Covid-19 Forecasting by Using Some Supervised Machine Learning Techniques

Salema Younus
College of Electrical and Electronics Technology (CEET), Benghazi, 5213.
salema_younus@ceet.edu.ly

Omar M. Sallabi
University of Benghazi
omar.sallabi@uob.edu.ly

Tarig El-Shheibia
University of Benghazi
Tarig.elshheibia@uob.edu.ly

Abstract:

COVID-19 is such a dangerous virus that the World Health Organization (WHO) has declared it a pandemic. Many countries have taken policies to control the virus spread to overcome this global epidemic due to its rapid transmission from one person to another. As the number of individuals displaying symptoms has risen worldwide, the Covid-19 virus has proven challenging to detect and treat at an early stage, which has become a major challenge. This study use supervised machine learning techniques, Decision Tree (DT), and Support Vector Machine (SVM) to construct an algorithm. The algorithm was trained and tested on the Mexican government website dataset. The results show a score of 64% for DT and 65% for SVM.

Keywords: *Covid-19, Coronavirus, Supervised Machine learning.*

الملخص

يعد كوفيد –19 فيروسًا خطيرًا لدرجة أن منظمة الصحة العالمية أعلنته وباءً. وقد اتخذت العديد من الدول سياسات للسيطرة على انتشار الفيروس ولعبت دورًا فاعلًا في التغلب على هذا الوباء العالمي نظرًا لانتقاله السريع من شخص إلى آخر ونظرًا لارتفاع عدد الأفراد الذين تظهر عليهم الأعراض في جميع أنحاء العالم، فقد أثبت فيروس كورونا أنه من الصعب اكتشافه وعلاجه في مرحلة مبكرة، والذي أصبح تحديًا كبيرًا، استخدمت هذه الدراسة تقنيات التعلم الآلي الخاضعة

للإشراف وبالتحديد تقنية شجرة القرار (DT) و تقنية آلة المتجهات الداعمة (SVM) ، وتم تدريبها واختبارها على مجموعة بيانات من موقع الحكومة المكسيكية، وتم الحصول علي نتائج بدقة 64٪ لـ DT و 65٪ لـ .SVM.

1. Introduction

The novel coronavirus disease 2019 (COVID-19) pandemic caused by the SARS-CoV-2 poses a critical and urgent threat to global health. The outbreak began in the People's Republic of China's Hubei region in early December 2019 and has since expanded internationally (Singhal,2020; Zoabi et al.,2021). As of October 2020, the number of patients confirmed to have the disease has more than 39,500,000 in >180 countries. Though the number of people infected is probably much higher, more than 1,110,000 people have died from COVID-19 (Zoabi et al.,2021).

Predicting the COVID-19 pandemic is crucial. The ability to foresee the disease, in particular, allows a country to respond correctly shortly. However, there are significant obstacles to predicting this condition. (Mondal et al.,2021).

The tracking of infected people, the lack of a definitive treatment option, the increased risk of fatality for the elderly or those with other serious diseases, the 2-week incubation period, and the lack of accuracy in the available datasets, period, and strictness of the lockdown are just a few of the challenges. (Mondal et al.,2021).

There has recently been much interest in discovering effective COVID-19 management solutions. One major obstacle is the growing and changing the number of COVID-19-related data, which makes creating appropriate solutions difficult. Artificial Intelligence (AI) can assist in managing COVID-19 under certain scenarios. AI's subsets, Machine Learning (ML) and Deep Learning (DL) approaches can help mobilize and save time in the medical, logistic, human resources, and other fields. It should be noted that ML and DL are commonly utilized to find patterns in data samples in various medical systems. Machine learning and deep learning can map patterns from big data sets and improve quickly as more data becomes available. Moreover, AI can ensure reliability in predicting the spread of the virus, classifying suspected patients, and data-driven diagnosis of COVID-19 (Mondal et al.,2021).

In the literature, some studies are reported that focus on machine learning applied to COVID-19. In one study (Asl et al.,2020), Fuzzy logic (FL) was chosen as one of the best methodologies for describing systems with a lot of ambiguity and complexity. We plan to use the benefits of FL in making decisions about instances that require ICU treatment. An interval type-2 fuzzy expert system is proposed in this work to predict ICU admission in COVID-19 patients.

In this approach (Zhang and Liu,2021), a hybrid intelligent model is proposed to simulate the spreading of COVID-19. First, consider the effect of control measures, such as government investment, media publicity, medical treatment, and law enforcement, in spreading the epidemic. Then, the infection rates are optimized using a genetic algorithm (GA). A modified susceptible-infected-quarantined-recovered (SIQR) epidemic spreading model is developed. In addition, the SIQR model incorporates the long short-term memory (LSTM) to create a hybrid intelligent model. Further improving system model characteristics will obtain the perfect prediction model and control measures. The suggested hybrid intelligence algorithm offers a strong prediction capacity, according to simulation findings.

In (Alamsyah et al.,.2021) proposed this study predicts COVID-19 with 25 attributes by adjusting the learning parameter of a Recurrent Neural Network to find the optimum parameter. After the training phase, the learning rate, hidden layer, and maximum epoch parameters generated the optimum value for building the best model of RNN. The values are 0.3 for the learning rate, 6 for the hidden layer, and 450 for the maximum epoch. The research results show that the percentage of accuracy is 88.

(Wibowo,2021) This approach has developed a model for predicting infected cases. The recovery cases from COVID-19 and mortality for each province in Indonesia using the Long Short-Term Memory (LSTM) machine learning method. The results of the model evaluation of this method used the root mean squared error (RMSE) approach.

An ANN architecture (Shawaqfah and Almomani,2021) was created to forecast the devastating effects of a pandemic outbreak in Qatar, Spain, and Italy. The prediction model was validated and tested using official statistics data collected from each country till July 6th. The root-mean-square error was used to assess the model's sensitivity (RMSE). For the dates

studied, the mean absolute percentage error and the regression coefficient index R2 gave remarkably accurate values of 0.99 for the expected correlation between infected and dead cases. COVID-19's tested and validated growth model for these countries demonstrated the effects of government and medical sector initiatives to mitigate the pandemic effect. Moreover, the effort to decrease the spread of the virus reduces the death rate. The differences in the spread rate were related to different exogenous factors (such as social, political, and health factors, among others) that are difficult to measure.

In this work, Decision Tree (DT) and Support Vector Machine (SVM) are used, and trials on the database from the Internet from the Mexican government website.

2. The Proposed Model

The structure of the proposed methodology includes the main stages that occur after the database of covid-19 acquisition: input covid-19 database, description of the database, and attributes resulting in three cases of covid, data processing, forecasting using DT, and SVM. The main methodology of this work is illustrated in Figure. 1.

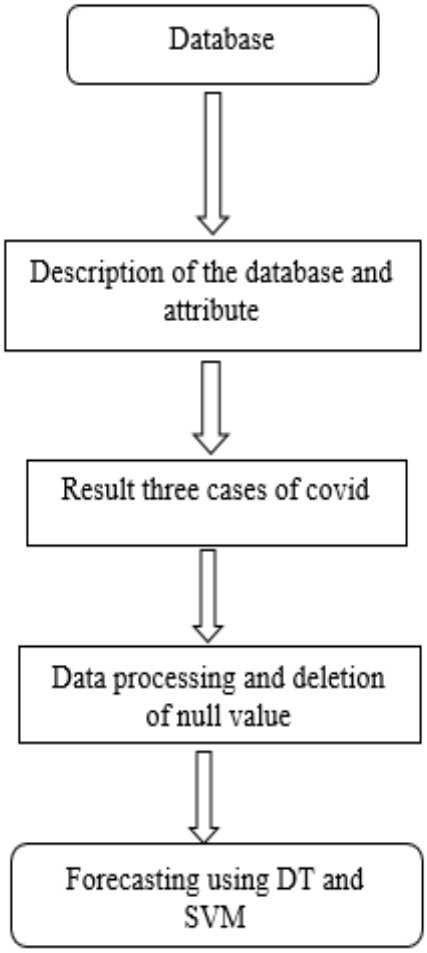

Figure.1. Stages of proposed covid-19 forecasting

2.1 The Database

There are 566602 objects and 23 attributes in the samples obtained from the Mexican government website. The data comprises the patient's personal information, kind, whether or not the patient is hospitalized, and the patient's cases (negative, positive, or waiting). In addition, the onset of symptoms, intubation, and chronic disorders, such as (asthma, hypertension, diabetes, and cardiovascular).

2.2 Supervised Learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from training examples and labeled training data. Algorithms that require external aid are known as supervised machine learning algorithms. The training and testing datasets are separated from the input dataset, and the output variable in the training dataset has to be predicted or categorized. For prediction or classification, all algorithms learn patterns from the training dataset and apply them to the test dataset (Mahesh.,2018). The supervised machine learning algorithms workflow is shown in Figure2.

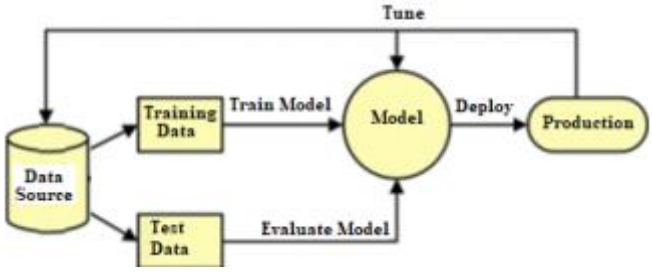

Figure 2. The Supervised Learning Workflow.

2.2.1 Decision Tree

Decision trees (DT) are trees that group attributes by sorting them based on their values. The decision tree is mostly used for classification. Nodes and branches make up each tree, each node represents attributes in a group that is to be classified, and each branch represents a value that the node can take (Dey,2016). An example of a decision tree is given in Figure.3.

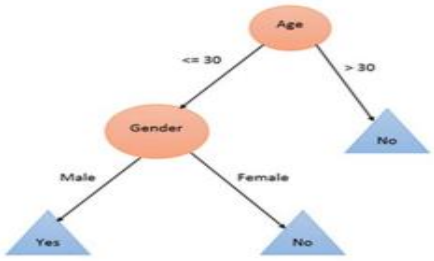

Figure .3. Decision Tree

2.2.2 Support Vector Machine

Another widely used state-of-the-art machine learning technique is the Support Vector Machine (SVM). Support vector machines are supervised learning models with related learning algorithms for classification and regression analysis in machine learning. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using the kernel trick. They implicitly map their inputs into high-dimensional feature spaces. It involves drawing lines between classes. The margins are drawn so that the distance between the margin and the classes is maximum, reducing the classification error (Mahesh.,2018).

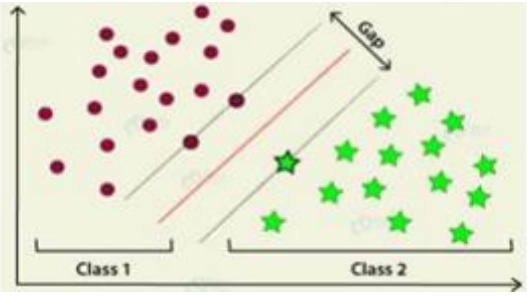

Figure .4. Support Vector Machine

3. Experimental Result

The entered data were classified into inpatient cases to three cases negative, positive, or waiting.

	covid_res
1	279035
2	220657
3	66910

Figure .5.a Test Result

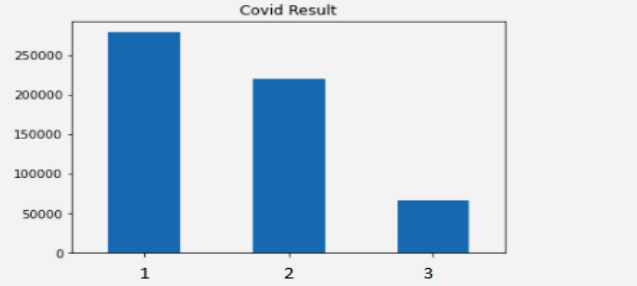

Figure 5.b. Covid Result

The results shown in Figures (5a&5b) are acceptable to some extent and very close. The accuracy was when using the DT at 64% and 65% when using SVM**.**

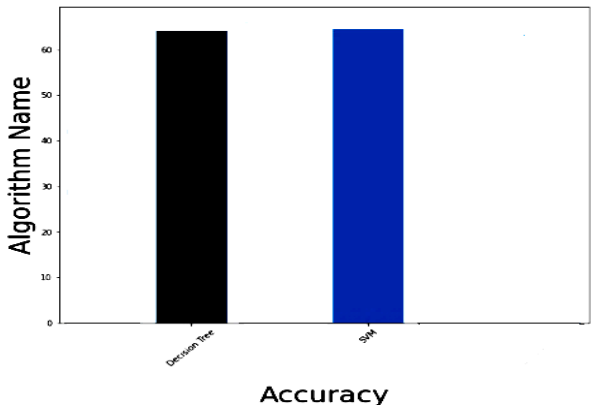

Figure. 5. Algorithms and Accuracy.

4. Conclusions and Future Work

The world has suffered from the COVID-19 pandemic, and infections are increasing rapidly. Predicting it has become an inevitable necessity to preserve human life. In this study, supervised machine learning techniques for forecasting specifically were DT and SVM for training and testing on data from the Internet. The results of training and testing were acceptable to some extent. Still, it needs more training so that we can get more accurate results.

The results of this work can be improved if the algorithms are trained more. They can also be improved using deep learning techniques because the algorithm can accurately predict by processing its data. The artificial neural network structure consists of multiple inputs, outputs, and hidden layers. The layer contains modules that convert the input data into information that the next layer can use for a particular predictive task.

References

- Alamsyah, A., Prasetiyo, B., Al Hakim, M. F., & Pradana, F. D. (2021). Prediction of COVID-19 Using Recurrent Neural Network Model. *Scientific Journal of Informatics*, *8*(1), 98-103.

- Asl, A. A. S., Ershadi, M. M., & Sotudian, S. (2021). Fuzzy Expert Systems for Prediction of ICU Admission in Patients with COVID-19. *arXiv preprint arXiv:2104.12868*.
- Dey, A.(2016). Machine Learning Algorithms: A Review. International Journal of Computer Science and Information Technologies, 7 (3), 1174-1179.
- Mahesh, B. (2018). Machine Learning Algorithms-A Review. *International Journal of Science and Research (IJSR).[Internet]*, *9*, 381-386.
- Mondal, M., Bharati, S., & Podder, P. (2021). Diagnosis of COVID-19 Using Machine Learning and Deep Learning: A Review. *arXiv preprint arXiv:2110.14910*.
- Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). *The Indian journal of pediatrics*, *87*(4), 281-286.Learning: A Review of Learning Types,www.preprints.org.
- Shawaqfah, M., & Almomani, F. (2021). Forecast of the Outbreak of COVID-19 Using Artificial Neural Network: Case Study Qatar, Spain, and Italy. *Results in Physics*, 104484.
- Wibowo, F. W. (2021, March). Prediction modeling of COVID-19 on provinces in Indonesia using long short-term memory machine learning. In *Journal of Physics: Conference Series* (Vol. 1844, No. 1, p. 012006). IOP Publishing.
- Zoabi, Y., Deri-Rozov, S., & Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *Npj digital medicine*, *4*(1), 1-5.
- Zhang, G., & Liu, X. (2021). Prediction and control of COVID-19 spreading based on a hybrid intelligent model. *Plus one*, *16*(2), e0246360.