Journal of Emerging Computer Technologies
Research Article
Received: *2022-06-01* | ***Reviewing:*** *2022-06-02* & *2022-07-18* | ***Accepted:*** *2022-07-18* | ***Online:*** *2022-07-19* | ***Issue Date:*** *2022-06-30*

Symptom Based COVID-19 Prediction Using Machine Learning and Deep Learning Algorithms

Nesibe Yalçın
Department of Computer Engineering
Erciyes University
Kayseri, Türkiye
nesibeyalcin@erciyes.edu.tr
0000-0003-0324-9111

Sibel Ünaldı
Department of Electrical Electronics Engineering
Bilecik Şeyh Edebali University
Bilecik, Türkiye
sibel.unaldi@bilecik.edu.tr
0000-0001-9948-4284

***Abstract*— Research studies are carried out in many areas of science to cope with the impacts of the COVID-19 crisis in the world. Machine learning can be used for purposes such as understanding, addressing, fighting, and preventing - controlling COVID-19. In this research, the presence of COVID-19 has been predicted using K Nearest Neighbor, Support Vector Machines, Logistic Regression, and Multilayer Perceptual Neural Networks machine learning and Gated Recurrent Unit (GRU) and Long Short-Term Memory deep learning algorithms. A publicly available dataset that includes various features (i.e. wearing masks, abroad travel, contact with the COVID patient) and symptoms (i.e. breathing problems, fever, and dry cough) is used for the COVID-19 diagnosis prediction. The learning algorithms have been compared according to the evaluation metrics. The experimental results have been shown that GRU deep learning algorithm is more reliable with a prediction accuracy of 98.65% and a loss/mean squared error of 0.0126.**

Keywords*— *COVID-19, deep learning, symptom, machine learning, prediction

I. INTRODUCTION

COrona **VI**rus **D**isease 2019 (COVID-19) epidemic caused by SARS-CoV-2 still brings many problems globally. This epidemic, which emerged in Wuhan, Hubei Province of China in 2019 December, has spread worldwide rapidly. A novel coronavirus, whose symptoms may include dry cough, fever, and anosmia, was identified on 7 January 2020 [1, 2]. The World Health Organization (WHO) announced a pandemic on 11 March 2020 [3]. In October 2020, the total number of patients exceeded 39,500,000 [4]. According to WHO [5], as of 7 January 2022, a total of 298,915,721 COVID-19 cases and 5,469,303 COVID-19 deaths were announced globally, while the number of COVID-19 cases in the last 7 days was 13,307,762 and the number of COVID-19 deaths was 40,868. This pandemic emphasizes the ability of viral spread from animals to cause significant disease in humans [6].

Machine learning, a subfield of artificial intelligence, is a method that enables machines to produce new solutions based on previous solutions [7]. Machine learning can play an important role in research and predictions of COVID-19 or other diseases. It can be used to analyze, evaluate and triage COVID-19 cases by integrating into health provider programs and strategies [8]. Machine learning based applications/platforms show a huge potential for accelerating COVID-19 diagnosis and treatment [8-10]. It can be interpreted that the machine learning methods will be useful in improving the diagnostic accuracy by using it together with Polymerase Chain Reaction (PCR) test or other tests [11]. Prediction of diagnosis according to symptoms in pandemics is of great importance in terms of both initiating treatments with early diagnosis and creating highly accurate alternatives that can alleviate the workload of healthcare professionals.

In the literature, many studies are aiming to obtain faster and more accurate results for COVID-19 diagnosis, including machine learning and deep learning based on artificial intelligence principles [12-14]. In addition, Computed Tomography and X-ray medical images are used to accurately segment infected parts with artificial intelligence to increase the efficiency of COVID-19 diagnosis [15, 16]. In [8], it is aimed to figure out the role of machine learning algorithms in different studies dealing with COVID-19. Supervised learning algorithms have presented better results with 92.9% test accuracy compared to unsupervised learning algorithms. Reference [17] have studied on COVID-19 dataset and developed a model utilizing the Support Vector Machine (SVM) to estimate patients as COVID or not. An accuracy of 87% has been achieved in estimating 3 cases: not infected, mildly infected, and severely infected. In [18], Prophet, Random Forest, AutoRegressive Integrated Moving Average (ARIMA), Polynomial Regression, and Linear Regression models have been built up to detect COVID-19 confirmed cases in the USA using machine learning algorithms and Polynomial Regression has outweighed the other algorithms by giving the best estimates. Reference [19] has created a model for the future COVID-19 forecast of 7 countries (including Turkey) considering the number of cases. In addition to classical forecasting methods, machine learning methods have been implemented to a COVID-19 dataset and Facebook's Prophet method has given the lowest forecasting error for all countries. Various supervised machine learning algorithms have been applied to estimate COVID-19 in [20] worked on a COVID-19 dataset. The algorithms' performance has been evaluated using 10-fold cross validation, and after comparing all experiments, the highest accuracy rate has been obtained by SVM with 98.81%. Reference [21] have applied Naïve Bayes, Logistic Regression (LR), SVM, Decision Tree, and K Nearest Neighbors (KNN) machine learning algorithms for the determination of COVID-19 patients and have studied on a worldwide accessible dataset. In this prediction study based on their symptoms, Naïve Bayes and Decision Tree having an accuracy of 93.70% present the best performance. Reference [22] has proposed forecast models including Long Short - Term Memory (LSTM), bidirectional LSTM, ARIMA, and support vector regression for COVID-19 prediction. Bidirectional LSTM outperforms better for pandemic

Cite (APA): Yalçın, N., Ünaldı, S. (2022). Symptom Based COVID-19 Prediction Using Machine Learning and Deep Learning Algorithms. *Journal of Emerging Computer Technologies*, 2(1), 22-29.

prediction in planning and management in the public health system. Recent studies on the detection of COVID-19 are summarized in Table I.

In the study on human infection caused by 2019 novel coronavirus (2019-nCoV) [27], treatment and clinical features, and epidemiological, radiological, and laboratory characteristics of the 2019-nCoV infected patients have been reported. To clarify the clinical and epidemiological characteristics of 2019-nCoV, [28] also analyzed demographic, epidemiological, radiological, and clinical features and laboratory data of coronavirus patients. A patient with no history of diabetes, hepatitis, or tuberculosis is studied in [6]. This patient was admitted to the hospital 6 days after the onset of coronavirus disease and reported fever, dizziness, cough, and serious respiratory syndrome at presentation. The dataset used in [4] includes 8 basic features: demographic information (gender and age 60+), clinical symptoms (cough, fever, sore throat, shortness of breath, headache), and known contact with a confirmed COVID patient.

TABLE I. SUMMARY OF LITERATURE REVIEW ON COVID-19 DETECTION

Reference	Month - Year	Dataset	Description	Methods	Results
[4]	January 2021	Records from tested individuals	Prediction of COVID-19 diagnosis based on symptoms	Gradient Boosting	0.90area under a receiver operating characteristic (AUROC), 0.66 are under precision-recall curve (AUPRC)
[13]	February 2022	Laboratory blood tests	A novel Deep Neural Network (DNN) modelfor early COVID-19 diagnosis	LR, KNN, Decision Tree, Extremely Randomized Trees, SVM, Naïve Bayes, Random Forest, XGBoost, LSTM, DNN, Recurrent Neural Network (RNN), and Convolutional Neural network (CNN)	The proposed DNN model achieved an accuracy of 93.33%.
[14]	January 2022	RT-PCRvirology test results	A deep learning model to improve COVID-19 diagnostic performance	LSTM	The model exceeded a sensitivity of 90%
[15]	October 2021	X-ray and CT-scan medical images	Deep learning methods applied to medical images for COVID-19 detection	CNN models, VGG16, DenseNet121, ResNet50, ResNet152, and Fast.AI ResNet	High accuracy of 99%
[17]	May 2021	Extracted critical symptoms	Detection of COVID-19 from the symptoms	SVM	An accuracy of 87%
[18]	January 2021	Confirmed cases	Prediction of COVID-19 cases	Random Forest, Polynomial Regression, Linear Regression, Prophet, and ARIMA	A mean absolute error (MAE) of 1.86%
[19]	May 2020	Confirmed cases from 7 countries	Prediction of possible confirmed cases and mortality numbers	SVM, Holt-Winters, Prophet, and LSTM	The prophet model presented the lowest RMSE for all countries.
[20]	May 2021	Possible factors	Detection of COVID-19 presence	SVM, KNN, J48 Decision Tree, Random Forest, and Naïve Bayes	An accuracy of 98.81%
[21]	May 2021	Patient recordscontaining symptoms and actual results	Determination of COVID-19 patients among various age groups	SVM, KNN, Decision Tree, LR, and Naïve Bayes	An accuracy of 93.70%
[22]	August 2020	Confirmed and recovered cases	Prediction for COVID-19	LSTM, GRU, and Bidirectional-LSTM	MAE value of 0.007 and R^2 value of 0.9997
[23]	July 2020	Laboratory findings	Clinical predictive modelsto estimate COVID-19 infection	Artificial Neural Network (ANN), CNN, RNN, LSTM, SNNLLSTM, and CNNRNN	Deep learning models have an accuracy of over 84%.
[24]	January 2020	Clinical features	Prediction of COVID-19 mortality	MLPNN, KNN, J48 Decision Tree, Random Forest, Naïve Bayes, LR, and XGBoost	An accuracy of 95.03%
[25]	October 2020	Multiple features of patients	Prediction of COVID-19 risk	LR	An accuracy of 92%.
[26]	December 2021	Clinical symptoms	Symptom based prediction model for diagnosis of COVID-19 in children	Random forest, LR, MLP, SVM, Boosted Trees	Area under ROC of 0.65
This study	-	Symptoms and various features	Prediction of COVID-19 presence	KNN, SVM, LR, MLPNN, GRU and LSTM	Prediction accuracy of 98.65%, AUROC of 0.989 and AUPRC of 0.998 with 95% confidence interval (CI)

This research aims to analyze and estimate the COVID-19 presence based on the symptoms & features. For this purpose, KNN, LR, SVM, Multilayer Perceptron Neural Network (MLPNN) machine learning algorithms, and LSTM and Gated Recurrent Unit (GRU) deep learning algorithms have been used. GRU has been announced to be an appropriate algorithm for COVID-19 diagnosis prediction due to its better accuracy. The following section details the COVID-19 dataset used in the research. Section III provides the information about the preparation of the study and briefly explains the used algorithms and evaluation metrics. In Section IV, the experimental results are given and evaluated. The final section concludes and provides suggestions for future study.

II. COVID-19 DATASET

In this study, it is used a publicly available dataset entitled "Symptoms and COVID Presence" from Kaggle [29]. The dataset is updated on 2020-08-18. It covers data between 2020-04-17 and 2020-08-29. It contains 20 features that indicate the presence of various symptoms and 1 class feature (the person has COVID or not). The total number of examples in the dataset is 5434, the number of COVID patients is 4383 (80.7%), and the number of healthy people is 1051 (19.3%), as shown in Fig. I. The presence of COVID in potential patients is indicated as "Yes" or "No".

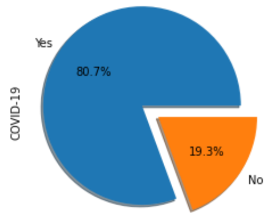


FIG. I. THE PRESENCE OF THE COVID-19

Different people are affected by COVID-19 in various ways. Infected patients develop various levels of symptoms. In addition to several symptoms of COVID-19 infection such as shortness of breath, fever, and dry cough, some infected people have experienced fatigue, anosmia (loss of taste or smell), and muscle aches [30]. The correlation between each feature/symptom and coronavirus disease has been computed and the obtained correlation coefficients have been listed in Table II. In addition, the number and rates of COVID-19 positive and negative cases are given according to the status of each feature/symptom.

TABLE II. THE BASIC STATISTICS OF THE COVID-19 DATASET

Feature & Symptom	Correlation Coefficient	Status	Total (n=5,434)		COVID-19 Positive (n=4,383)		COVID-19 Negative (n=1,051)	
			n	*%*	*n*	*%*	*n*	*%*
Breathing problem	0.444	Yes	3,620	66.6	3,369	76.9	251	23.9
		No	1,814	33.4	1,014	23.1	800	76.1
Fever	0.353	Yes	4,273	78.6	3,757	85.7	516	49.1
		No	1,161	21.4	626	14.3	535	50.9
Dry cough	0.464	Yes	4,307	79.3	3,878	88.5	429	40.8
		No	1,127	20.7	505	11.5	622	59.2
Sore throat	0.503	Yes	3,953	72.7	3,669	83.7	284	27
		No	1,481	27.3	714	16.3	767	73
Running nose	-0.006	Yes	2,952	54.3	2,375	54.2	577	54.9
		No	2,482	45.7	2,008	45.8	474	45.1
Asthma	0.09	Yes	2,514	46.3	2,124	48.5	390	37.1
		No	2,920	53.7	2,259	51,5	661	62.9
Chronic lung disease	-0.057	Yes	2,565	47.2	2,008	45.8	557	53
		No	2,869	52.8	2,375	54.1	494	47
Headache	-0.028	Yes	2,736	50.3	2,177	49.7	559	53.2
		No	2,698	49.7	2,206	50.3	492	46.8
Heart disease	0.027	Yes	2,523	46.4	2,064	47.1	459	43.7
		No	2,911	53.6	2,319	52.9	592	56.3
Diabetes	0.041	Yes	2,588	47.6	2,131	48.6	457	43.5
		No	2,846	52.4	2,252	51.4	594	56.5
Hypertension	0.103	Yes	2,663	49	2,258	51.5	405	38.5
		No	2,771	51	2,125	48.5	646	61.5
Fatigue	-0.044	Yes	2,821	51.9	2,228	50.8	593	56.4
		No	2,613	48.1	2,155	49.2	458	43.6
Gastrointestinal	-0.003	Yes	2,551	46.9	2,054	46.9	497	47.3
		No	2,883	53.1	2,329	53.1	554	52.7
Abroad travel	0.444	Yes	2,451	45.1	2,451	55.9	0	0
		No	2,983	54.9	1,932	44.1	1,051	100
Contact with COVID patient	0.357	Yes	2,726	50.2	2,582	58.9	144	13.7
		No	2,708	49.8	1,801	41.1	907	86.3
Attended large gathering	0.39	Yes	2,510	46.2	2,442	55.7	68	6.5
		No	2,924	53.8	1,941	44.3	983	93.5
Visited public exposed places	0.12	Yes	2,820	51.9	2,403	54.8	417	39.7
		No	2,614	48.1	1,980	45.2	634	60.3
Family working in public exposed places	0.16	Yes	2,262	41.6	1,994	45.5	268	25.5
		No	3,172	58.4	2,389	54.5	783	74.5
Wearing masks	-	Yes	0	0	0	0	0	0
		No	5,434	100	4,383	100	1,051	100
Sanitization from market	-	Yes	0	0	0	0	0	0
		No	5,434	100	4,383	100	1,051	100

According to Table II, there is a strong positive relationship between COVID-19 presence and sore throat, dry cough, and breathing problems. A sore throat was observed in 83.7% of those infected with the coronavirus. 88.5% and 85.7% of patients with COVID-19 have dry cough and fever complaints, respectively. The percentage of infections in people who have had close contact with COVID-19 patients is 94.71%. People who have recently traveled abroad and about 80.7% of people not wearing masks are infected.

III. Methods

This research presents machine learning / deep learning based models to detect COVID-19 diagnosis. The deep learning and machine learning algorithms involved in the research have been built with Python programming language and the Google Colab platform (a free online cloud-based product from Google Research) has been selected to execute the algorithms. Popular Python libraries such as NumPy, Pandas, Keras, Scikit-Learn, and Matplotlib have been used to develop the models. KNN, LR, SVM, and MLPNN supervised machine learning algorithms, and LSTM and GRU deep learning algorithms have been selected for this study to construct the prediction model. The models have been trained on the training dataset using the learning algorithms and then the trained models have been tested with the testing dataset. The architecture of the proposed prediction model is illustrated in Fig. II.

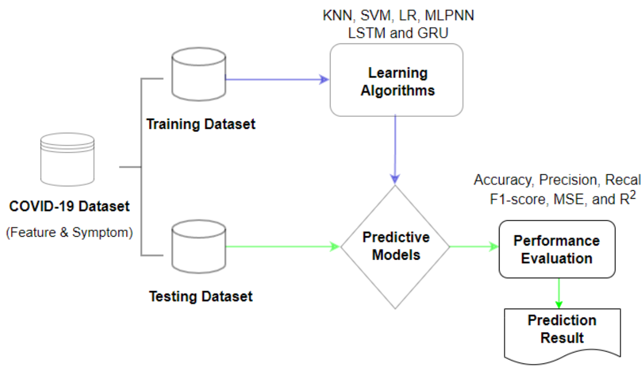


Fig. II. Graphical representation of the COVID-19 prediction model

KNN is the oldest classification algorithm and has some advantages simplicity in terms of complexity and quick calculation time [31]. The number of neighbors, k is a hyperparameter for building the prediction model in KNN. SVM is one of the widely preferred machine learning algorithms for classification or prediction problems. It uses the maximum margin concept and converts low-dimensional input space to higher dimensional space to create separable classes, depending on kernel functions [17]. MLPNN is a feed-forward neural network model using a backpropagation algorithm for training [32]. It consists of input, hidden (at least), and output layers and aims to minimize the difference between the target (desired output) and the output of the network [33]. LR, despite its name, is a linear model for predicting classes rather than regression. It is also a commonly studied simple machine learning algorithm for binary classification. It describes the relationship between at least one independent variable and a categorical dependent variable [25]. Recurrent Neural Networks (RNNs) are an extension of feedforward neural networks. LSTM and GRU networks are popular RNN architectures. LSTM is structurally composed of 3 main gates, namely forget gate, input gate, and output gate [22]. It takes into account crucial lessons acquired from previous experiences [19], unlike conventional neural networks. GRU is a gating mechanism in RNN and is less complex than LSTM. GRU has reset gate and update gate (combination of input gate and forget gate) [34].

Various metrics have been used to describe and evaluate the performance of each algorithm. Confusion matrix is often used to describe and illustrate the performance of the prediction/classification methods. As shown in Table III, it presents a summary table about the number of incorrect and correct predictions. True Negative (TN) and True Positive (TP) are the total numbers of correctly predicted negative and positive examples, respectively. False Negative (FN) and False Positive (FP) are incorrect predictions. False Negative (FN) and False Positive (FP) are the total numbers of incorrectly predicted positive and negative examples, respectively.

Table III. The Structure of The Confusion Matrix

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

Accuracy, a common evaluation metric, is the ratio of accurate predictions (TP and TN) over all (correctly and incorrectly) predictions. As can be seen in (1), it is computed depending on the confusion matrix.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

Precision, recall, and F1-score performance metrics have been also used to determine the algorithm(s) making the most accurate predictions and can be defined by (2), (3), and (4), respectively. Precision is the ratio of correct positive predictions to all positive predictions. Recall is the ratio of positive predictions to the total positive examples. F1-score measures the harmony and the balance of precision and recall metrics.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{F1} - \text{score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

Mean Square Error (MSE) has been used as a loss function for computing the loss between the real values and predictions. It is defined by (5).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)^2 \tag{5}$$

where$n$ is the total number of examples in the dataset and i is the index (i = 1, 2, 3, …, n). $t_i$is the target (actual, desired) output value, and y_i is the predicted output using the learning algorithm for i. example. Explanatory coefficient R^2 is computed by (6) and t_{ort} is the average of the target output values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(t_i - y_i)^2}{\sum_{i=1}^{n}(t_i - t_{ort})^2} \quad (6)$$

The dataset has been divided into two subsets (85% - 15%): a training dataset and a testing dataset. There are 5,434 examples in the dataset, the training dataset includes examples of 4,618 (3,716 COVID and 902 healthy) and the testing dataset examples of 816 (667 COVID and 149 healthy). Then the dataset has been passed on to the learning algorithms.

IV. Results and Discussion

The development of all the models has been performed under the Google Colab environment. In order to decide the optimal value of k in KNN, error rates have been calculated for all k neighbor numbers between 1 and 40, and the k value corresponding to the lowest error rate has been determined as 3. Euclidean distance is used as the distance function in KNN. As a result of experiments, the polynomial kernel is selected as the kernel function in SVM. Different neural network models (i.e., a maximum number of hidden layers/neurons and epochs, activation function) have been designed to determine the best network structure. After extensive experiments, the number of hidden layers is 2 and the numbers of hidden neurons are 64 and 32, respectively. The rectified linear unit (relu) activation function has been utilized for the hidden layers in MLPNN. The learning rate for weight updates between layers is a constant of 0.001. The summaries of the developed LSTM and GRU model architectures are depicted in Fig. III. The output shape and number of parameters in each layer can be seen clearly.

```
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 20, 100)           1000000

lstm_2 (LSTM)                (None, 20, 100)           80400

dropout_6 (Dropout)          (None, 20, 100)           0

lstm_3 (LSTM)                (None, 50)                30200

dropout_7 (Dropout)          (None, 50)                0

dense_4 (Dense)              (None, 10)                510

dense_5 (Dense)              (None, 1)                 11

=================================================================
Total params: 1,111,121
Trainable params: 1,111,121
Non-trainable params: 0
```

(a)

```
Layer (type)                 Output Shape              Param #
=================================================================
gru_2 (GRU)                  (None, 20, 100)           30900

dropout_4 (Dropout)          (None, 20, 100)           0

gru_3 (GRU)                  (None, 100)               60600

dropout_5 (Dropout)          (None, 100)               0

dense_3 (Dense)              (None, 1)                 101

=================================================================
Total params: 91,601
Trainable params: 91,601
Non-trainable params: 0
```

(b)

Fig. III. The developed (a) LSTM and (b) GRU models

Fig. IV provides a complete inside into the test results obtained after applying machine learning algorithms in the diagnosis prediction. KNN and MLPNN are more satisfactory in terms of the TP and TN values, respectively. It has been observed that KNN classifies the infected people less incorrectly and LR predicts a higher number of healthy people as ill in comparison to other algorithms.

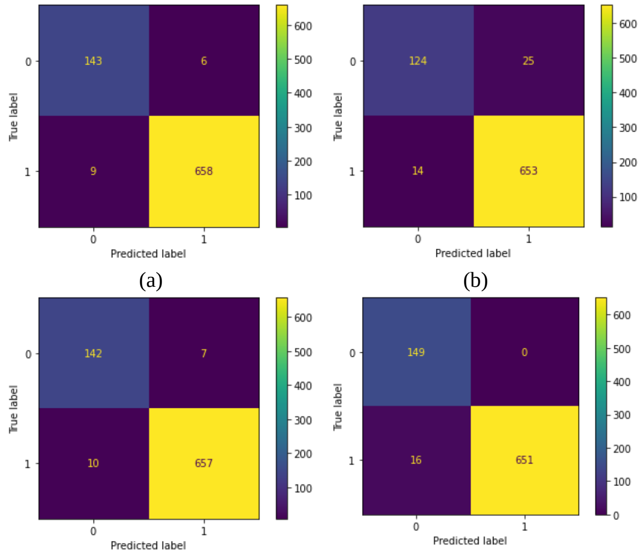


Fig. IV. The confusion matrices obtained after training on the dataset using (a) KNN, (b) LR, (c) SVM, and (d) MLPNN algorithms

The direct comparison of the studied machine learning algorithms' performance for the COVID-19 prediction is presented in Fig. V. The R^2 obtained for the training dataset and testing accuracy values show a similar distribution.

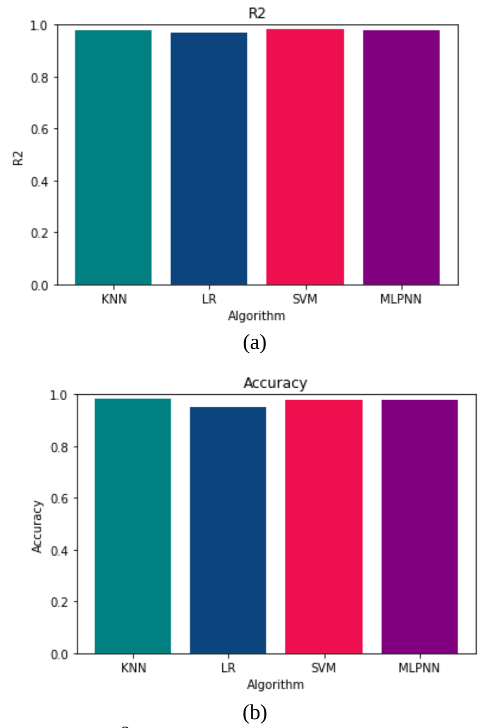
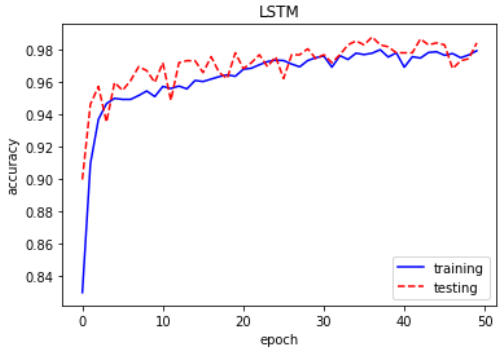
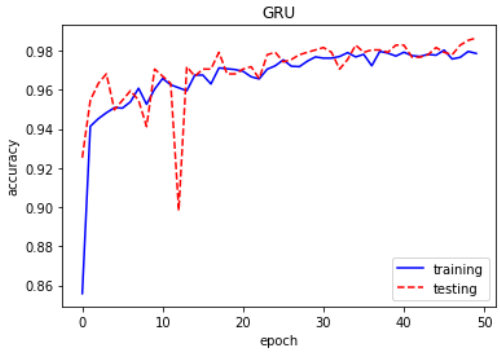


Fig. V. The (a) R^2 and (b) accuracy graphics of the machine learning algorithms

Fig. VI and Fig. VII provide comparisons between LSTM and GRU deep learning algorithms in terms of the most important performance metrics “accuracy” and “MSE” for training and testing datasets. From the overall comparison, it can be observed that both LSTM and GRU have prediction accuracy of more than 98% and also offer acceptable performance with a loss/MSE of about 1% in both training and testing.

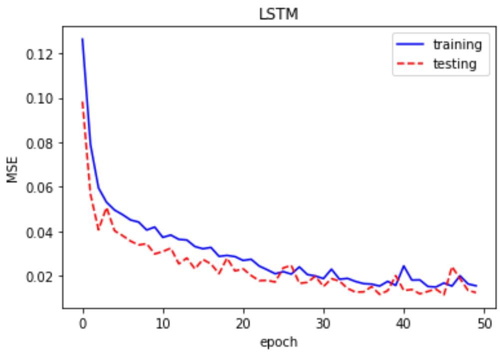


(a)

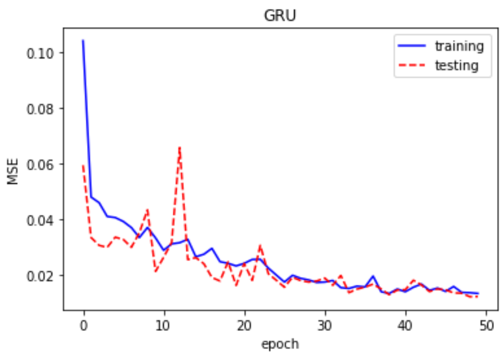


(b)

FIG. VI. THE OBTAINED ACCURACY RESULTS USING (a) LSTM AND (b) GRU ALGORITHMS



(a)



(b)

FIG. VII. THE OBTAINED LOSS GRAPHICS USING (a) LSTM AND (b) GRU ALGORITHMS

The obtained test results using the different performance metrics are presented comparatively in Table IV to give an idea about the prediction success of the symptom-based COVID-19 diagnosis for each algorithm. All of the used algorithms produce promising results with an accuracy of above 95%. The results show that GRU having a 98.65% accuracy is the best learning algorithm. That is to say, it has the highest level of accuracy when compared with KNN, LR, SVM, MLPNN, and LSTM.

TABLE IV THE SUCCESS OF THE ALGORITHMS

Algorithm	**Accuracy (%)**	**Precision**	**Recall**	**F1-score**	**Loss (MSE)**
KNN	98.16	0.991	0.987	0.989	0.0184
LR	95.22	0.963	0.979	0.971	0.0478
SVM	97.92	0.989	0.985	0.987	0.0208
MLPNN	98.04	1.000	0.976	0.988	0.0196
LSTM	98.41	1.000	0.98	0.99	0.0125
GRU	**98.65**	0.997	0.986	0.992	0.0126

TABLE V. COMPARISON WITH DIFFERENT LEARNING ALGORITHMS ON THE SAME DATASET

Reference	**Algorithms**	**Best Algorithm**	**Highest accuracy (%)**
[20]	J48 Decision Tree, Naïve Bayes, SVM, KNN, Random Forest	SVM	98.81
This study	KNN, LR, SVM, MLPNN	KNN	98.16
This study	LSTM, GRU	GRU	98.65

Villavicencio et al. [20] has also used the same dataset [29] for predicting the COVID-19 infected patients. Table V summarizes the performance of the studies on the same dataset with different algorithms in terms of accuracy rates. It can be observed that LSTM and GRU are provided 98.65% and 98.41% accuracy success for COVID-19 prediction, respectively. The SVM algorithm has an accuracy of 98.81% and is the best method reported in [20] for the detection of the potential presence of COVID-19. GRU prediction model have run 50 epochs and 30 times. According to the results of the best run, the model has predicted with the AUROC of 0.97-0.989 and AUPRC of 0.993 and 0.998 with 95% CI: 97.9% - 98.7% accuracy, 98.0% - 98.6% recall (sensitivity), and 95.5% - 99.4% specificity.

The accuracy, precision, and recall values computed for common machine learning algorithms (SVM and KNN) used in this study and [20] are compared comprehensively in Table VI. The results show that SVM in [20] and KNN in our study have better performance than the other algorithms used in each study. The better values are produced in our study when the same parameters are used for SVM and KNN algorithms. This study is achieved a higher prediction accuracy than the study of [20] when the polynomial kernel function is used for SVM and the number of nearest neighbors, k is selected as 3, and cross-validation is not performed for KNN.

TABLE VI. A COMPREHENSIVE COMPARISON OF THE STUDIES USED THE SAME DATASETS

Reference	Algorithm	Accuracy (%)	Precision	Recall
[20]	KNN (k = 1,10-fold cross-validation)	**98.69**	0.987	0.987
[20]	KNN (k = 3, no cross validation)	97.57	-	-
This study	KNN (k = 3, no cross validation)	**98.16**	0.987	**0.989**
[20]	SVM (Pearson VII universal kernel, 10-fold cross-validation)	**98.81**	0.988	**0.988**
[20]	SVM (Polynomial kernel, 10-fold cross-validation)	95.48	-	-
This study	SVM (Polynomial kernel no cross-validation)	95.48	-	-
This study	SVM (Polynomial kernel, no cross-validation)	**97.92**	**0.989**	0.985

V. CONCLUSION

This research aims to analyze and estimate the diagnosis of COVID-19 based on COVID-19 symptoms using several machine learning and deep learning algorithms. KNN, LR, SVM, MLPNN, GRU, and LSTM algorithms have been used for the COVID-19 diagnostic estimation. This research has been carried out on a worldwide available COVID-19 database for the diagnosis of this viral disease and has shown promising results with high accuracy, precision, recall, F1-score, and MSE. The best performance has been obtained by GRU (98.65% accuracy) and the lowest accuracy by LR (95.22%). The second-best results in terms of prediction success and error rate have been presented by LSTM.

The results demonstrate the capability of machine learning and deep learning algorithms in predicting COVID-19. The study shows that using the learning algorithms together with other tests such as PCR can be a good alternative in terms of increasing the diagnostic accuracy. In the future, this study can be extended to address COVID-19 variants in the COVID-19 health care applications.

AUTHORS' CONTRIBUTIONS

Nesibe YALÇIN: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Performing the algorithms, Evaluating the results with performance metrics, and Writing the manuscript.

Sibel ÜNALDI: Conceptualization, Validation, Investigation, Resources, Analyzing the results, and Writing the manuscript

CONFLICT OF INTEREST

There is no conflict of interest in this study.

REFERENCES

[1] N.Alballa and I. Al-Turaiki, "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review", Inform. Med. Unlocked, vol. 24, pp. 100564 (1-17), 2021.

[2] C. I.Paules,H. D. Marston, and A. S.Fauci, "Coronavirus infections - more than just the common cold", JAMA: J. Am. Med. Assoc., vol. 323, pp. 707-708, 2020.

[3] WHO, "Virtual press conference on COVID-19 - 11 March 2020", 25 January 2022, Available online: https://www.who.int/docs/default-source/coronaviruse/transcripts/who -audio-emergencies-coronavirus-press-conference-full-and-final-11mar2020.pdf, 2020.

[4] Y.Zoabi, S.Deri-Rozov, andN.Shomron, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms", NPJ Digit. Med., vol. 4, pp. 3 (1-5), 2021.

[5] WHO, "WHO coronavirus disease (COVID-19) dashboard", 13 January 2022, Available online: https://covid19.who.int/, 2022.

[6] F. Wu et al., "A new coronavirus associated with human respiratory disease in China", Nature, vol. 579 (7798), pp. 265-269, 2020.

[7] O. Sevli andV. G.Başer, "COVID-19 salgınına yönelik zaman serisi verileri ile Prophet model kullanarak makine öğrenmesi temelli vaka tahminlemesi", European Journal of Science and Technology, vol. 19, pp. 827-835, 2020.

[8] A. S. Kwekha-Rashid, H. N. Abduljabbar, andB. Alhayani, "Coronavirus disease (COVID-19) cases analysis using machine-learning applications", Appl. Nanosci., pp. 1-13, 2021.

[9] M.Naseem, R.Akhund, H.Arshad, andM. T. Ibrahim, "Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: a Scoping review", J. Prim. Care Community Health, vol. 11, pp. 1-11, 2020.

[10] Jamshidi M. et al., "Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment", IEEE Access, vol. 8, pp. 109581–109595, 2020.

[11] E.Dinçmen, "Makine öğrenmesi ve Covid-19", 20 January 2022, Available online: https://www.isikun.edu.tr/web/1695-15661-1-1/isik_universitesi/hakkinda/yonetim__ idari_birimler__kurumsal_iletisim_daire_baskanligi__basinda_isik_universitesi__isik_yazilari/makine_ogrenmesi_ve_covid-19 (2022)

[12] Z. A. A. Alyasseri et al., "Review on COVID-19 diagnosis models based on machine learning and deep learning approaches", Expert Syst., pp. e12759 (1-32), 2021.

[13] S. B.Rikan, A. S.Azar, A.Ghafari, J. B.Mohasefi, and H.Pirnejad, "COVID-19 diagnosis from routine blood tests using artificial intelligence techniques", Biomed. Signal Process. Control, vol. 72,pp. 103263 (1-16), 2022.

[14] Y. Lee et al., "The application of a deep learning system developed to reduce the time for RT-PCR in COVID-19 detection", Sci. Rep., vol. 12, pp. 1234 (1-10), 2022.

[15] D.Yang, C.Martinez, L.Visuña, H.Khandhar, C.Bhatt, and J.Carretero, "Detection and analysis of COVID-19 in medical images using deep learning techniques", Sci. Rep., vol. 11, pp. 19638 (1-13), 2021.

[16] F.Zhang, "Application of machine learning in CT images and X-rays of COVID-19 pneumonia", Medicine, vol. 100 (36), pp. e26855 (1-13), 2021.

[17] S.Guhathakurata, S.Kundu, A.Chakraborty, and J. S.Banerjee, "A novel approach to predict COVID-19 using support vector machine", Data Science for COVID-19, pp. 351-364, 2021.

[18] N. S.Özen, S.Saraç, and M.Koyuncu, "COVID-19 vakalarının makine öğrenmesi algoritmaları ile tahmini: Amerika Birleşik Devletleri örneği", European Journal of Science and Technology, vol. 22,pp. 134-139, 2021.

[19] R. Ünlü and E.Namlı, "Machine learning and classical forecasting methods based decision support systems for COVID-19", Comput., Mater. Contin., vol. 64(3), pp. 1383-1399, 2020.

[20] C. N.Villavicencio, J. J. E.Macrohon, X. AInbaraj., J. -H.Jeng, and J. -G.Hsieh, "COVID-19 prediction applying supervised machine learning algorithms with comparative analysis using WEKA", Algorithms, vol. 14(7), pp. 201 (1-22), 2021.

[21] M. Malik et al., "Determination of COVID-19 patients using machine learning algorithms", Intell. Autom. Soft Comput., vol. 31(1), pp. 207-222, 2022.

[22] F.Shahid, A.Zameer, and M.Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM", Chaos Solit. Fractals, vol. 140, pp. 110212 (1-9), 2020.

[23] T. B.Alakus and I.Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection", Chaos Solit. Fractals, vol. 140, pp. 110120 (1-7), 2020.

[24] K.Moulaei, M.Shanbehzadeh, Z.Mohammadi-Taghiabad, and H.Kazemi-Arpanahi, “Comparing machine learning algorithms for predicting COVID-19 mortality”, BMC Medical Inform. Decis. Mak., vol. 22, pp. 2(1-12), 2022.

[25] A. B.Majumder, S.Gupta, D.Singh, and S.Majumder, “An intelligent system for prediction of COVID-19 case using machine learning framework-logistic regression”, J. Phys. Conf. Ser., vol. 1797 (1),pp. 012011(1-9), 2021.

[26] J. M. Antoñanzas et al. “Symptom-Based Predictive Model of COVID-19 Disease in Children”, Viruses, vol. 14, pp. 63, 2022.

[27] C. Huang et al., “Clinical features of patients infected with 2019 novel Coronavirus in Wuhan, China”, The Lancet, vol. 395(10223), pp. 497-506, 2020.

[28] N. Chen et al., “Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study”, The Lancet, vol. 395(10223),pp. 507-513, 2020.

[29] Kaggle, “Symptoms and COVID Presence”, 8 January 2022. Available online: https://www.kaggle.com/hemanthhari/symptoms-and-covid-presence

[30] A. Tsatsakis et al., “SARS-CoV-2 pathophysiology and its clinical implications: An integrative overview of the pharmacotherapeutic management of COVID-19”, Food Chem. Toxicol., vol. 146,pp. 111769, 2020.

[31] E. Karaahmetoğlu, S.Ersöz, A. K.Türker, V.Ateş, and A. F.İnal, “Evaluation of profession predictions for today and the future with machine learning methods: emperical evidence from Turkey”, Journal of Polytechnic, (in press)

[32] I.Balikci Cicek and Z.Kucukakcali, “Classification of prostate cancer and determination of related factors with different artificial neural network”, Middle Black Sea Journal of Health Science, vol. 6(3), pp. 325-332, 2020.

[33] N.Yalcin, G.Tezel, and C.Karakuzu, “Epilepsy diagnosis using artificial neural network learned by PSO”, Turk. J. Electr. Eng. Comput. Sci., vol. 23(2), pp. 421-432, 2015.

[34] S.Cakir, S.Toklu, and N.Yalcin, “RPL attack detection and prevention in the Internet of Things networks using a GRU based deep learning”, IEEE Access, vol. 8, pp. 183678-183689, 2020.