Mustansiriyah University, College of Engineering

https://doi.org/10.31272/conf.CESC.14

The 2nd International Conference of Engineering Science (CESC)

May 2024

Predicting Correct Medical Procedure for COVID-19 Patients Using Machine Learning

Hiba Alyaqoobi [1], Mohammad Azzeh [2], Methaq Kadhum[3*], Mohammad Hijjawi[4] and Seveem Omran[5]

[1] Department of Computer Science, Applied Science Private University, Amman, Jordan
[2]Department of Data Science, Princess Sumaya University for Technology, Amman, Jordan
[3]Department of College of Engineering, Mustansiriyah University, Baghdad, Iraq
[4] Department of Computer Science, Applied Science Private University, Amman, Jordan
[5] Alkarkh Hospital, Iraqi Ministry of Health, Baghdad, Iraq

Abstract: Since the beginning of the COVID-19 (Coronavirus Disease 2019) global pandemic, patient cases have continuously increased. To avoid the spread of COVID-19 at that time the policymakers recommend different public health actions such as lockdown and social distancing, which led to significant re-ductions in all life aspects. Health officials warn that designated treatment centers are also overwhelmed with COVID-19 patients. This has increased the pressure on hospitals and all kinds of healthcare system resources. There is a statistical increase before and after the pandemic about the increasing comfort of many people for telemedicine and there is a clear increase in the use of telemedicine, especially after the pandemic Therefore, there is a desperate need to reduce such pressure and avoid extra costs on patients and hospitals in the future, if a similar disease appears this study can be used to reduce the risks. This paper provides a solution using machine learning to identify the correct procedure for COVID-19 patients when they visit hospitals. This solution is supposed to reduce the crowdedness in hospitals, and the cost of treatments and relieve pressure on healthcare systems, especially human resources. The proposed system has been constructed and evalu-ated based on a dataset collected from hospitals in Iraq. We investigated the importance of the selected features on the accuracy of the proposed system. The results revealed that the proposed system is predictive and offers im-provements in terms of quick procedure prediction for COVID-19 patients who visited the hospital. We also found that using a few features can harm the proposed prediction models using the top 6 features would seem more effi-cient and accurate. However, since the collected dataset is from one country, further experiments on the proposed system are needed on global datasets to ensure its performance.

Keywords: COVID-19, Information Gain, Machine Learning, Naïve Bayes, Support Vector Machine

1. Introduction

This At the onset of the pandemic in order to slow down the infection transmission of COVID-19, the governments have considered numerous public health actions such as social distancing, lockdown, and closing of facilities, etc. Along with all these pre-cautions, health officials warned that designated medical treatment centers would be overwhelmed with COVID-19 patients (Iwendi *et al.*, 2020)(Fayyoumi, Idwan and Aboshindi, 2020). Increasing the number of infected people who require medical attention/Hospitalization has led to growing global concerns about delays in diagnosis and taking appropriate medical responses at the right time. Thus, leading to inefficient utilization of hospital resources and causing various health problems for medical specialists (Abdulah and Mohammed, 2020)(Naudé, 2020). The research conducted by Abdulah & Mohammad (Abdulah and Mohammed, 2020) concluded that the effect

*Corresponding author email: methaq_kadhum@yahoo.com

of COVID-19 on the mental state of medical specialists had been shown to include frustration, stress, insomnia, anxiety, fear, and symptoms of depression. Mental condition problems affect the higher cognitive processes, comprehension, and focus of healthcare workers and affect the general well-being (Abdulah and Mohammed, 2020). Predictors of job satisfaction and exhaustion in a long-term study of nurses and other healthcare professionals during the COVID-19 pandemic where the most powerful indicators of burnout and satisfaction were symptoms of depression and organizational structure(Guastello *et al.*, 2024). Therefore, they urgently need to reduce the additional costs, whether on the patient or the hospital, to reduce the pressure and help medic al specialists make clinical decisions in a short time and with limited information. Re-searchers in various countries have used many prediction models to develop appropriate judgments and laws that will aid in the sup-pression of the COVID-19 pandemic.

Therefore, medical specialties, especially Emergency Room (ER), need a system to help them predict the correct procedure within the clinical decision-making process with limited time and limited information without directly touching the patient for fear of acquiring the infection. Naudé et al. (Naudé, 2020)discussed six areas where machine learning can contribute to the fight against COVID-19, such as alerts and early warnings in prediction, monitoring, diagnosis and prognosis, community intervention, data dashboards, and treatments. Many researchers try to help in decreasing the spread of COVID-19 by proposing a technique to predict the infected patients from non-infected ones (Fayyoumi, Idwan and Aboshindi, 2020)(Arpaci *et al.*, 2021). Such techniques are helpful to reduce the spread of COVID-19, but they could not prevent the spread of this disease(Nguyen *et al.*, 2020). Thus, the main problem in hospitals at that time is that they have a problem identifying the correct procedure for each patient when s/he visited the hospital. Usually, when a patient visits the hospital with a confirmed COVID-19 virus, doctors ask for symptoms, an x-ray, and some blood tests to know the best procedure for this patient. The standard procedure for COVID-19 patients is as follows: 1) isolation in hospital, 2) isolation in-home, or 3) respiratory care unit (RCU), as seen in Fig.1. Doctors often decide on the procedure in the ER based on x-ray and some blood tests. This process may take a long time until the results are revealed, consequently increasing crowdedness and social mixing.

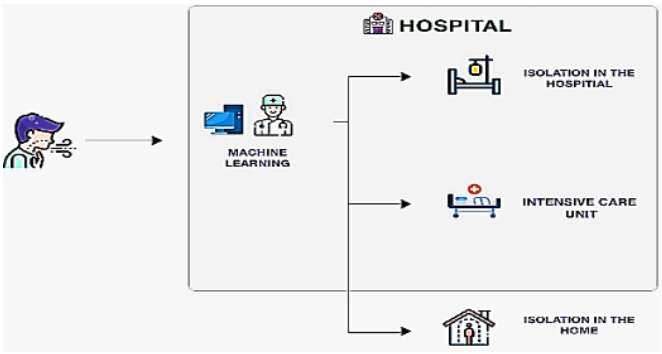


Figure 1. The Patient Care Pathway (In Hospital) During COVID-19 Pandemic

The proposed solution is designed to help doctors identify the correct procedure or actions for COVID-19 patients when visiting hospitals using machine learning. This area of research is still yet unexplored. The system is supposed to work on the limited information provided by patients to produce a quick but reliable decision. The dataset used in this study was collected from the Public Health Unit of the Iraqi Ministry of Health /Al-Karkh. The data structure was determined based on a discussion with the doctor in charge of the Public Health Unit. Around 900 patients record has been collected to build various machine

learning models. We also investigate the importance of designated features on the accuracy of the constructed machine learning models by using Information Gain (IG) algorithm(Karegowda, Manjunath and Jayaram, 2010). Then one can select the threshold for which features must be involved during the learning and testing process. The purpose of this investigation is to see which and how many features can work well for predicting the correct procedure. All results have been statistically tested to ensure the significance of our proposed system.

Above all, besides building a decision-making system that can predict the correct procedure using machine learning, this study aims to answer the following research questions:

RQ1. Is there evidence that machine learning can support procedure prediction when using shallow data descriptions that include the main COVID-19 symptoms?

RQ2. Do the designated features have a significant impact on the accuracy of the constructed prediction models?

The remaining of this paper is organized as follows: The related work is presented in section 2. Section 3 introduces the collected dataset. Section 4 presents the ma-chine learning algorithms that are used in this study. The research methodology is presented in Section 5, while the experimental results and discussion are reported in Section 6. The threats to validity are discussed in section 7. Finally, Section 8 draws the conclusion and future works.

2. Related Work

In 2020, several researchers proposed a study of Machine learning applications and prediction models for the COVID-19 pandemic. Malki et al.(Malki *et al.*, 2020) used linear models and machine learning algorithms to analyze the association between weather variables and the distribution of COVID-19 cases. The data for COVID-19 cases is from the Johns Hopkins Center and Kaggle. Essential features for forecasting the COVID-19 death rate are the weather data from the historical database that hit temperature and humidity. The findings showed that the higher the temperature value, the lower the number of infections. S. Debnath et al. (Debnath *et al.*, 2020) discuss patient risk stratification due to lower and higher risk of impairment. For 5700 hospitalized COVID-19 patients, the data was from the New York City area. (Zhu et al., 2024)discuss after the COVID-19 pandemic, tele mental health practitioners use telecommunication on a daily basis, and they plan to keep using it in their practices after the pandemic. This assumption is more common among some provider categories. The outcome for the hospital includes resource allocation, care prioritization, and prognosis estimation using the ML/AI model used. Kavadi et al.(Kavadi *et al.*, 2020) discuss a partial derivative regression and nonlinear machine learning (PDR-NML) approach to predict COVID-19 cases from the Indian population, which found that the ML method outperformed excellent methods in the Indian population. E. Fayyoumi. [4] use ML and statistical modeling to predict COVID-19 patients by collecting 105 samples for patient symptoms by online questionnaire. The data was used as an input for various prediction models based on Logistic Regression and ML model (SVM and Multi-Layer Perceptron). These models were utilized to predict potential patients of COVID-19 based on their symptoms. The MLP (Multi-Layer Perceptron) has shown the best accuracy (91.62%) compared to the other models. Meanwhile, the SVM has shown the best

precision (91.67%). A. K. Das(Das, Mishra and Gopalan, 2020) deploys the best performing ML algorithms to predict mortality among 3,524 confirmed COVID-19 patients in South Korea; they compare the performance result of five ML algorithms (SVM, logistic regression, K nearest neighbor, random forest, and gradient boosting). The logistic regression algorithm was the best performer (area under ROC curve = 0.830) deployed as the online COVID-19 Community Mortality Risk Prediction tool named CoCoMoRP. Also, many researchers analyze COVID-19 diagnosis by X-rays images dataset because it is conveniently available, quickly, and at low costs. M. Alazab. (Haritha, Praneeth and Pranathi, 2020)analyze the incidence of COVID-19 distribution for 1000 X-ray images based on a deep convolutional neural network (CNN), the prophet algorithm(PA), and autoregressive integrated moving average (ARIMA) model to detect COVID19 patients and it led to many results including F-measure range of 95–99% and the coastal areas are heavily impacted by the COVID-19 spread .S. Minaee.(Minaee *et al.*, 2020)apply deep learning models to detect COVID-19 patients from 5000 X-rays images, and as a result, the archived a sensitivity rate of 98% and a specificity rate of around 90%.Our primary focus is on researchers analyzing COVID-19 diagnosis dataset performance comparison of different machine learning algorithms and many diseases prediction. Finally, communication between ML data scientists and policymakers is necessary for the success of these distinguished endeavors to succeed in overcoming the crisis.

3. Collected Data and features

The data used in this study was collected from the Public Health Unit of the Iraqi Ministry of Health /Al-Karkh (from 3/5/2020 to 26/7/2020). The information was taken for the infected patients from many hospitals on the AL-Karkh side(Ali Abed, Hussein Ewaid and Al-Ansari, 2019) of Baghdad with their identities and addresses kept anonymous to avoid legal issues and not to inflict physical and emotional harm on the patients. A total of 915 patient records are in the dataset. The description of the collected dataset is shown in Table 1, which contains 14 features of symptoms as input. It is worth noting that Oxygen Saturation (SpO2) is a measurement of how much oxygen your blood is carrying as a percentage of the maximum it could carry. In COVID-19 patients, Oxygen saturation (SpO2) was at least at the upper limit of the recommended 92–96% range(Shenoy, Luchtel and Gulani, 2020). The possible class labels for procedure variables are 1) Isolation in the hospital, 2) Intensive Care Units (ICU), and Isolation in the home. As we know, the data must be processed before entering ML algorithms to ensure performance of algorithms. The dataset has been re-processed for machine learning algorithms, such as the records with missing values being deleted even though we can impute them. Still, we favor excluding them because it is not legal to change data related to patients. All categorical variables have been encoded using the One-Hot encoder algorithm.

Table 1. Description of the collected dataset

No.	Symptoms	Description
1	Fever	{yes: temp>37.5, no: otherwise}
2	Dry Cough	{yes, no}
3	Age	Continuous variable
4	Chronic Disease	{yes, no}
5	Nasal Congestion	{yes, no}
6	Hearing Problem	{yes, no}
7	Loss of Smell	{yes, no}
8	Diarrhoea	{yes, no}
9	Vomiting	{yes, no}
10	Body Ache Fatigue	{yes, no}
11	Sore Throat	{yes, no}
12	Travel History	{yes, no}
13	Oxygen Saturation (SpO_2)	Continuous variable
14	Breathing Problem	{yes, no}
15	Procedure	{Isolation in hospital, Isolation in home, ICU}

Below we analyze the relationship between the two most factors (Fever and SpO2) that are usually used to identify COVID-19 patients. In Fig.2 shows the bar plot of patients with fever clustered based upon the actual selected procedure (i.e., temperature>37.5); we can observe that above 50% (majority of them) are isolated at home. We also noted that all patients had fever symptoms during their visits to the hospital.

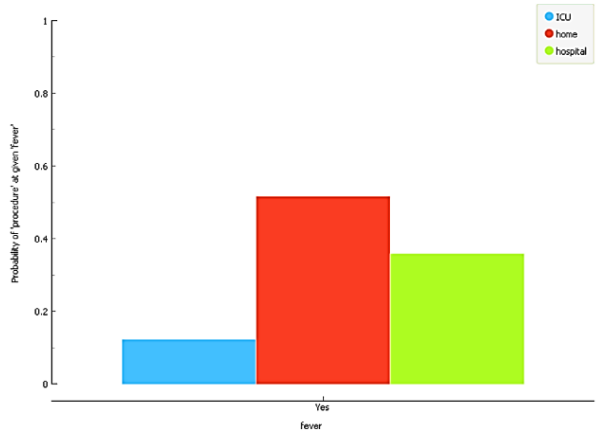


Figure 2. Bar plot for fever feature grouped by procedure label.

The Fig.3 shows the bar plot of patients with SpO2 clustered based upon the actual selected procedure (i.e., yes when SpO2<92 and no otherwise). We can see that most patients with no problem with breathing can stay isolated at home. Whereas most patients with breathing problems are isolated in the hospital, and some of them are isolated in the ICU. Just very few of them are isolated homes. This is a clear indication that the SpO2 variable plays an essential role in identifying COVID-19 patients.

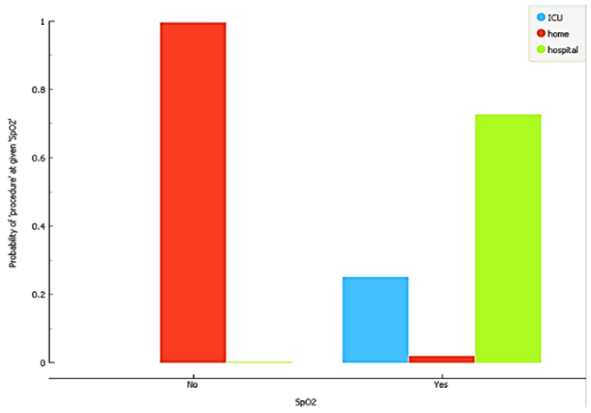


Figure 3. Bar plot for SpO2 feature grouped by procedure label.

The Fig.4 shows the bar plot of patients with loss of smell clustered based upon the actual selected procedure. We can notice that most of the patients with no problem with smell are isolated at home. In contrast, there is no clear pattern for patients with no smell problem. This factor cannot give a clear indication of a straight forward procedure.

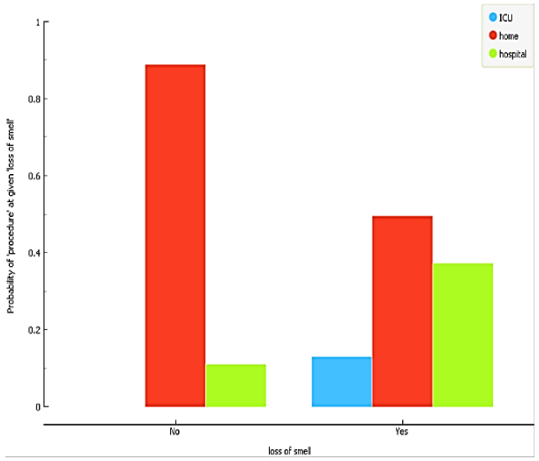


Figure 4. Bar plot for loss of smell feature grouped by procedure label.

The Fig.5 shows the bar plot of patients with sore throats clustered based upon the actual selected procedure. We can notice that most patients with no sore throat problems are isolated at home. In contrast, most patients with sore throat problems are isolated at the hospital, and some are isolated in the ICU.

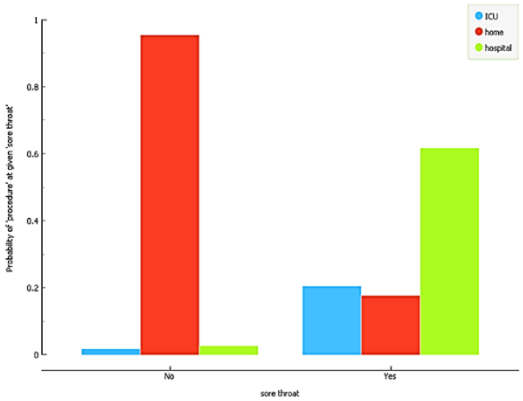


Figure 5. Bar plot for sore throat feature grouped by procedure label.

4. Algorithms Selection

Two machine learning algorithms have been used in this study: Support Vector Machine (SVM) and Naïve Bayes (NB) and one feature ranking algorithm, Information Gain. These algorithms have been selected as they have commonly used for many health-related problems and offer great accuracy improvement in their predictions. The problem studied in this paper is considered multi-label classification because the output variable is not continuous and contains three labels.

4.1 Support Vector Machine

The Support Vector Machine (SVM) is a state-of-the-art classification method that supports linear and nonlinear learning(Shenoy, Luchtel and Gulani, 2020). The SVM is particularly well suited for classifying complex, small-size data sets and even with outlier detection. The SVM mainly works by separating data points using an optimal hyperplane with a maximum margin the hyperplane boundary of decision between the two classes of data (Shenoy, Luchtel and Gulani, 2020)(Claesen *et al.*, 2014). For example, in fig. 6 the SVM tries to find the optimal hyperplane, the maximum distance between the linear model, and the "support points" near the decision boundary. If there are no points near the margin, the derived hyperplane will divide the labels perfectly, reducing the classification error, Also, the SVM can control the balance by hyperparameter for soft margin classification. It also helps in locating the hyperplane to avoid overfitting (variance) and underfitting (Bias) by allowing misclassification inside the soft margin (Shenoy, Luchtel and Gulani, 2020).

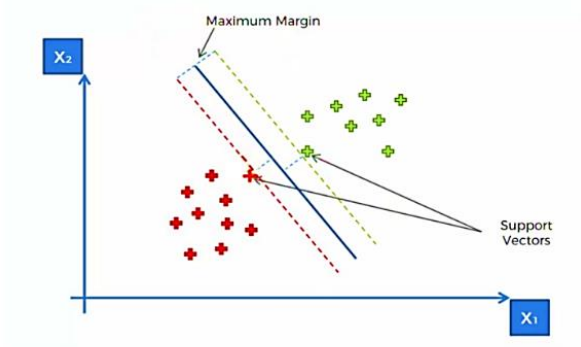


Figure 6. Support Vector Machine (SVM) Description Adapted from (Claesen *et al.*, 2014)

The optimal boundary between the outputs depends on the selected kernel function to systematically find the support vector classifier in high dimensions such as (Linear, Radial Basis Function (RBF), Gaussian, Polynomial, and Sigmoid). These kernels attempt to optimize the

separation boundaries between the data points by starting with data from a low dimension and moving the data into a higher dimension. The Radial Basis Function (RBF) kernel has been chosen for our experiment because it moves data from the lower dimension to the higher dimension based on the similarity between data points, as shown in Eq. (1)(Shenoy, Luchtel and Gulani, 2020).

$$k(x_i, x_j) = e^{-y\|x_i - x_j\|^2} \quad (1)$$

Where: x_i and x_j are two data point in the original space. y is a random number $y > 0$ *or* $y = \frac{1}{2\sigma^2}$

Our research's dependent variable (procedure) is multiclass classification (isolation at home, isolation at hospital, ICU). The SVM classification works with multiclass by decomposing the multiclassification problem into several binary classification problems. To classify data points from M classes, two breakdown approaches can be used:

1.One-to-One: The hyperplane is separated into two classes by ignoring the points in the third class. This means that only the points of the two classes in the present division are taken into account(Malki *et al.*, 2020).

2.One-to-Rest: The hyperplane separates a class and all others at the same time. So, the separation takes all points into account and divides them into two groups, the first group for the class points and the second group for all other points(Malki *et al.*, 2020).

4.2 Naïve Bayes Classifiers

Naïve Bayes classifiers are based on applying Bayes (probability theorem) as shown in Eq. (2). This classifier works under the assumption that all the features are completely independent of each other. It can work well when data has too many categorical features. An observation, X will be classified according to which C_i gives the highest $P(C_i|X)$for it(Claesen *et al.*, 2014).

$$P(C_i|X) = \frac{P(X|C_i).P(C_i)}{P(X)}, \quad P(X) \neq 0 \quad (2)$$

Where: C_i: i^{th}class, X: feature vector of independent variables.$P(C_i|X)$:the probability of X being of classC_i $P(X|C_i)$: the probability of getting X given that the class is C_i.

4.3 Information Gain

Information Gain is used mainly in constructing a decision tree. However, it can also select features by first ranking them based on entropy metrics. The features are ranked based on the entropy assessment between the input and output features. Usually, an expert must determine the threshold for which top features to construct prediction models. It is well known that using a few features could produce a good performance and all features. In this study, we gradually change the threshold from 2 to 10 for each threshold, we select the features placed above the threshold to construct the prediction model [16].

5. Research Methodology

This section describes the research methodology of this paper. First, the data has been processed to ensure the quality of data (In Fig.7). Although there are different techniques for handling missing values, we favor deleting all instances with missing values because the number of missing records is very few. On the other hand, we apply the one-hot encoder technique to convert categorical variables to numeric variables because SVM can work with numeric values. Second, the data was split into training and testing data using repeated 10-folds cross-validation to divide the dataset into ten groups. In

each iteration, one group is holdout as test data set and the remaining groups as a training dataset. Then a model is fitted to the training dataset and evaluated on the test datasets. We have run this experiment 10 times for each model, and the accuracy values were recorded in each iteration to evaluate the model's accuracy. This procedure is repeated ten times to avoid any potential bias in selecting training data. Both SVM and NB methods have been used to develop pre-diction models. The following configuration parameters were applied to the SVM with cost=1.00, the kernel type selected as RBF separates a class with approach (One-to-Rest), and the optimization parameter to numerical tolerance equal to 0.0010 with 100 iteration limits. In contrast, there are no configuration parameters for the Naïve Bayes classifiers model.

To examine the impact of features on the constructed models, we use Information Gain to rank features using entropy measure. Then, each time we select a feature selection threshold starting from the top 2 to top 10 features. For each feature subset, we re-run the above procedure, and we record the evaluation results. This enables us to investigate the importance of our designated features on early decision-making accuracy for COVID-19 patients. These features in hand are supposed to provide quick decisions, but we do not know which combination of features would enhance the accuracy of the prediction. Furthermore, we use the Mann-Whitney test to investigate the significant difference between prediction models.

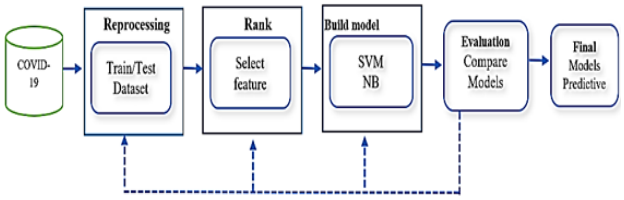


Figure. Machine learning model construction and validation process

As mentioned above, multiple evaluation measures have been adopted to evaluate the performance of the constructed prediction models. These measures have been selected carefully to reflect the predictive performance of each model [17]. The first measure is the Accuracy metric which evaluates the proportion of correctly identified examples to all in the dataset, as shown in Eq.3(Nassif *et al.*, 2018). This accuracy is useful when there is no class imbalance problem, as in our case. The following measure is Recall which evaluates the proportion of predicted positive examples to all actual positive examples as shown in Eq.4[17]. Precision measure evaluates the percentage of correctly predicted positive examples to all positive predicted examples, as shown in Eq.5(Yekkehkhany *et al.*, 2014). F1 measure makes compromises between Recall and Precision, as shown in Eq.6[17]. Finally, the Specificity measure evaluates the percentage of correctly identified negative examples to all actual negative examples, as shown in Eq.7. Finally, Area Under Curve (AUC) measures the area under Operating Curve [18]. The AUC measure will compare SVM and NB concerning the stability of top k feature subsets.

The above-mentioned evaluation measures are supposed to work well with binary classifiers. Still, since our problem is multi-label classification, we used a weighted average to aggregate Recall and precision for all labels, which considers the number of actual examples that belong to each label. To accomplish that, we computed Re-call and precision for each class label then we used a weighted average to aggregate the final recall value.

$$Accuracy = \frac{TN + TP}{TN + FN + TP + FP} \quad (3)$$

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

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

Where: TP is True Positive, TN is True Negative, FN is False Negative and FP is False Positive.

6. Experimental Results and Discussion

This section presents and comments on the obtained results. Since this study examines the performance of machine learning algorithms for predicting correct procedures for COVID-19 patients, it is also essential to investigate the importance of collected features on the accuracy of this prediction model. Therefore, we first consulate the Information Gain algorithm to rank collected features based on their strong relationship with the target variable, as explained in section 5. Table 2 shows the importance of the features sorted by the Information Gain ratio. From the table, we can notice that SpO2 is the top-ranked feature which is the indicator for detecting the correct procedure that must be monitored for the patient.

In contrast, the least helpful feature was the fever variable. This finding is non-trivial because SpO2 is a strong sign of COVID-19 disease, while fever is common for many diseases. However, low SpO2 can also be noticed for other respiratory diseases, but it is usually common for COVID-19 patients. Therefore, this feature helps determine the severity of patient situations and could improve procedure decision-making accuracy. Surprisingly, the presence or absence of other chronic diseases offers another vital sign for making the correct decision for the correct procedure. The other signs essential for diagnosing COVID-19 diseases, such as Diarrhea and Vomiting symptoms, are less critical for determining the correct procedure. Nevertheless, the importance of these features can be accurately determined after constructing the prediction models. In this study, we use ranked features from table 2 to make different feature subsets used to construct various prediction models. In other words, in every experiment, we use top k features to build a procedure prediction model. For example, when we say Top2, that means we select SpO2 and Chronic Diseases features. To investigate different thresholds, we consider k from 2 to 10 and use all 14 input features.

Table 2. Ranks of Features based on Information Gain Ratio.

Rank	Features	Information Gain ratio
1	SpO_2	0.935
2	Chronic Diseases	0.607
3	Body Ache Fatigue	0.515
4	Sore Throat	0.510
5	Age	0.281
6	Vomiting	0.211
7	Diarrhea	0.201
8	Nasal Congestion	0.101
9	Loss of Smell	0.094
10	Hearing Problem	0.078
11	Breathing Problem	0.024
12	Dry Cough	0.023
13	Travel History	0.019
14	Fever	0.000

For each feature subset, we construct two prediction models using SVM and NB; then, we validate them using six evaluation measures. To present these results in detail, we first record the evaluation results for each type of procedure (i.e., class label) as shown in Tables 3, 4, and 5. Then we present aggregate results, which is the weighted average of all class labels results as shown in Table 6. Table 4 shows the evaluation results for the ICU procedure label. The results of both SVM and NB are shown in terms of every evaluation measure. The horizontal line represents the Top k features that are employed to construct that prediction model. The text in bold represents the best accuracy values between SVM and NB in terms of annotated evaluation measures. The first observation is that the Accuracy and AUC results for both SVM and NB are suitable for predicting ICU labels. The second observation is that there is no superiority consistent between SVM and NB and all feature subsets. The accuracy values for both models along all feature subsets are relatively close with no clear preference.

In contrast, the recall values, which represent the percentage of observations correctly predicted as ICU, are small for low and high dimensionality and acceptable for the top 5 and 6 sets. The same trend can be noticed for precision results. To confirm that, we draw a relationship between AUC

values and top k features subsets. We chose AUC because it depicts relative trade-offs between Recall and false-positive classification rates for every possible threshold. Therefore, using all features with SVM and NB works well and produces an acceptable AUC value. The Mann-Whitney significant test between SVM and NB in terms of AUC (p-value= 0.87) shows no significant difference between them, confirming that both models can produce relatively similar predictions along with all feature subsets.

Table 3. Evaluation results for ICU Label

	Accuracy		Recall		Precision		F1		Specificity		AUC	
	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB
Top2	**0.86**	0.85	0.10	**0.13**	0.29	**0.30**	0.14	**0.18**	**0.97**	0.96	0.81	**0.82**
Top3	**0.86**	0.84	**0.23**	0.15	**0.39**	0.26	**0.29**	0.19	**0.95**	0.94	0.79	**0.83**
Top4	0.84	**0.86**	0.26	0.26	0.31	**0.39**	0.28	**0.31**	0.92	**0.94**	0.79	**0.82**
Top5	0.88	**0.89**	**0.48**	0.40	0.50	**0.56**	**0.49**	0.47	0.93	**0.95**	0.86	**0.87**
Top6	0.88	0.88	**0.52**	0.40	0.50	0.50	**0.51**	0.45	0.93	**0.94**	0.87	**0.88**
Top7	0.86	**0.89**	**0.35**	0.32	0.44	**0.59**	0.39	**0.42**	0.94	**0.97**	**0.88**	0.86
Top8	0.86	**0.87**	**0.39**	0.32	0.44	**0.47**	**0.41**	0.38	0.93	**0.95**	**0.88**	0.86
Top9	**0.89**	0.88	**0.35**	0.29	**0.59**	0.53	**0.44**	0.38	**0.97**	0.96	**0.87**	0.85
Top10	**0.87**	0.86	**0.35**	0.27	**0.46**	0.43	**0.40**	0.33	0.94	**0.95**	**0.87**	0.84
All	0.88	**0.90**	**0.35**	0.31	0.51	**0.70**	0.42	**0.43**	0.95	**0.98**	**0.87**	0.86

Table 4 shows the evaluation results for the home procedure label. We can notice that accurate results for both SVM and NB are high, confirming the constructed prediction models' ability to predict isolated at-home procedures. Unlike ICU procedure, the Recall and precision are almost high along with all feature subsets. However, the AUC of SVM and NB for predicting home labels are relatively similar to those obtained for the ICU label. The Mann-Whitney Rank significant test between SVM and NB in terms of AUC (p-value= 0.88) shows no significant difference be-tween them, confirming that both models can produce relatively similar predictions along with all feature subsets.

Table 4. Evaluation results for home Label

	Accuracy		Recall		Precision		F1		Specificity		AUC	
	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB
Top2	0.93	0.93	**0.91**	0.89	0.95	**0.97**	0.93	0.93	0.95	**0.97**	0.80	**0.81**
Top3	**0.91**	0.90	**0.87**	0.85	0.96	0.96	**0.91**	0.90	0.96	0.96	0.81	0.81
Top4	**0.88**	0.87	**0.81**	0.77	0.95	**0.96**	**0.88**	0.86	0.95	**0.97**	0.88	0.88
Top5	0.88	0.88	0.82	**0.83**	**0.95**	0.93	**0.88**	0.87	**0.95**	0.93	0.85	0.85
Top6	0.87	**0.91**	0.81	**0.89**	**0.94**	0.93	0.87	**0.91**	**0.94**	0.93	0.91	**0.93**
Top7	0.89	**0.92**	0.83	**0.88**	0.96	0.96	0.89	**0.92**	0.96	0.96	**0.89**	0.80
Top8	0.86	**0.90**	0.77	**0.85**	0.95	0.95	0.85	**0.90**	0.95	0.95	**0.87**	0.85
Top9	0.92	**0.94**	0.89	**0.93**	0.95	0.95	0.92	**0.94**	0.95	0.95	0.81	**0.92**
Top10	0.90	**0.92**	0.89	**0.92**	0.91	**0.93**	0.90	**0.92**	0.90	**0.93**	0.88	**0.94**
All	0.91	**0.97**	0.87	**0.98**	0.95	**0.96**	0.91	**0.97**	0.95	0.95	0.88	**0.93**

Results for hospital labels are less accurate than that of label home, as shown in Table 5. Specifically, we can notice that the accuracy for SVM and NB along all feature subsets is around 0.85-0.87. Precision is also less accurate along with all feature subsets. These results revealed that hospital label is not firmly related to feature subsets as for home label. The Mann-Whitney significant test between SVM and NB in terms of AUC shows no significant difference between them, confirming that both models can produce relatively similar predictions along with all feature subsets.

Table 5. Evaluation results for hospital Label

	Accuracy		Recall		Precision		F1		Specificity		AUC	
	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB
Top2	0.86	**0.87**	0.94	**0.97**	0.74	0.74	0.83	**0.84**	0.81	0.81	**0.90**	0.86
Top3	**0.86**	0.84	0.94	0.94	**0.74**	0.71	**0.83**	0.81	**0.81**	0.79	**0.87**	0.82
Top4	**0.85**	0.82	0.93	**0.94**	**0.73**	0.67	**0.82**	0.79	**0.81**	0.74	0.87	**0.91**
Top5	**0.87**	0.83	0.92	**0.89**	**0.76**	0.71	**0.83**	0.79	**0.84**	0.80	**0.89**	0.86
Top6	0.87	**0.88**	**0.92**	0.90	0.77	**0.80**	0.84	**0.85**	0.85	**0.87**	**0.93**	0.91
Top7	**0.86**	0.85	0.93	0.93	**0.74**	0.73	**0.83**	0.82	**0.82**	0.81	0.88	**0.92**
Top8	0.85	**0.87**	0.94	**0.95**	0.72	**0.75**	0.82	**0.84**	0.80	**0.82**	**0.94**	0.91
Top9	0.87	**0.88**	**0.94**	0.93	0.76	**0.78**	0.84	**0.85**	0.83	**0.85**	0.84	**0.88**
Top10	0.84	**0.87**	0.83	**0.89**	0.75	**0.77**	0.79	**0.83**	0.85	0.85	**0.87**	0.85
All	0.87	**0.89**	0.93	0.93	0.76	**0.80**	0.84	**0.86**	0.83	**0.87**	0.83	**0.91**

Table 6 shows the aggregated evaluation results for three procedure labels. The results of Recall and precision were aggregated using a weighted average. The aggregated results confirm that both prediction models are capable of predicting three labels accurately. Studying the relationship between feature subsets and evaluation results shows no clear, stable pattern, so we could not identify a definite accurate model among SVM and NB or feature subsets. In addition, the margin between SVM and NB and all feature subsets concerning evaluation results is trivial and negligible. Finally, The Mann-Whitney Rank significant test between SVM and NB in terms of AUC shows no significant difference between them, confirming that both models can produce relatively similar predictions along with all features subsets.

Table 6. Evaluation results for all Labels

	Accuracy		Recall		Precision		F1		Specificity		AUC	
	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB
Top2	0.89	**0.90**	0.82	0.82	0.79	**0.80**	0.80	0.80	0.90	**0.91**	0.86	0.86
Top3	**0.89**	0.87	**0.82**	0.80	**0.81**	0.78	**0.80**	0.78	**0.91**	0.90	0.84	**0.86**
Top4	**0.87**	0.85	**0.79**	0.77	0.79	0.79	**0.78**	0.76	**0.90**	0.88	0.82	**0.88**
Top5	**0.88**	0.86	**0.81**	0.80	**0.82**	0.80	**0.81**	0.79	**0.91**	0.89	0.87	**0.90**
Top6	0.87	**0.89**	0.81	**0.83**	0.83	0.83	0.81	**0.83**	0.91	0.91	**0.93**	0.89
Top7	0.88	**0.89**	0.81	**0.83**	0.81	**0.83**	0.80	**0.82**	0.90	**0.91**	**0.91**	0.89
Top8	0.86	**0.88**	0.79	**0.82**	0.80	**0.82**	0.79	**0.81**	0.89	**0.91**	0.81	**0.89**
Top9	0.90	**0.91**	0.84	**0.85**	0.84	0.84	0.83	**0.84**	0.91	**0.92**	0.90	**0.91**
Top10	0.87	**0.89**	0.80	**0.83**	0.80	**0.81**	0.80	**0.82**	0.89	**0.90**	**0.88**	0.86
All	0.89	**0.93**	0.83	**0.88**	0.83	**0.87**	0.82	**0.86**	0.91	**0.93**	0.82	**0.87**

Here we re-visit the proposed research questions to answer them based on our findings.

RQ1. Is there evidence that machine learning can support procedure prediction when using shallow data description that includes main COVID-19 symptoms?

Yes, we showed that both prediction models work well in terms of all evaluation measures. Although there is variation between labels, we can say that the obtained results for all labels are somewhat high and accepted. Strictly speaking, we can observe that most AUC values for two prediction models along all feature subsets range from 0.8 to 0.95, which is reasonable enough to judge both prediction models' superiority.

RQ2. Do the designated features have a significant impact on the accuracy of the constructed prediction models?

Indeed, we showed that machine learning could work well for predicting procedure labels, but there was no clear pattern for which feature subsets can boost performance accuracy. From the overall results, we can confirm that using the Top 6 features would produce accurate but acceptable results for both prediction models. These Top 6 features (SpO2, Chronic Diseases, Body Ache Fatigue, Sore Throat, Age, and Vomiting) achieved an IG ratio above 0.21. Concerning the difference be-tween SVM and NB along with all feature subsets, we did not notice any significant difference between them, so both can produce relatively similar predictions. The final remark is that using all features would work better than using Top2 or Top 3 features, which confirms that all collected features are working collaboratively to support decision-making.

7. Threats to Validity

The main threats to the validity of this study are the collected dataset. The dataset employed in this paper was collected in Iraq from the Ministry of Health in Bagdad. The accuracy of collected data depends on the person responsible for data entry. Furthermore, the derived conclusion cannot be easily generalized because of the locality of data representing a small population in a small georgical region. However, offering such data from hospitals or public health officials is not easy. Still, since this project collaborates with health care researchers in the Ministry, they did their best to collect it. The second threat is the designated features. The features that de-scribe our datasets were designated in collaboration with researchers in the Ministry of health based on the main symptoms of COVID-19 patients. These features might not be only the ones that can help in deciding on the correct procedure. Still, they are often available without the need to expose patients to further blood tests or X-ray images to get more diagnosis features. Finally, the way we designed the experiments might not give a complete picture about the performance of procedure prediction for two reasons: 1) the choice of machine learning algorithms, and 2) the way we select feature subsets. One can find other machine learning that might work well with this kind of dataset and generate many feature subsets by ignoring the ranking of these features.

8. Conclusion

Since the WHO announced the COVID-19 global pandemic in 2020, at that time patient cases have been continuously increasing. All countries took different actions to prevent and stop the virus's quick spread, including lockdown and social distancing, which led to significant reductions in all aspects of our lives. Health officials warn that health systems are overwhelmed with COVID-19 patients even with new mutated strains of Covid-19 have emerged, including the Omicron and Delta strains reducing or impeding efforts to eradicate the disease. The quick spread of the new

strain of COVID-19 virus has increased the pressure on hospitals and all kinds of health care system resources. Therefore, there is a desperate need to reduce such pressure and avoid extra costs for patients and hospitals. This paper proposed a new solution to predict the correct medical procedure for COVID-19 patients when they visit emergency rooms in hospitals. The proposed system is not a replacement for medical specialists, but it is decision-making to assist them with limited information and time. Two machine learning classification algorithms were used to construct procedure prediction models based on a dataset collected from the Ministry of Health in Iraq. To investigate the performance of the proposed models, we examine the impact of top feature subsets on the accuracy of these models. We have found that the constructed algorithms can produce an accurate prediction for each class label and acceptable accuracy for the aggregated labels. We could not find a clear pattern for which feature subset is most stable for SVM and NB regarding the optimal feature subset. However, we recommend not to use a small number of features or to use all features, but to use the top 6 features would work well with two models. Remarkably, there was a fluctuation in AUC along with all feature subsets. Also, the Mann-Whitney test did not find any significant difference between predictions generated by SVM and those generated by NB Finally, since COVID-19 is mutating and new generations are emerging, collecting data with new features is required. For example, one can collect data about X-ray images, blood tests, nursing assessments, laboratory data, and more medical test information about infected patients. These data may increase the predictive performance of procedure prediction. Also, the data collected in this study cannot be generalized because it was collected from a small geographical region in the same population. Therefore, one should collect data from different countries to make the out-come of the procedure prediction more generalizable.

Acknowledgement

The authors are grateful to the Applied Science Private University, Amman, Jordan, for the financial support granted to cover the publication fee of this research article. In addition, we want to thank Dr. Seveem Omran Jasim, responsible for the Public Health Unit in the Iraqi Ministry of Health, for the unconditional assistance to provide us with the necessary dataset without hesitation to complete this study.

References

1. Abdulah, D. M. and Mohammed, A. A. (2020) 'The consequences of the COVID-19 pandemic on perceived stress in clinical practice: Experience of Doctors in Iraqi Kurdistan', *Romanian Journal of Internal Medicine*, pp. 1–22. doi: 10.2478/rjim-2020-0020.
2. Ali Abed, S., Hussein Ewaid, S. and Al-Ansari, N. (2019) 'Evaluation of Water quality in the Tigris River within Baghdad, Iraq using Multivariate Statistical Techniques', in *Journal of Physics: Conference Series*. Institute of Physics Publishing, p. 072025. doi: 10.1088/1742-6596/1294/7/072025.
3. Arpaci, I. *et al.* (2021) 'Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms', *Multimedia Tools and Applications*, 80(8), pp. 11943–11957. doi: 10.1007/s11042-020-10340-7.
4. Claesen, M. *et al.* (2014) 'Fast Prediction with SVM Models Containing RBF Kernels'. Available at: http://arxiv.org/abs/1403.0736 (Accessed: 24 April 2021).
5. Das, A. K., Mishra, S. and Gopalan, S. S. (2020) 'Predicting CoVID-19 community mortality risk using machine learning and development of an online prognostic tool', *PeerJ*, 8, p. e10083. doi: 10.7717/peerj.10083.
6. Debnath, S. *et al.* (2020) 'Machine learning to assist clinical decision-making during the COVID-19 pandemic', *Bioelectronic*

Medicine, 6(1), pp. 1–8. doi: 10.1186/s42234-020-00050-8.
7. Fayyoumi, E., Idwan, S. and Aboshindi, H. (2020) 'Machine learning and statistical modelling for prediction of Novel COVID-19 patients case study: Jordan', *International Journal of Advanced Computer Science and Applications*, 11(5), pp. 122–126. doi: 10.14569/IJACSA.2020.0110518.
8. Guastello, A. D. *et al.* (2024) 'Predictors of professional burnout and fulfilment in a longitudinal analysis on nurses and healthcare workers in the COVID-19 pandemic', *Journal of Clinical Nursing*, 33(1), pp. 288–303. doi: 10.1111/jocn.16463.
9. Haritha, D., Praneeth, C. and Pranathi, M. K. (2020) 'Covid prediction from x-ray images', in *Proceedings of the 2020 International Conference on Computing, Communication and Security, ICCCS 2020*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICCCS49678.2020.9276795.
10. Iwendi, C. *et al.* (2020) 'COVID-19 patient health prediction using boosted random forest algorithm', *Frontiers in Public Health*, 8(July), pp. 1–9. doi: 10.3389/fpubh.2020.00357.
11. Karegowda, A. G., Manjunath, A. S. and Jayaram, M. A. (2010) 'Comparative Study of Attribute Selection Using Gain Ratio and Correlation Based Feature Selection', *International Journal of Information Technology and Knowledge Management*, 2(2), pp. 271–277.
12. Kavadi, D. P. *et al.* (2020) 'Partial derivative Nonlinear Global Pandemic Machine Learning prediction of COVID 19', *Chaos, Solitons and Fractals*, 139, p. 110056. doi: 10.1016/j.chaos.2020.110056.
13. Malki, Z. *et al.* (2020) 'Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches', *Chaos, Solitons and Fractals*, 138, p. 110137. doi: 10.1016/j.chaos.2020.110137.
14. Minaee, S. *et al.* (2020) 'Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning', *Medical Image Analysis*, 65, p. 101794. doi: 10.1016/j.media.2020.101794.
15. Nassif, A. B. *et al.* (2018) 'Machine Learning Classifications of Coronary Artery Disease', *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing, iSAI-NLP 2018 - Proceedings*, pp. 1–6. doi: 10.1109/iSAI-NLP.2018.8692942.
16. Naudé, W. (2020) 'Artificial Intelligence against COVID-19: An Early Review', *IZA Discussion Paper*, (13110). Available at: www.iza.org (Accessed: 19 November 2020).
17. Nguyen, D. C. *et al.* (2020) 'Blockchain and AI-based solutions to combat Coronavirus (COVID-19)-like epidemics: A survey', *2020*, (April), pp. 1–15. Available at: https://www.techrxiv.org/articles/preprint/Blockchain_and_AI-based_Solutions_to_Combat_Coronavirus_COVID-19_-like_Epidemics_A_Survey/12121962/files/22292202.pdf (Accessed: 30 June 2021).
18. Shenoy, N., Luchtel, R. and Gulani, P. (2020) 'Considerations for target oxygen saturation in COVID-19 patients: Are we under-shooting?', *BMC Medicine*, 18(1), p. 260. doi: 10.1186/s12916-020-01735-2.
19. Yekkehkhany, B. *et al.* (2014) 'A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data', in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, pp. 281–285. doi: 10.5194/isprsarchives-XL-2-W3-281-2014.
20. Zhu, D. *et al.* (2024) 'Exploring telemental health practice before, during, and after the COVID-19 pandemic', *Journal of Telemedicine and Telecare*, 30(1), pp. 72–78. doi: 10.1177/1357633X211025943.

Author Biographies

Mohammad Azzeh is a professor of Data Science at Princess Sumaya University for Technology. Dr. Azzeh holds PhD in computing from University of Bradford, UK and MSc in Software Engineering from University of the West of England, UK. He is currently working as



a faculty staff member in Data Science department. His research interests focus on Data Science, Machine Learning, Data mining, Empirical Software Engineering, Mining Software Repositories. Dr. Azzeh is an invited referee for high quality journals and PC member of international conferences. He was a guest editor in Journal of Neural Computing and Applications (Springer). Dr. Azzeh was conference chair of the 7th and 8th international conference of Computer Science and Information Technology (2016 and 2018). Dr. Azzeh co-organized and co-chaired several special sessions/workshops: Computational Intelligence Applications in Software Engineering (CIASE 2013) at the 3rd IEEE International Conference on Communications and Information Technology (ICCIT 2013), Machine Learning for Predictive Models (MLPM 2013, 2014) at the IEEE International Conference on Machine Learning and Applications (ICMLA 2013 to 2021). He was also publicity chair of CSIT 2014 conference. Dr. Azzeh published over 40 research articles in reputable journals and conferences such as IET Software, Software: Evolution & Process, Empirical Software Engineering Applied Soft Computing and Systems & Software, Science of Computer Programming, Journal of Software Quality.