access and available dataset for public use [3]. Moreover, COVID-Net is able to predict the COVID-19 and provides a deep explanation of critical indicators for COVID cases which could enhance the radiography examination. A computer-aided tool based on AI is developed by Gozes, et al. 2020[4]. The developed tool analyses the CT images to detect and identify Coronavirus cases automatically. The obtained results reveal that the diagnosis accuracy of CT images is over 99% using datasets of Chinese COVID-19 cases with potential factors of 92.2% for specificity and 98.2% for sensitivity. Also, Chen et al. 2020 [8], proposed a diagnostic model based on deep learning to proposed to diagnosis the COVID-19 using good quality of CT chest images. The proposed models produce good accuracy, sensitivity, specificity results equal to 95.24%, 100% and 93.55% respectively for each patient. A prediction approach based on ML techniques is presented by Yan et. al.2020 [9], the proposed approach predicts the severe cases for the infected patients in Wuhan City. The predation process based on three clinical indices that tracking the severe COVID-19 progress, and even to identify the potential death cases. The proposed model is useful in the early diagnosis to avoid the health consequences of the COVID-19 and increasing the chances of life.

Zheng et al. 2020 propose another COVID-19 diagnostic tool based on supervised Deep Learning [17]. The proposed tool detects COVID-19 through identify Weak Label for 3D CT Chest imaging. A pre-trained UNet divides the chest images specifically the lung area into 3D segments. The deep neural network used 3d lung segments to detect the COVID-19. The tool is trained using collected 499 CT images for the moths of December 2019 and January 2020. The tool tested on collected CT images for the months of January and February 2020. The proposed method achieves 0.959 for the area under the ROC Curve (ROC AUC) and 0.976 for Precision-Recall curves (PR AUC). Wang et al. 2020 propose a new study on investigation the use of deep learning methods for diagnosis of the COVID-19 [18]. The study is applied to CT COVID-19 images to provide an automatic diagnosis before the laboratory test to speed up the clinical investigation process. The validation results demonstrate the proposed model yields accuracy up to 89.5% and sensitivity equal to 0.87 and specificity equal to 0.88. While the testing results show the proposed model yields accuracy up to 79.3% and sensitivity equal to 0.67 and specificity equal to 0.83. Xu et al. 2020 that employ Deep Learning method to diagnose COVID-19 using chest CT images conduct another study. The aim of the study is distinguishing between the novel coronavirus and Influenza-A virus. The results of evaluation achieve accuracy of up to 86.7 %. As reported in the scientific literature, artificial intelligence (AI) and Machine learning techniques, particularly deep learning method, are applied for COVID-19 diagnosis using chest radiography imaging such as CT screening images. The evaluation results demonstrate those methods are promising in term of delivering an automatic COVID-19 diagnosis with high accuracy in a short time. [19].

Despite that, the best of the authors' knowledge, these AI-based systems are non-open sourced and not publicity available which prevent the research to access and investigate them for further research. Recently, many researchers tend to enrich the open sources with AI and ML solutions for COVID-19 diagnosis that used radiography imaging screening. Chest X-ray Dataset of COVID-19 is now being open access and available for public use. The quality of the AI algorithms and, in particular, the ML classifiers that used for COVID-19 diagnose should be evaluated and benchmarked to guarantee that fulfilled the main requirements in term of the reliability and the complexity of time. According to Yan et al. 2020 [9], the AI and ML models based on radiography-driven could deliver an accurate and efficient solution for COVID-19 which help to early diagnosis. There are a massive number of that conducted on diagnosis COVID-19 radiography imaging. However, regardless of the considerable effort in proposing of new COVID-19 diagnosis tasks of radiography imaging, no research has produced an incorporated solution for evaluation criteria covering for evaluating and benchmarking the models of diagnosis and supports medical organization administrators and various users identify an appropriate model. This study tries to fill the research evaluation and benchmarking gap as regards the COVID-19 through radiography imaging diagnosis tasks.

III. Multi-criteria Decision Making (MCDM)

According to [20], MCDM is an extension of decision theory covering any multi-objective decision. It is a methodology used to evaluate alternatives, which are often conflicting criteria, and to integrate them into a distinct universal appraisal. MCDM is a common decision-making method widely used in the operational research field to handling the decision-making criteria issue [21, 22, 23]. Moreover, MCDM utilises a variety of criteria that help in the planning and building the structure and finding the solution for the decision-makers [24, 25]. In general, MCDM can be expressed as a method to conduct quantitative and qualitative comparison based on a different set of criteria concurrently [26, 27]. As the MCDM is more efficient and has a capability to improving the quality of the decisions compared to the traditional procedures, it becomes widely used by the decision-makers [28]. The MCDM aims to assists the decision-makers and data scientists to select the best choices among different alternatives. It aims also to categorize feasible alternative within set existing options then ranking their performances descending [29, 30]. The best feasible alternatives will be scored a higher rank.

The ranking aspect is the main principle for any MCDM and DM. It is based on the evaluation matrix (EM) within different criteria [31, 32]. In EM, n criteria and m alternative are created. The crossover of every criterion with every alternative is donated as xij. The matrix of (xij) _ (m*n) is presented with its parameters in the equation 1.

$$\boldsymbol{D} = \begin{array}{c c} & \begin{array}{cccc} C_1 & C_2 & \cdots & C_n \end{array} \\ \begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_m \end{array} & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \end{array},$$

$$W = [w_1 \ w_2 \ \cdots \ w_n], \qquad (1)$$

A1, A2… Am represent the possible alternatives to be ranked by some image such as dehazing algorithms [33]. *C1, C2… Cn* represent provided criteria and used to measure every alternative. Xij represent the score of alternative Ai that associated with criterion Cj while Wj is value of criterion $Cj's$ weight. To make a ranking process, many methods in the scientific literature apply to make the ranking process inducing the addition of weights, the maximization index, and the normalization process.

In the scientific literature also there are many MCDM techniques including hierarchical adaptive weighting(HAW), analytic hierarchy process (AHP), the weighted product method(WPM), the VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR), the weighted sum model(WSM), simple additive weighting(SAW), the best-worst method (BWM), Multiplicative exponential weighting(MEW), the technique order of preference by similarity to ideal solution (TOPSIS) and the analytic network process(ANP) [32-44]. The huge number of MCDM methods makes the selection task more difficult as each method has its advantages and disadvantages [42, 45]. The selection of the best MCDM technique is a critical task. To the best of authors' knowledge, there is no MCDM method is proposed for the evaluation and benchmarking and ranking COVID19 diagnostic models.

IV. Proposed Methodology

To best of our knowledge, this is the first research being suggested to evaluate and benchmark of COVID19 diagnosis models based on X-Ray chest image. This study aimed to use AI to ease the problem of professional interpretations shortage for X-Ray chest images while the epidemic is still spreading rapidly. The proposed benchmarking methodology for COVID19 diagnosis models shows in Figure1.

A. *Machine learning models*

From an automatic learning perspective, COVID-19 diagnosis can be viewed as a problem of classification or clustering. On the other hand, on the vast set of file data absence and presence, we formed a model which can reduce the diagnosis problem. This problem could be reduced to only one diagnosis for known families-having a limited class is set, including the sample diagnosis of COVID-19, it's simpler to find the correct class, and the result will become more accurate than with algorithms of clustering. The theoretical background is provided in this section regarding all the methods that have been used in this research. Twelve Machine Learning Algorithms that will be used as alternatives in the DM are discussed for comparative analysis. Naive Bayes, Neural Network, SVM(linear), RBF, kNN, SGD, Logistic Regression, Random Forest, Decision Tree, AdaBoost, CN2 rule inducer, and SVM(Polynomial).

- **Naïve Bayes (NB)** is another basic supervised machine-learning model based on the theorem of Bayes, with the assumptions of observations independency. Main advantages NB that its learning model is straightforward and requires no complex calculation of the iterative parameters. The NB model, despite its simplicity, can outperform more sophisticated models of machine learning. Bayesian Bayes main problem is to obtain the most appropriate network data testing for training data [46].
- **Neural Network (NN)** Is an Artificial Intelligence (AI) numerical learning techniques family (e.g. simulated annealing and genetic algorithm) and is adapted from neural networks of human biological. This includes several computational nonlinear elements that shape neurons or nodes in the network, and weighted interconnections linked the nodes. In applications requiring data processing or performing tasks with a high complexity level, NN is especially useful which makes it impractical or difficult to solve by human hand. One of NN's primary properties is its ability to conclude a function from observations. Many fields of research used the NN including clustering, classification, approximation, compression, filtering, and separation of blind sources [2, 10, 47].
- **Support Vector Machine (SVM)** is a well-known supervised machine learning model which is mainly utilized for problems of binary prediction. The fundamental idea of this model is based on the hyper-plane and margin principles. The process of learning involves finding a linear separator (also known as a hyperplane) that separates the training data while maximizing the margin between these training data and the hyperplane. In some cases, in its original representation, SVM cannot clearly find a linear separation between the data. Thus, a training data transformation suggested by Vapnik [1] is carried out from the original space to another higher dimensional space to be able to find a linear separator between the groups. A kernel function such as the quadratic, Gaussian, or polynomial kernel functions may be used to make this transformation.

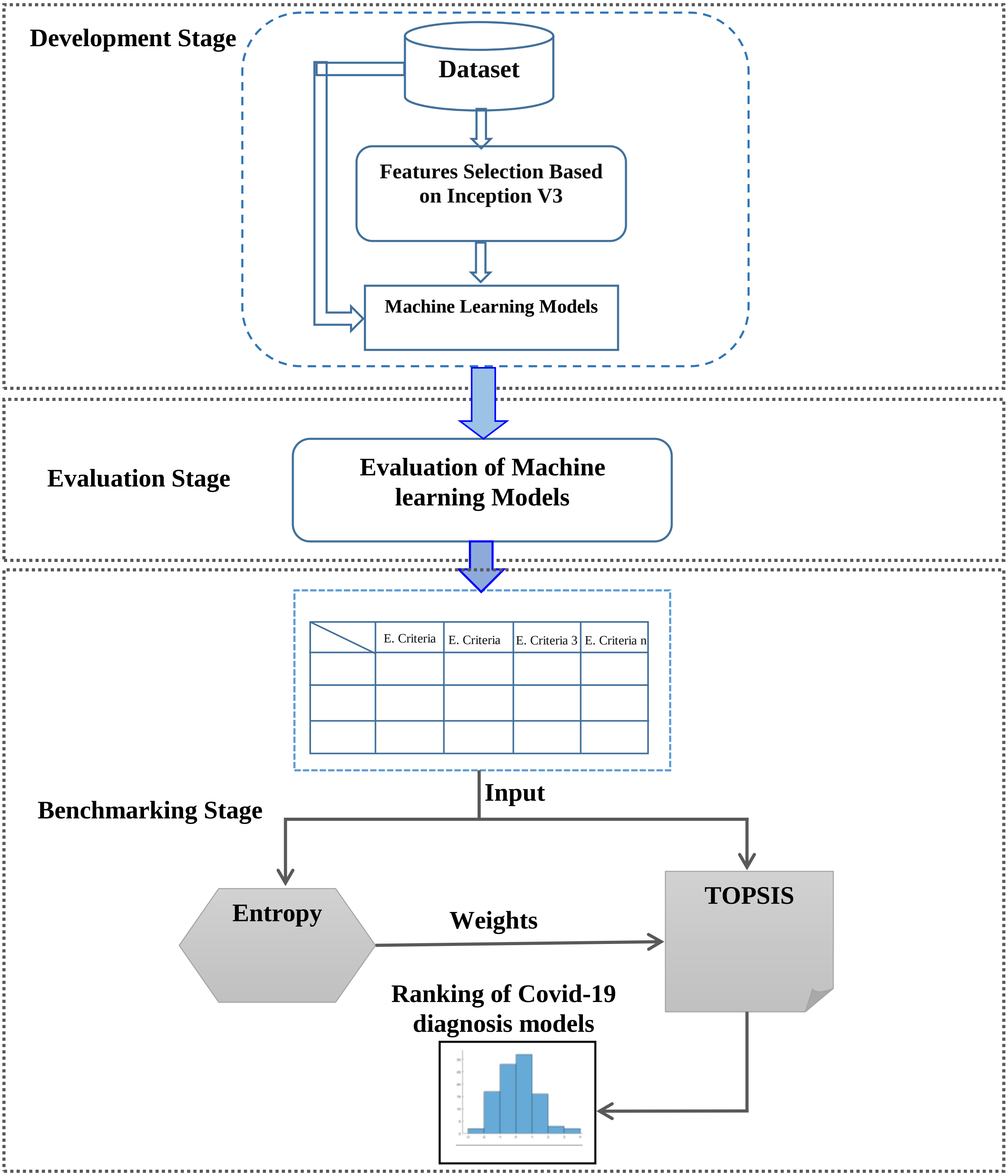

Figure1: The proposed benchmarking methodology for COVID19 diagnosis models

- **Radial basis function (RBF)** The SVM kernel efficiency has been proven on both linear and nonlinear results. With this algorithm, the Radial base function was implemented to classify nonlinear data. Kernel function demonstrates a very significant role in bringing data into the space of features. Attributed to the kernel function, flexible nature of SVMs implicitly transforms data into a Higher-dimensional feature space. The dimensional input space is reduced by a linear solution in a higher-dimensional function space which is originally associated with a nonlinear solution. This is because, in hydrology and hydraulics, which are usually nonlinear, the SVM approach is considered a reasonable choice. Several approaches use the nonlinear kernel function to solve regression problems in a strategy [48].
- **K-nearest neighbors (KNN)** uses a rule of distance closeness for each dataset to data having an equal property. It is possible to classify data which have no classes by their closest neighbor observing. Data equality can be achieved by distance measurement, usually using Euclidean measurement. We use 1 neighbor (k=1) in this study to measure the distance between the data and without weighting the distance [49].
- **Stochastic Gradient Descent (SGD)** is considered as basic but very successful approach discriminative learning of loss functions convex like logistic regression and (linear) SVM. While in the machine learning environment SGD has been running for a long time, just recently it has gained significant attention in the large-scale context learning. SGD has been applied significantly to solve large-scale problems and machine learning sparse often encountered in the classification of text and the processing of natural languages [50].
- **Logistic Regression (LR)** in the models of Machine Learning, Logistic regression, despite having the word 'regression' in its name, is a kind of parametric classification model. This indicates that the models of logistic regression are models with a confirmed constant parameters number that depends on the input features number, and they predict categorically, such as whether or not a plant exists in a certain species. LR is one of the popular and basic classification problems solving algorithm [51].
- **Random Forests:** A classification process uses an ensemble method widely in which uses various decision trees to classify data. It creates the templates of bootstrap from original data of the random forest and for each bootstrap template it grows a regression tree or raw classification. Rather than selecting the only best predictors for disclosure, it considers every node. It implements a random predictors selection and selects the preferable split among them [52].
- **Decision Tree:** is a component of a flowchart in which each internal node represents the attribute test, each class label and branch represents the test outcome and is represented by each leaf node. Classification rules represent the paths from the root to leaf. The decision tree consists of three types of nodes, the first node is decision-making (usually denoted by square), the second one is the chance node that is used for shaping the options (usually denoted by a circle) and the last one is end node for representing the action (usually denoted by a triangle) [53].
- **AdaBoost,** Adaptive Boosting short, is a meta-algorithm of machine learning. Robert Schapire and Yoav Freund have created this algorithm who received the Gödel Prize for their work in 2003. It can be used to enhance performance, in conjunction with several other forms of learning algorithms [53].
- **CN2 rule inducer algorithm** is a rule induction learning algorithm. It is designed to be implemented even when the data training is imperfect. It is based on the ideas of the ID3 algorithm and AQ algorithm. As a result, it generates a set of rules like the one provided by AQ but it has the capability of noisy data handling such as ID3. To produce a list of classification rules, the algorithm must be given a set of examples, Training Set, that have already been categorized. A set of conditions, the Simple Condition Set, that can be applied to any set of instances, alone or in combination, is predefined to be used for classification [53].
- **SVM(Polynomial)** the polynomial kernel in machine learning, is a kernel function widely used for (SVMs) and also other kernelized models, in which the vectors similarity (training samples) are characterized in a feature-space above the original variable's polynomials, permitting nonlinear models to be learnt. Intuitively, to determine their similarity, the polynomial kernel will not look only at the specified input samples features but also at their combinations. Those combinations are classified as features interaction in the analysis regression context. The implicit feature space of a polynomial kernel is identical to that of polynomial regression but without the combinatorial blow-up in learned parameters number. When Booleans features input (binary-valued), then the features suit in logical input of the conjunctions [53].

B. *Evaluation Criteria*

The performance of each method was evaluated at this phase to see which method could achieve the best result. We used the accuracy parameters; F1 score, ROC, Recall, Precision, False Positive, and True Positive from the confusion matrix to evaluate each of the methods used in this research. Confusion matrix contains FN (False Negative), FP (False Positive), TN (True Negative), and TP (True Positive).

- **Accuracy** refers to a measurement's closeness parameter when reading the data value against the actual data values.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100 \qquad (2)$$

- The **Precision** metric tests the proportion of related subjects. It tests the classifier's ability to reject irrelevant subjects. The calculation of metric is as follows

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

- **The Recall metric** measures the proportion of identified relevant subjects. It measures the classifier's ability to produce all applicable subject. The calculation of metric is as follows

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (4)$$

- **ROC curve** is a performance measurement at several threshold classification problem setting. ROC is a curve of probability, and AUC is a metric of degree or separability. It shows the capability of the model to distinguish among classes. Higher the AUC, the better a model is to predict 1s as 1s and 0s as 0s. By contrast, the higher the AUC, the better the model is to differentiate between disease patients and no illnesses.
- **The F1 score** can be viewed as a recall and precision weighted average, where a score of F1 achieves its worst at 0 and best value at 1. The recall and precision relative contribution to the score of F1 are similar. The F1 score formula is:

$$F1 = \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (5)$$

- **TP**: output predicted as true positive (TP), we concluded that the subject of coronavirus disease is classified correctly, and subjects are affected by a coronavirus.
- **TN**: output predicted as true negative (TN), we concluded that a healthy subject is classified correctly and that the subject is safe.
- **FP**: output predicted as false positive (FP), we concluded that a safe subject is wrongly labeled (Type 1 error) as having coronavirus disease.
- **FN**: output predicted as false negative (FN), we concluded that a coronavirus disease is classified incorrectly (Type 2 error) as having no coronavirus disease because the subject is healthy.

C. COVID-19 Dataset

Concerning the advent of the new coronavirus epidemic, datasets are now being generated and annotated. The data sources and knowledge needed to label the particular data relating to the latest human strain of the virus are very limited. As such, clinically applicable learning at this early stage of data collection cannot be done because there are no appropriate samples. Nonetheless, the use of X-ray images, obtained from data given by [54], was employed in this study. The data provides a wide COVID-19 samples range alongside ARDS, MERS, and SARS cases. Images of ARDS, MERS, and SARS from different countries have been withdrawn. The experiment was performed using a total of 50 X-ray images, of which 25 were positively COVIDE-19 cases, while the other 25 were normal cases. Varied images resolution is used between 700p and 3342p. The data set was used for training, and various learning models were then used for performance evaluation. Dataset sample is shown in Figure 2.

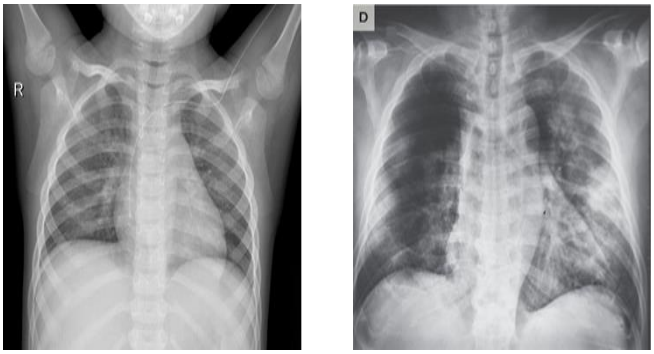

Figure 2. Samples of X-ray dataset [54]

The use of standard InceptionV3 model [55], that has been used in many medical imaging applications, is used as a feature extractor in this work [56]. This model was particularly used to detect pneumonia via Chest X-Rays [55]. Google has developed the Inception V3 architecture with weights pre-trained on ImageNet, which is an extended version of Inception V1. Originally the model's training was performed using some very powerful machines on more than one million photos from 1,000 students.

D. Construction of Decision Matrix(DM)

The decision matrix of COVID-19 diagnostic designed rely on the overlap between the evaluation criteria, namely, accuracy, precision, accuracy, recall, FP, FN, TP, TN, AUC, score and time consuming, and the sophisticated 12 models of classification. The proposed decision matrix structure is presented in Table 1. The upper row exemplifies the key evaluation criteria, and different alternatives developed models represented on the left to the first column for classification. In this DM, the values (data) indicate the results of the evaluation for all models of classification that have been developed according to all evaluation criteria. All evaluation criteria have been used to evaluate each model of classification, in which the category of reliability with four sub-criteria sets are represented by error rate and parameter’s (behaviour, relationship, and matrix). Firstly, the parameter matrix (TP, TN, FN, and FP) is generated, and these parameters represent the basic sub-criteria in the criteria reliability group. Since this study tackled the issue of classification the, one-verse all approach is utilized to measure the criteria's reliability set. The confusion matrix is converted into three confusion matrices according to the aforementioned requirements, and every matrix specifies the

Table 1: Structure of a decision matrix

Criteria / Model	Recall	Precision	Accuracy	F-Score	AUC	FPR	FNR	TNR	TPR	Time Complexity
Model1	RV (M1/TS)	PV(M1/TS)	AV(M1/TS)	FV(M1/TS)	AuV(M1/TS)	FPV(M1/TS)	FNV(M1/TS)	TnV(M1/TS)	TpV(M1/TS)	TCV(M1/TS)
Model2	RV(M2/TS)	PV(M2/TS)	AV(M2/TS)	FV(M2/TS)	AuV(M2/TS)	FPV(M2/TS)	FNV(M2/TS)	TnV(M2/TS)	TpV(M2/TS)	TCV(M2/TS)
Model3	RV(M3/TS)	PV(M3/TS)	AV(M3/TS)	FV(M3/TS)	AuV(M3/TS)	FPV(M3/TS)	FNV(M3/TS)	TnV(M3/TS)	TpV (M3/TS)	TCV(M3/TS)
Model4	RV(M4/TS)	PV(M4/TS)	AV(M4/TS)	FV(M4/TS)	AuV(M4/TS)	FPV(M4/TS)	FNV(M4/TS)	TnV(M4/TS)	TpV (M4/TS)	TCV(M4/TS)
Model5	RV(M5/TS)	PV(M5/TS)	AV(M5/TS)	FV(M5/TS)	AuV(M5/TS)	FPV(M5/TS)	FNV(M5/TS)	TnV(M5/TS)	TpV (M5/TS)	TCV(M5/TS)
Model6	RV(M6/TS)	PV(M6/TS)	AV(M6/TS)	FV(M6/TS)	AuV(M6/TS)	FPV(M6/TS)	FNV(M6/TS)	TnV(M6/TS)	Tp(VM6/TS)	TCV(M6/TS)
. . . .	....	....	....	....	....	....	....	....	....	....
Model n	RV(Mn/TS)	PV(Mn/TS)	AV(Mn/TS)	FV(Mn/TS)	AuV(Mn/TS)	FPV(Mn/TS)	FNV(Mn/TS)	TnV(Mn/TS)	Tp(Mn/TS)	TCV(Mn/TS)

M:Model, **RV**: Recall Value, **PV**: Precision Value, **AV**: Accuracy Value, **FV**: F-Score Value, **AuV:** AUC Value, **FPV**: FPR Value, **FNV**: FNR Value, **TnV**: TNR Value **TpV**: TPR Value,, **TCV**: Time Complexity Value; **n**: number of algorithms, **TS**: Test samples.

particular parameters for a class of COVID-19 (non-COVID-19 and COVID-19). Within the reliability group, the remaining sub-criteria are determined for each matrix using a particular formula, based on the three confusion matrices. Therefore, values will be calculated separately for each classification model to produce the values taking into consideration the decision matrix input. Lastly, the procedure of time complexity calculation is based on two elements elapsed time: data set sample input and result outcomes. The computation method for the sampling procedure depends on the sample size and number as shown in the below equation:

$$Tprocess = To - Ti \qquad \textbf{(6)}$$

To indicates the time of processing to get outputs and Ti is the sample input time. Through the experimental process, the complexity of time is determined using software named Weka. Three main issues are discussed in Section 2, the matrix of the proposed decision considers as following: (1) the trade-off and conflict between criteria evaluation, (2) multiple criteria evaluation pattern, and (3) the criteria significance. MCDM is used to resolve such problem as discussed in the following section.

The decision matrix takes into account the basic component of the methodology for benchmarking and evaluation. Decision criteria and alternatives are the core aspects of the decision matrix. The criteria, in the current case, reflect the used metrics to measure the consistency of models for diagnosis. The following paragraph explains the methods followed for diagnosis models development and evaluation and the design of the decision matrix.

1) Development of diagnosis models

Designing models for classifying involves a process of three-steps. First of all, the target dataset is prepared, which involves relevant features selection. Secondly, training (learning process) is accomplished by evaluating the instances of a training dataset, which includes the creation of a machine-learning class. Each instance must belong to a predefined class, so it is supposed that each instance belongs to a predefined class. Third, algorithms of machine learning are performed in combination with other separate datasets, sometimes referred to as datasets testing. In this stage is consistent with the intent of estimating ML methods. Will the execution may seem 'appropriate' for the diagnosis model, the model may be used in future cases classification where the class label is uncertain. Eventually, models of diagnosis that provide an appropriate result may be considered a suitable model of diagnosis. COVID-19 Dataset has employed the use of X-ray images, and these images were derived from data given by [54]. The data provides a wide COVID-19 samples range alongside ARDS, MERS, and SARS cases. Images of ARDS, MERS, and SARS from different countries have been withdrawn. The experiment was performed using a total of 50 X-ray images, of which 25 were positively COVIDE-19 cases, while the other 25 were normal cases. Varied images resolution is used between 700p and 3342p. The data set was used for training, and various learning models were then used for performance evaluation. In addition to their unfavourable impact on diagnosis performance, the diagnosis model may cause conflict. In addition, given that irrelevant cases can result in overfitting, Positive effect can be achieved by decreasing the number of normal cases. This method will reduce calculated input to a minimum. This favourable impact the total diagnosis results and performance [3]. The COVID-19 that are highly important to diagnosis groups, which are called insightful instances, are chosen in this study.

A total of 12 diagnosis models are developed based on 12 well-known algorithms for machine learning, which were used in some of the previous researches and satisfactory results demonstrated when used in the COVID-19 dataset diagnosis. These algorithms include SVM(Polynomial), CN2 rule inducer, AdaBoost, Tree of Decision, Random Forest, Logistic Regression, Logistic Regression, Logistic Regression, SGD, kNN, RBF, SVM (linear), Neural Network, and Naive Bayes. To develop models of diagnosis, the data set is divided into two parts. The main part is used for the aim of training, and the second part is employed for the aim of testing. The training set is utilized to train the algorithms of machine learning, and the other part of the data set (testing set) is used to evaluate the trained algorithms of machine learning. Using the 12 diagnosis models, the test dataset is categorized into two groups, namely COVID cases, and normal.

E. Integrated Entropy and TOPSIS

The proposed benchmarking and evaluation methodology are developed using the techniques of MCDM. Based on the integration of TOPSIS and Entropy, this methodology is developed for alternatives weighting, ranking, and selecting the superior one in the matrix proposed for the decision. The following steps are discussed below. Entropy and TOPSIS are the appropriate methods for ranking and benchmarking of models for diagnosis. The TOPSIS approach is a recommended rating mathematical model and overcome particular problems relevant to (1) conflict and trade-off, (2) encountered evaluation of multi-criteria in the suggested decision matrix. The Entropy is often proposed to implement criteria weighing, and (3) the value of relational criteria to the suggested matrix of decisions. Consequently, it is justified to combine TOPSIS and Entropy methods to rank models of diagnosis and benchmark them.

1) Entropy Method

The method of Entropy should not only be used for quantitative measurement of data volume but also to objectively compute the proportional weight information [57]. Entropy was originally supposed a physical phenomenon implies of the degree of turbulence to the numerator or the probability scale under a given condition. If the values of entropy are smaller, the numerator grades are proportionality greater, indicating the nearest possible entropy to the perfect. Conversely, the numerator grades inflect more irregularly if entropy values are higher. Hence, the weight of the entropy method was implemented to get the weighting of every attribute. Furthermore, entropy can be used in information theory to calculate the predicted value of information in a specific message. Entropy is a criterion for the amount of uncertainty in information theory that a discreet distribution of probabilities represents [58]. For measuring the entropy values, every attribute has a value assigned calculated by each alternative. The entropy values are then compared with each criterion, and relative levels of significance (i.e. relative weight) of each other are calculated. Entropy weight is then derived from information of the measurement matrix, which is part of the values of objective weight. The calculation process for the Entropy Weighting method is as follows.

The set of alternatives decision matrix contains some details; entropy may be employed as a method in evaluating criteria [58]. The entropy value can be used to calculate the knowledge quality of project results with criteria j. However, *Pij's* significance is calculated by the differential of all alternative's outcomes, indicating why *Pij* can be adapted by criterion *j* to the average inherent information provided by the alternatives collection. m alternatives and n attributes (criteria) are decision matrix D.

$$D = \begin{array}{c} \\ A_1 \\ A_2 \\ \vdots \\ A_m \end{array} \begin{bmatrix} X_1 & X_2 & \cdots & X_n \\ X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix}. \tag{7}$$

The project results of the *j*, *Pij*, the attribute can be described as

$$P_{ij} = \frac{X_{ij}}{\sum_{i=1}^{m} X_{ij}}, \quad \forall_{i}, j. \tag{8}$$

The entropy *Ej* of the attribute *j* project outcome set is

$$E_j = -k \sum_{i-1}^{m} P_{ij} \, \ell n \, P_{ij}, \qquad \forall j. \tag{9}$$

Noteworthy, the range values of entropy must be between $0 \leq E_j \leq 1$, and *k* is a positive constant, $k = \frac{1}{\ell nm}$ (*m* decision alternatives indicated).

The degree of diversification *dj* of the information issued by attribute j results can be calculated as

$$d_j = 1 - E_j, \quad \forall j. \tag{10}$$

The optimal weight collection, rather than the equal weight, is

$$W_j = \frac{d_j}{\sum_{j=1}^{n} d_j}, \qquad \forall j. \tag{11}$$

2) *TOPSIS Method*

Similarity technique suggested by Hwang and Yoon (1981) for Ideal Solution (TOPSIS), which allows policymakers to evaluate the Ideal positive (A*) solution and ideal negative (A-) solution. TOPSIS is premise-based that preferred alternative will be the smallest distance from the ideal and the farthest from the perfect negative solution [58, 59]. The solution is as follows:

Step 1*:* Standardized construction of the appraisal matrix and use the formula below:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \tag{12}$$

On the standardized appraisal matrix, x_{ij} is the *i* alternative to be evaluated according to criterion *j* alternatives, *j* stands for performance criteria, and *i* denote the alternatives. Therefore, every attribute has a similar unit length as the vector.

Step 2*:* Create the matrix of standardized weighted evaluations. Incorporate the model of entropy weight as an efficient method for objective evaluation of entropy weight *w=(w*1*,w*2*, . . . ,wj , . . . ,wn)* and $\sum_{j=1}^{m} W_{j=1}$. The weight of entropy multiplied by a standard evaluation matrix could be represented as:

$$V = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1n} \\ V_{21} & V_{22} & \cdots & V_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ V_{m1} & V_{m2} & \cdots & V_{mn} \end{bmatrix} = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \cdots & w_n r_{1n} \\ w_1 r_{21} & w_2 r_{22} & \cdots & w_n r_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ w_1 r_{m1} & w_2 r_{m2} & \cdots & w_n r_{mn} \end{bmatrix} \tag{13}$$

Step 3*:* Determine the positive ideal solution (*A*∗) and negative ideal solution (*A*−)

$$A^* = \{[max_i V_{ij} | j \in J'] | i = 1,2,\dots,m\} = \{V_1^*, V_2^*, \dots, V_j^*, \dots, V_n^*\} \tag{14}$$

$$A^- = \{[max_i V_{ij} | j \in J'] | i = 1,2,\dots,m\} = \{V_1^-, V_2^-, \dots, V_j^-, \dots, V_n^-\} \tag{15}$$

Step 4*:* Calculate the measure of separation (S∗i, S−i). The distance between A∗ and A- alternative can be measured as

$$S_i^* = \sqrt{\sum_{j=1}^{n} (V_{ij} - V_i^*)^2}, \qquad i = 1,2,\dots,m, \tag{16}$$

$$S_i^- = \sqrt{\sum_{j=1}^{n} (V_{ij} - V_i^-)^2}, \qquad i = 1,2,\dots,m, \tag{17}$$

Step 5*:* Estimate relative proximity for positive of the optimal solution. Defines a closeness coefficient for evaluating the order of ranking of all alternatives between S* and S−i have been determined for each alternative Ai I = 1, 2, . . ., m). The coefficient of closeness is determined for every alternative as

$$C_i^* = \frac{S_i^-}{S_i^* + S_i^-}. \tag{18}$$

Ai alternative is nearer to A and further towards A− as C range is approaching 1.

Step 6*:* Rate the Order of Choice. Determine the order of ranking of all alternatives depends on the coefficient closeness C∗i and choose the optimal one from a selection of applicable alternatives.

Even conforming the intentions of decision-makers and simply reasonable, this approach is limited in its inability to deal with the problem of materialization, requiring other approaches combination, like Entropy and TOPSIS combination. However, as shown by recent research, to be clearest, TOPSIS can be utilized for objects comparison that is good or bad in the quality situation [60]. Because TOPSIS is used to determine the ideal solution positive and the ideal solution negative, the assessment of alternatives takes various perspectives. A decision-maker engaged in decision-making may, therefore, prevent various rates of demand neglecting, and ultimately provide the superior alternative[61]. Furthermore, realized that TOPSIS offers an efficient method of deriving these problems from a consensus solution. The introduction of the consistency function and the method of analytical hierarchy, besides the fuzzy theory, to assess the characteristics of the informative decision systems [62]. During the meantime, TOPSIS can be utilized to effectively coordinate the information systems. Deng et al. (2000) demonstrated in a case study on China the feasibility of applying the TOPSIS methods to solving an inter-company problem related to the comparison. Comparing objective weighting methods shows that the entropy calculation correlates favourably with other approaches for the case study conducted with the updated TOPSIS approach [62]. This analysis, therefore, applies entropy to obtain objective weight values in order to achieve uniformity between experts. A calculation procedure to rank the alternatives is performed via TOPSIS. By combining the TOPSIS and entropy approach, this study will help hospital managers not only evaluate various criteria decisions systematically and objectively but also create a consistent standard model and efficiently make decisions [63,64]. The entropy method is then coupled with the TOPSIS model to pick the most suitable COVID-19 diagnosis model.

V. RESULTS AND DISCUSSION

The results of benchmarking methodology for Selection of optimal COVID-19 diagnostic model based on Entropy and TOPSIS Methods are presented in this section. Section 5.1 describes the DM data results were generated from the criteria and diagnosis methods. In section 5.2, the Entropy results for data normalization is discussed and the TOPSIS method results were presented and discussed in details in Subsection 5.3.

A. *Decision matrix d data*

The data used in our research is produced from the crossover between the diagnosis methods and the criteria. Table 2 presented all the details for the decision matrix that include twelve machine-learning classifiers. The evaluation process of ML models was done based on ten evaluation criteria. The upper row exemplifies the key evaluation criteria, and different alternatives developed models represented on the left to the first column for diagnosis. All of AUC, Accuracy, F-Score, Recall, TNR, and TPR are considered as benefit criteria where the maximum value is desired. While each of FPR, FNR, and Time-consuming criteria are considered as cost criteria where the minimum value is desired. The findings of the integration of Entropy–TOPSIS are discussed in the next section.

Table 2: Decision matrix data

Classifier	AUC	Accuracy	F-Score	Precision	Recall	FPR	FNR	TNR	TPR	Time
Naive Bayes	0.971	0.973	0.973	0.975	0.973	5.0	0	100	95	0.14
Neural Network	0.988	0.987	0.987	0.987	0.987	2.6	0	100	97.4	0.55
SVM(linear)	0.988	0.983	0.98	0.981	0.98	3.8	0	100	96.2	0.19
RBF	0.979	0.96	0.96	0.962	0.96	5.0	0	100	95	0.16
kNN	0.97	0.94	0.939	0.94	0.94	5.3	5.4	94.6	94.7	0.9
SGD	0.94	0.94	0.939	0.940	0.94	5.2	4.1	95.9	94.8	0.18
Logistic Regression	0.931	0.88	0.88	0.88	0.88	6.4	4.2	95.8	93.6	0.9
Random Forest	0.933	0.847	0.846	0.85	0.847	7.0	12.7	87.3	93	0.9
Tree	0.84	0.827	0.826	0.827	0.827	18.8	15.7	84.3	81.2	0.9
AdaBoost	0.814	0.813	0.813	0.814	0.813	16.7	20.5	79.5	83.3	0.15
CN2 rule inducer	0.80	0.72	0.72	0.73	0.72	29.4	27.6	75.4	70.6	7.57
SVM(Polynomial)	0.975	0.62	0.555	0.784	0.62	5.3	44.3	55.7	94.7	0.18

B. *Entropy weighting results*

The decision matrix of COVID-19 diagnostic designed rely on the overlap between the evaluation criteria, namely, accuracy, precision, accuracy, recall, FP, FN, TP, TN, AUC, score and time consuming, and the sophisticated 12 models of diagnosis. In this step, data normalization is demonstrated in table 3; and criterion weighting has been done objectively for 10 criteria for each model as shown in table 4.

Table3: The normalized data

Classifier	AUC	Accuracy	F-Score	Precision	Recall	FPR	FNR	TNR	TPR	Time
Naive Bayes	0.0872	0.0928	0.0934	0.0914	0.0928	0.0452	0.0072	0.0936	0.0872	0.0110
Neural Network	0.0888	0.0941	0.0947	0.0925	0.0941	0.0235	0.0072	0.0936	0.0894	0.0432
SVM (linear)	0.0888	0.0937	0.0941	0.0919	0.0934	0.0344	0.0072	0.0936	0.0883	0.0149
RBF	0.0880	0.0915	0.0921	0.0902	0.0915	0.0452	0.0072	0.0936	0.0872	0.0126
kNN	0.0872	0.0896	0.0901	0.0881	0.0896	0.0480	0.0390	0.0885	0.0869	0.0708
SGD	0.0845	0.0896	0.0901	0.0881	0.0896	0.0471	0.0296	0.0898	0.0870	0.0142
Logistic Regression	0.0837	0.0839	0.0845	0.0825	0.0839	0.0579	0.0303	0.0897	0.0859	0.0708
Random Forest	0.0838	0.0807	0.0812	0.0797	0.0808	0.0633	0.0917	0.0817	0.0854	0.0708
Tree	0.0755	0.0788	0.0793	0.0775	0.0789	0.1701	0.1134	0.0789	0.0745	0.0708
AdaBoost	0.0731	0.0775	0.0780	0.0763	0.0775	0.1511	0.1480	0.0744	0.0765	0.0118
CN2 rule inducer	0.0719	0.0686	0.0691	0.0684	0.0687	0.2661	0.1993	0.0706	0.0648	0.5951
SVM(Polynomial)	0.0876	0.0591	0.0533	0.0735	0.0591	0.0480	0.3199	0.0521	0.0869	0.0142

Table 4: The e and entropy weights values

	AUC	Accuracy	F-Score	Precision	Recall	FPR	FNR	TNR	TPR	Time
(e)values	0.9989	0.9966	0.9957	0.9982	0.9967	0.8847	0.7702	0.9954	0.9985	0.6176
Entropy weights	0.00141	0.00448	0.00582	0.00240	0.00447	0.15422	0.30741	0.00618	0.00207	0.51155

Tables 3 & 4 illustrate the gained values of (e) and weights of the Entropy for the criteria that have been obtained using Equations (7)– (11). The Time-consuming criteria achieved the maximum Entropy followed by each FNR and FBR. On another hand, AUC achieved the minimum Entropy weight. Based on the results, the criteria with the highest weight of Entropy are considered as the highest important criteria, whilst the criteria with the lowest weight of Entropy are of least importance. In another word, the group of cost criteria have gained higher importance than the group of benefits criteria.

C. *Ranking Results based on TOPSIS Method*

By combining the TOPSIS and entropy approach, this study will help a recommendation promotes the medical professional to use radiography screening as the major method for diagnosis some epidemic diseases such as COVID-19 not only evaluate various criteria decisions systematically and objectively but also create a consistent standard model and efficiently make decisions. The results of ML models ranking for COVID-19 diagnostic obtained from the weighted decision matrix was demonstrated in this part. The DM weighted was collected and generated by utilizing Equations (12) and (13). TOPSIS identified the best and worst performances of the COVID-19 diagnosis model for each criterion. Table 5 shows the result of the weighted DM, Ideal best, and Ideal worst value.

Table 5: The result of the weighted DM, Ideal best and ideal worst value

Classifier	AUC	Accuracy	F-Score	Precision	Recall	FPR	FNR	TNR	TPR	Time
Naive Bayes	0.000426	0.001429	0.001864	0.000771	0.001431	0.018479	0.000000	0.001982	0.000622	0.009168
Neural Network	0.000434	0.001450	0.001890	0.000781	0.001452	0.009609	0.000000	0.001982	0.000638	0.036018
SVM (linear)	0.000434	0.001444	0.001877	0.000776	0.001441	0.014044	0.000000	0.001982	0.000630	0.012443
RBF	0.000430	0.001410	0.001839	0.000761	0.001412	0.018479	0.000000	0.001982	0.000622	0.010478
kNN	0.000426	0.001381	0.001798	0.000744	0.001382	0.019588	0.027606	0.001875	0.000620	0.058939
SGD	0.000413	0.001381	0.001798	0.000744	0.001382	0.019219	0.020960	0.001900	0.000621	0.011788
Logistic Regression	0.000409	0.001293	0.001685	0.000696	0.001294	0.023654	0.021471	0.001898	0.000613	0.058939
Random Forest	0.000410	0.001244	0.001620	0.000673	0.001246	0.025871	0.064925	0.001730	0.000609	0.058939
Tree	0.000369	0.001215	0.001582	0.000654	0.001216	0.069483	0.080262	0.001671	0.000532	0.058939
AdaBoost	0.000357	0.001194	0.001557	0.000644	0.001196	0.061721	0.104800	0.001575	0.000546	0.009823
CN2 rule inducer	0.000351	0.001058	0.001379	0.000578	0.001059	0.108659	0.141097	0.001494	0.000462	0.495745
SVM (Polynomial)	0.000428	0.000911	0.001063	0.000620	0.000912	0.019588	0.226471	0.001104	0.000620	0.011788
Ideal best	0.000434	0.001450	0.001890	0.000781	0.001452	0.009609	0.000000	0.001982	0.000638	0.009168
Ideal worst	0.000351	0.000911	0.001063	0.000578	0.000912	0.108659	0.226471	0.001104	0.000462	0.495745

According to the Table 5, each COVID-19 diagnosis model has shown significant variance for each criterion in terms of ideal best and worst best value. On the other hand, each of Naive Bayes, Neural Network, SVM (linear), and RBF have gained same ideal best value in terms of FNR and TNR criteria which draw a significant challenge for select the best COVID-19 diagnosis model.

However, each performance was then compared with the ideal and worst performance. Si⁻ represents the closeness of a COVID-19 diagnosis model to the worst performance, and Si* represents the closeness of a COVID-19 diagnosis model to the best performance. However, the Equations (16) and (17) were used to calculate Si* and Si-. Finally, the closeness coefficient (Ci*) values for the twelve diagnosis models for COVID-19 diagnosis were calculated using the Equation (18). Therefore, diagnosis models were ranked and determined based on the TOPSIS principle. Table 6 shows the result.

Table 6: Ranking of COIVID19 diagnosis models

Classifier	S_i^*	S_i^-	C_i^*	Rank
Naive Bayes	0.0089	0.5442	0.9840	2
Neural Network	0.0269	0.5220	0.9511	5
SVM (linear)	**0.0055**	**0.5421**	**0.9899**	**1**
RBF	0.0090	0.5431	0.9838	3
kNN	0.0578	0.4881	0.8942	7
SGD	0.0232	0.5333	0.9583	4
Logistic Regression	0.0560	0.4900	0.8974	6
Random Forest	0.0834	0.4730	0.8501	8
Tree	0.1118	0.4623	0.8052	10
AdaBoost	0.1170	0.5031	0.8113	9
CN2 rule inducer	0.5162	0.0854	0.1419	12
SVM(Polynomial)	0.2267	0.4921	0.6846	11

The COVID-19 closest diagnosis model with the best performance and farthest from the worst performance will be the selected as the best COVID-19 diagnosis model. Visibly in table 5, the SVM (linear) outranks the other eleven diagnosis models which obtained closet value from an ideal solution and farthest value from the negative solution. Therefore, SVM (linear)is selected as the best diagnosis model for COVID-19 based on our case study data.

VI. LIMITATION AND FUTURE WORK

The proposed benchmarking methodology can solve problems of evaluation and benchmarking for the diagnostic models for COVID-19. In future studies the following aspects should be considered:

Firstly, other types of COVID-19 diagnostic models can be evaluated and benchmarked using the proposed methodology. Secondly, in the future studies, additional criteria will be studied and added into our proposed DM which can support the process of evaluation and benchmarking. Lastly, the use of the new methodology can be employed in benchmarking approaches based on the proposal of new decision matrices.

VII. CONCLUSION

Due to the globally spreading of COVID-19, the studies related to the automated diagnosis of COVID-19 have been notably increasing. The increase in the proposed diagnosis models has also raised a question of which diagnosis model is the best for a single diagnosis task and what type of criteria should be dependent. Therefore, an evaluation and benchmarking scheme is needed to solve this problem. In this study, we aim to bridge the gap of MCDM research and COVID-19 diagnostic models that are identified in the review of the scientific literature. There is no study has investigated and dealing with the COVID-19 diagnosis models in term of evaluation and benchmarking. In the scientific literature, the main evaluation criteria of diagnosis models for COVID-19 are only reviewed with no raising any research and challenges issues to bridge the gap. An intelligent MCDM based methodology is proposed for evaluating, benchmark, rank the diagnostic models for COVID-19. In this paper, we describe the mechanism to develop the proposed methodology. We construct decision matrix to mix the different twelve COVID-19 diagnostic models with ten evaluation criteria. Entropy and TOPSIS are integrated into our proposed methodology. The final weight results from Entropy steps were presented, showing the importance of the diagnosis model criteria. Subsequently, the TOPSIS approach was used to rank and select the best COVID-19 diagnosis models according to the quantitative information of the measured criteria. Even with comprehensive evaluation and significant results have presented by [65],comparing with our study, the benchmarked study has included several issues. First, no such mechanism has adopted to define the importance of evaluation criteria. Second, not easy to rank and select the best ML model based on multiple evaluation criteria and tackle the issue of variation among these criteria. However, our study has tackled all mentioned issues based on proposed integrated platform.

Finally, results confirm that (1) The Hybridization of the Entropy and TOPSIS can effectively solve the COVID-19 diagnosis models selection challenge. (2) The ranks of the COVID-19 diagnosis models obtained from TOPSIS showed that the best diagnosis model was SVM (linear) and the worst one was SVM(Polynomial). However, the values of the used models might be changed due to the availability of the data in terms of the type and volume.

In practical view, the huge number of ML learning models have drawn a great challenge to administration departments of medical organizations to select the most optimal model, which consider as the main problem of this study. Moreover, the improper selection of a diagnosis model for COVID-19 may be costly for medical organizations especially at meanwhile time where there is a great necessity for more accurate and fast diagnosis model. The methodology allows them to evaluate and benchmark a variety of COVID-19 diagnostic models to decide on selecting the model that meets the health organization’s needs and save the time and cost by adopting a reliable methodology for ML models selection. Furthermore, our methodology can be applied for benchmarking diagnosis models based on CT images, which can practically help the administration departments of medical organizations to select the most optimal COVID-19 diagnosis model. Finally, our study may draw a new line in the evaluation and benchmark of the ML models for COIVID-19.

References

[1] Abd Ghani, M.K., Mohammed, M.A., Arunkumar, N., Mostafa, S.A., Ibrahim, D.A., Abdullah, M.K., Jaber, M.M., Abdulhay, E., Ramirez-Gonzalez, G. and Burhanuddin, M.A., 2020. Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. *Neural Computing and Applications*, *32*(3), pp.625-638.

[2] Mohammed, M.A., Al-Khateeb, B., Rashid, A.N., Ibrahim, D.A., Ghani, M.K.A. and Mostafa, S.A., 2018. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. Computers & Electrical Engineering, 70, pp.871-882.

[3] Wang, L. and Wong, A., 2020. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images. arXiv preprint arXiv:2003.09871.

[4] Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P.D., Zhang, H., Ji, W., Bernheim, A. and Siegel, E., 2020. Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. arXiv preprint arXiv:2003.05037.

[5] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q. and Cao, K., 2020. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Radiology, p.200905.

[6] MV Villarejo, BG Zapirain, AM Zorrilla.,2013. Algorithms based on CWT and classifiers to control cardiac alterations and stress using an ECG and a SCR. Sensors 13 (5), 6141-6170.

[7] Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Chen, Y., Su, J., Lang, G. and Li, Y., 2020. Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv preprint arXiv:2002.09334.

[8] Chen, J., Wu, L., Zhang, J., Zhang, L., Gong, D., Zhao, Y., Hu, S., Wang, Y., Hu, X., Zheng, B. and Zhang, K., 2020. Deep learning-based model for detecting 2019

novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. medRxiv.

[9] Yan, L., Zhang, H.T., Xiao, Y., Wang, M., Sun, C., Liang, J., Li, S., Zhang, M., Guo, Y., Xiao, Y. and Tang, X., 2020. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv.

[10] Mohammed, M.A., Ghani, M.K.A., Arunkumar, N.A., Hamed, R.I., Abdullah, M.K. and Burhanuddin, M.A., 2018. A real time computer aided object detection of nasopharyngeal carcinoma using genetic algorithm and artificial neural network based on Haar feature fear. Future Generation Computer Systems, 89, pp.539-547.

[11] Garcia-Chimeno, Y., Garcia-Zapirain, B., Gomez-Beldarrain, M. et al., 2017. Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data. BMC Med Inform Decis Mak 17, 38.

[12] Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G. and Tan, W., 2020. Detection of SARS-CoV-2 in different types of clinical specimens. Jama.

[13] Ng, M.Y., Lee, E.Y., Yang, J., Yang, F., Li, X., Wang, H., Lui, M.M.S., Lo, C.S.Y., Leung, B., Khong, P.L. and Hui, C.K.M., 2020. Imaging profile of the COVID-19 infection: radiologic findings and literature review. Radiology: Cardiothoracic Imaging, 2(1), p.e200034.

[14] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X. and Cheng, Z., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 395(10223), pp.497-506.

[15] Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z. and Xia, L., 2020. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology, p.200642.

[16] Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P. and Ji, W., 2020. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology, p.200432.

[17] Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W. and Wang, X., 2020. Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label. medRxiv.

[18] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X. and Xu, B., 2020. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). medRxiv.

[19] Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Chen, Y., Su, J., Lang, G. and Li, Y., 2020. Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv preprint arXiv:2002.09334.

[20] R. L. Keeney and H. Raiffa, Decisions with multiple objectives: preferences and value trade-offs. Cambridge university press, 1993.

[21] S. Petrovic-Lazarevic and A. Abraham, "Hybrid fuzzy-linear programming approach for multi criteria decision making problems," arXiv preprint cs/0405019, 2004.

[22] J. Malczewski, GIS and multicriteria decision analysis. John Wiley & Sons, 1999.

[23] Xu, C., Wu, Y. and Dai, S., 2020. What are the critical barriers to the development of hydrogen refueling stations in China? A modified fuzzy DEMATEL approach. *Energy Policy*, *142*, p.111495.

[24] S. Zionts, "MCDM-If not a Roman Numeral, then what?," Interfaces, vol. 9, no. 4, pp. 94-101, 1979.

[25] Baltussen, R. and Niessen, L., 2006. Priority setting of health interventions: the need for multi-criteria decision analysis. Cost effectiveness and resource allocation, 4(1), p.14.

[26] Thokala, P., Devlin, N., Marsh, K., Baltussen, R., Boysen, M., Kalo, Z., Longrenn, T., Mussen, F., Peacock, S., Watkins, J. and Ijzerman, M., 2016. Multiple criteria decision analysis for health care decision making—an introduction: report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value in health, 19(1), pp.1-13.

[27] M. Oliveira, D. B. Fontes, and T. Pereira, "Multicriteria decision making: a case study in the automobile industry," Annals of Management Science, vol. 3, no. 1, p. 109, 2014.

[28] Jadhav, A. and Sonar, R., 2009, December. Analytic hierarchy process (AHP), weighted scoring method (WSM), and hybrid knowledge based system (HKBS) for software selection: a comparative study. In 2009 Second International Conference on Emerging Trends in Engineering & Technology (pp. 991-997). IEEE.

[29] Zaidan, A.A., Karim, H.A., Ahmad, N.N., Zaidan, B.B. and Kiah, M.M., 2015. Robust pornography classification solving the image size variation problem based on multi-agent learning. Journal of Circuits, Systems and Computers, 24(02), p.1550023.

[30] Nedher, A.S., Hassan, S. and Katuk, N., 2014. On multi attribute decision making methods: Prioritizing information security controls. Journal of Applied Sciences, 14(16), pp.1865-1870.

[31] Whaiduzzaman, M., Gani, A., Anuar, N.B., Shiraz, M., Haque, M.N. and Haque, I.T., 2014. Cloud service selection using multicriteria decision analysis. The Scientific World Journal, 2014.

[32] Yas, Q.M., Zaidan, A.A., Zaidan, B.B., Rahmatullah, B. and Karim, H.A., 2018. Comprehensive insights into evaluation and benchmarking of real-time skin detectors: Review, open issues & challenges, and recommended solutions. Measurement, 114, pp.243-260.

[33] Abdulkareem, K.H 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,2020.

[34] Aruldoss, M., Lakshmi, T.M. and Venkatesan, V.P., 2013. A survey on multi criteria decision making methods and its applications. American Journal of Information Systems, 1(1), pp.31-43.

[35] Opricovic, S. and Tzeng, G.H., 2004. Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European journal of operational research, 156(2), pp.445-455.

[36] Guo, S. and Zhao, H., 2017. Fuzzy best-worst multi-criteria decision-making method and its applications. Knowledge-Based Systems, 121, pp.23-31.

[37] Rezaei, J., 2015. Best-worst multi-criteria decision-making method. Omega, 53, pp.49-57.

[38] Zaidan, A.A., Zaidan, B.B., Albahri, O.S., Alsalem, M.A., Albahri, A.S., Yas, Q.M. and Hashim, M., 2018. A review on smartphone skin cancer diagnosis apps in evaluation and benchmarking: coherent taxonomy, open issues and recommendation pathway solution. Health and Technology, 8(4), pp.223-238.

[39] Salman, O.H., Zaidan, A.A., Zaidan, B.B., Naserkalid and Hashim, M., 2017. Novel methodology for triage and prioritizing using “big data” patients with chronic heart diseases through telemedicine environmental. International Journal of Information Technology & Decision Making, 16(05), pp.1211-1245.

[40] Jumaah, F.M., Zadain, A.A., Zaidan, B.B., Hamzah, A.K. and Bahbibi, R., 2018. Decision-making solution based multi-measurement design parameter for optimization of GPS receiver tracking channels in static and dynamic real-time positioning multipath environment. Measurement, 118, pp.83-95.

[41] Zaidan, A.A., Zaidan, B.B., Al-Haiqi, A., Kiah, M.L.M., Hussain, M. and Abdulnabi, M., 2015. Evaluation and selection of open-source EMR software packages based on integrated AHP and TOPSIS. Journal of biomedical informatics, 53, pp.390-404.

[42] Abdullateef, B.N., Elias, N.F., Mohamed, H., Zaidan, A.A. and Zaidan, B.B., 2016. An evaluation and selection problems of OSS-LMS packages. SpringerPlus, 5(1), p.248.

[43] Wu, Y., Xu, C., Zhang, B., Tao, Y., Li, X., Chu, H. and Liu, F., 2019. Sustainability performance assessment of wind power coupling hydrogen storage projects using a hybrid evaluation technique based on interval type-2 fuzzy set. Energy, 179, pp.1176-1190.

[44] Wu, Y., Xu, C. and Zhang, T., 2018. Evaluation of renewable power sources using a fuzzy MCDM based on cumulative prospect theory: A case in China. Energy, 147, pp.1227-1239.

[45] Kaya, İ., Çolak, M. and Terzi, F., 2018. Use of MCDM techniques for energy policy and decision-making problems: A review. International Journal of Energy Research, 42(7), pp.2344-2372.

[46] Mostafa, S.A., Mustapha, A., Khaleefah, S.H., Ahmad, M.S. and Mohammed, M.A., 2018, February. Evaluating the performance of three classification methods in diagnosis of Parkinson’s disease. In International Conference on Soft Computing and Data Mining (pp. 43-52). Springer, Cham.

[47] Obaid, O.I., Mohammed, M.A., Ghani, M.K.A., Mostafa, A. and Taha, F., 2018. Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer. International Journal of Engineering & Technology, 7(4.36), pp.160-166.

[48] Abdulkareem, K.H., Mohammed, M.A., Gunasekaran, S.S., Al-Mhiqani, M.N., Mutlag, A.A., Mostafa, S.A., Ali, N.S. and Ibrahim, D.A., 2019. A Review of Fog Computing and Machine Learning: Concepts, Applications, Challenges, and Open Issues. IEEE Access, 7, pp.153123-153140.

[49] Mohammed, M.A., Ghani, M.K.A., Hamed, R.I., Mostafa, S.A., Ibrahim, D.A., Jameel, H.K. and Alallah, A.H., 2017. Solving vehicle routing problem by using improved K-nearest neighbor algorithm for best solution. Journal of Computational Science, 21, pp.232-240.

[50] Klein, S., Pluim, J.P., Staring, M. and Viergever, M.A., 2009. Adaptive stochastic gradient descent optimisation for image registration. International journal of computer vision, 81(3), p.227.

[51] Zhang, Y., Kwon, D. and Pohl, K.M., 2017. Computing group cardinality constraint solutions for logistic regression problems. Medical image analysis, 35, pp.58-69.

[52] Ko, B.C., Kim, S.H. and Nam, J.Y., 2011. X-ray image classification using random forests with local wavelet-based CS-local binary patterns. Journal of digital imaging, 24(6), pp.1141-1151.

[53] Criminisi, A., 2016. Machine learning for medical images analysis.

[54] https://github.com/ieee8023/covid-chestxray-dataset.

[55] Chhikara, P., Singh, P., Gupta, P. and Bhatia, T., 2020. Deep Convolutional Neural Network with Transfer Learning for Detecting Pneumonia on Chest X-Rays. In Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals (pp. 155-168). Springer, Singapore.

[56] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[57] Shannon, C.E., 1948. A mathematical theory of communication. Bell system technical journal, 27(3), pp.379-423.

[58] Hsu, P.F. and Hsu, M.G., 2008. Optimizing the information outsourcing practices of primary care medical organizations using entropy and TOPSIS. Quality & Quantity, 42(2), pp.181-201.

[59] Tzeng, G.H. and Huang, J.J., 2011. Multiple attribute decision making: methods and applications. CRC press.

[60] Hsu, P.F. and Hsu, M.G., 2008. Optimizing the information outsourcing practices of primary care medical organizations using entropy and TOPSIS. Quality & Quantity, 42(2), pp.181-201.

[61] Yuan, B.J. and Chen, J.K., 2007. Adopting information systems for China enterprise investing using the Grey Relation Analysis. International Journal of Value Chain Management, 1(4), pp.416-438.

[62] Deng, H., Yeh, C.H. and Willis, R.J., 2000. Inter-company comparison using modified TOPSIS with objective weights. Computers & Operations Research, 27(10), pp.963-973.

[63] Mostafa, S.A., Mustapha, A., Mohammed, M.A., Hamed, R.I., Arunkumar, N., Ghani, M.K.A., Jaber, M.M. and Khaleefah, S.H., 2019. Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson’s disease. Cognitive Systems Research, 54, pp.90-99.

[64] Pirouz, B., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S. and Piro, P., 2020. Investigating a Serious Challenge in the Sustainable Development Process: Analysis of Confirmed cases of COVID-19 (New Type of Coronavirus) Through a Binary Classification Using Artificial Intelligence and Regression Analysis. Sustainability, 12(6), p.2427.

[65] Kassani, S.H., Kassasni, P.H., Wesolowski, M.J., Schneider, K.A. and Deters, R., 2020. Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning-Based Approach. *arXiv preprint arXiv:2004.10641*.