International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I11P204

Volume 70 Issue 11, 28-46, November 2022

Review Article

A Survey of Machine Learning Methods to Detect COVID-19 Severity, Mortality, and Vaccines Efficacy

Abdullah M. Baqasah

Department of Information Technology, College of Computers and Information Technology,
Taif University, Saudi Arabia

Corresponding Author: a.baqasah@tu.edu.sa

Received: 16 July 2022 Revised: 16 October 2022 Accepted: 07 November 2022 Published: 26 November 2022

Abstract *- In this paper, we tackle the problem of COVID-19 detection. We present a survey on machine learning (ML) and deep learning (DL) methods to predict different vaccines' severity, mortality, and efficacy. For severity, we study the spread of Alpha, Beta, Delta, and Gamma in the countries where the variant first appears, such as the United Kingdom, South Africa, India, and Brazil. For mortality, we present works that study the rate of mortality caused by each variant. Finally, we present an overview of methods that respond to the question: do the five vaccines—produced by—Moderna, Pfizer, Novavax, Johnson & Johnson, and Astra Zeneca slow down the progress of COVID-19 variants?*

Keywords *- Coronavirus, Covid-19, Machine learning, Deep learning, Vaccine efficacy, Mortality, Severity.*

1. Introduction

The coronavirus disease 2019 (COVID-19), or the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was the second coronavirus to emerge and spread globally, causing a public health crisis. In December 2019, the virus moved from bats to humans in Wuhan, China, and has since spread throughout the rest of the world. As of October 2022, there have been around 623 million COVID-19 confirmed cases, more than 6.5 million reported deaths, and around 604 million recovered cases [1]. The World Health Organization (WHO) created the SARS-CoV-2 variations naming system—Alpha, Beta, Delta, and Gamma [2]—to avoid referring to the variants by the country where they were originally identified. Governments worldwide have responded to the crisis by creating strategies to fight the pandemic or even help mitigate the territorial effects of the crisis. Several studies, such as [3-5], have discussed factors, such as biochemical, hematological, and immunological findings, that clinicians can utilize to predict COVID-19 severity and mortality. From a technical point of view, COVID-19 has accelerated using artificial intelligence (A.I.) technologies in healthcare [6]. Several have enabled significant progress in the fight against the pandemic.

Solicited to respond to the COVID-19 health crisis, A.I. has been effective in several aspects of the fight against the pandemic, whether to understand this new coronavirus, diagnose it to predict its evolution, slow down its spread, or accelerate other aspects of medical research. More than a year after the onset of the crisis, AI-based tools continue to multiply and deliver good results. However, the need to accelerate their deployment should not obscure the ethical questions they raise. Faced with COVID-19, researchers worldwide have quickly and massively mobilized, giving rise to an avalanche of scientific publications. However, since COVID-19 literature is abundant and freely accessible, it is also indigestible because it is challenging to navigate. This challenge sparked several initiatives, such as allowing articles to be categorized and rated or creating interactive visualizations. The challenge is to guide researchers and practitioners toward the research results most relevant to them and facilitate their interpretation. The techniques to predict COVID-19 severity, mortality, and vaccine efficacy are divided into two categories: traditional approaches, including the machine learning (ML) approach, and deep learning (DL) approaches.

This paper is organized as follows: Section 2 presents various prediction techniques used in ML and DL. Section 3 presents an overview of works related to COVID-19 variants' severity. Section 4 tackles the problem of mortality caused by COVID-19. Section 5 introduces the efficacy of vaccines in slowing the progression of COVID-19 variants. Section 6 is a discussion followed by Section 7, which concludes the paper and provides some suggested future works related to COVID-19 detection. To remove any confusion regarding the abbreviations used in this study, a list of abbreviations and their definitions are provided in Table 1.

Table 1. List of abbreviations and their definitions

Acronym	Definition
COVID-19-related abbreviations	
A.E.	Adverse Effect
CoVA	COVID-19 Acuity Score
COVID-19	Coronavirus Disease 2019
CXR	Chest X-Ray
hs-CRP	high-sensitivity C-Reactive Protein
ICU	Intensive Care Unit
LDH	Lactate Dehydrogenase
mRNA	messenger Ribonucleic Acid
MV	Mechanical Ventilation
PCR	Polymerase Chain Reaction
RCT	Randomized Clinical Trial
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SGTF	S-Gene Target Failure
Machine learning-related abbreviations	
A.I.	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under the Curve
CI	Confidence Interval
CNN	Convolutional Neural Network
DL	Deep Learning
D.T.	Decision Tree
GBDT	Gradient-Boosted Decision Trees
KNN	K-Nearest Neighbors
L.R.	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MLC	Machine Learning Classifier
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
Q.M.	Quantitative Matrix
ResNet	Residual Network
R.F.	Random Forest
RMSE	Root Mean Squared Error
SVM	Support Vector Machine
XGBoost	eXtreme Gradient Boosting
Organizations or entities	
CDC	Centers for Disease Control and Prevention
E.D.	Emergency Department
MGH	Massachusetts General Hospital
WHO	World Health Organization

2. Machine Learning Approaches

2.1. Traditional Approaches

ML approaches are considered traditional machine learning and are based on a learning-by-example paradigm; they require both training and testing sets. The algorithms for decision trees (D.T.s), support vector machines (SVMs), and k-nearest neighbors (KNNs) are covered in the following subsections.

2.1.1. Decision Trees

The D.T. technique is a supervised learning technique whose goal is to automatically calculate the values of the endogenous variable (to be predicted), fixed a priori, from other information (exogenous or predictive variables) [7]. The principle of D.T.s is based on the recursive partitioning of data. Partitioning aims to obtain homogeneous groups from the predicted variable's point of view. The result is a hierarchical chain of rules. A path, starting from the root to a tree leaf, constitutes an assignment rule of the type "If condition Then conclusion." All these rules constitute the prediction model [8], either a set of individuals or objects affected by the learning problem.

The learning algorithm takes a sample as the input, comprising N records (texts) classified (D_j, C_i), and provides a D.T. as output. The algorithm proceeds top-down; it starts from the root and then recursively chooses the children's label. Most of these classifiers use a representation of the binary document and are, therefore, created by binary trees. There are several algorithms, with the most popular being ID3 [9] and its successors C4.5 [10] and C5.0 [11].

2.1.2. Support Vector Machine

SVMs are a recent class of ML methods introduced by [12]. SVMs look for a "thick" decision surface to separate the training set points into two classes. With this technique, the model's parameters are learned using a training dataset. It uses a kernel function, which enables the best possible data separation. The SVM aims to find a classifier that best divides the data into these two classes and optimizes the distance between them. As seen in Figure 1, the hyperplane, a linear classifier used with an SVM, distinguishes between the two categories of points.

From Figure 1, we notice that the optimal hyperplane separating the points of two classes is the one that passes in the middle of the space between these classes. Support vectors are the closest instances necessary to identify the ideal hyperplane. The margin is the separation between this plane and the support vectors.

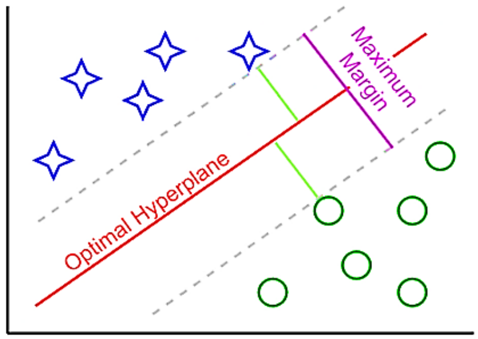


Fig. 1 Solution of a classification model using support vector machines

Table 2. Machine learning techniques

ML Technique	Description	Applications	Algorithms
SL	Labeled data The known number of classes Used to classify future observations	Classification Regression Estimation	Linear Regression SVM Descision Tree (DT)
UL	Unlabeled data An unknown number of classes Focus on finding patterns and gaining insights from the data Used to explore and understand data	Clustering Prediction	K-means Clustering Gaussian mixture Hidden Markov models
SSL	Combination of S.L. and U.L. methods (mix of a small amount of labeled data and unlabeled data)	Classification Clustering	Transductive SVM Graph-based algorithm Generative models
RL	Based on the system of rewards and punishment	Decision making	Markov decision process Brute force Q-learning

2.1.3. K-Nearest Neighbors

The basic idea of KNN is to predict the class or author of a text(T) as a function of the k closest neighbors already tagged in memory [13]. The learning phase consists of storing the labeled examples. The classification of new texts is done by calculating the distance between the vectorial representation of the document and that of each example in the corpus. Measuring the similarity between documents is one of the essential traits of this classifier. Since the texts are represented in vector form, points in an n-dimensional space, we determine the nearest neighbors by calculating the distance between these points.

Four distinct groups primarily comprise ML methods. Among them are supervised learning (S.L.), which employs labeled input and output data; unsupervised learning (U.L.), where the system analyzes and clusters unlabeled data; semi-supervised learning (SSL), where a small amount of labeled data is mixed in with a large amount of unlabeled data during training; and reinforcement learning (R.L.), where the machine interacts with a dynamic context. Table 2 summarizes the differences between ML techniques and lists their applications and algorithms.

2.2. Deep Learning Approaches

DL approaches represent an emerging branch of ML methods based on artificial neural networks (ANNs) [14]. This approach applies multiple layers of neural networks, which permits learning data representation in a supervised and unsupervised way with several levels of abstraction using a training set. DL approaches include numerous techniques, starting from basic ones such as feed-forward neural network (FFNN), recurrent neural network (RNN), long short-term memory (LSTM), or gated recurrent unit (GRU) to more complicated techniques. Figure 2 displays the overall DL text classification architecture (excerpted from [15]).

The represented architecture includes various layers:

- Dense Embeddings: This layer allows transforming text into real numbers.
- Hidden Layers: It is located between the algorithm's input and the output. These layers permit applying a nonlinear transformation of the inputs entered into the network.
- Output Units: According to the existing problem, the output layer could have either a Sigmoid function when we face a binary classification or a Softmax function for binary and multi-classification output.

Distinct approaches were developed and tested to predict the mortality or severity of COVID-19. The dataset essentially determines each method's use. A linguistic approach, for instance, is used when a strong lexicon is available. Due to their improved performance and ability to deliver more accurate findings, ML and DL approaches are more popular today. The DL model used in ResNet-50 and ResNet-152 is presented in the following section. Image classification is a technique for digitally extracting data from image classes.

Image classification can be supervised (each image is given a label) or unsupervised (the training model has learned to classify images based on patterns). We present ResNet-50 layers, a convolutional neural network (CNN) architecture. To sum up, CNN's feeds are:

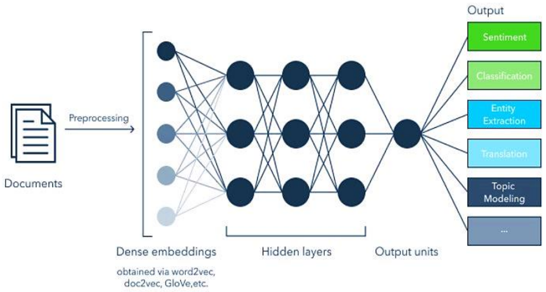


Fig. 2 General deep learning text classification architecture

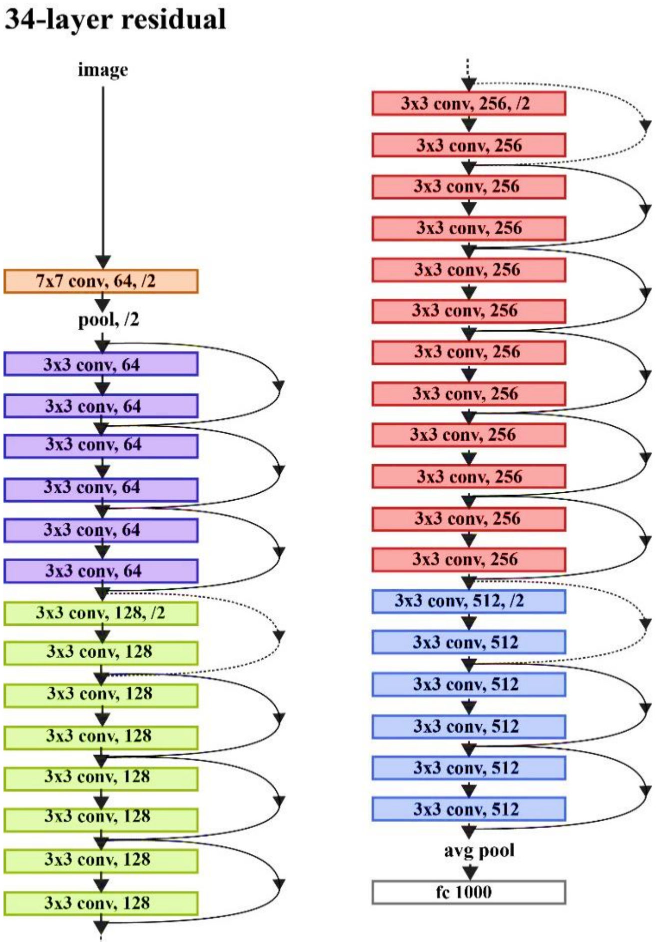


Fig. 3 The architecture of ResNet-50 with 34 residual layers

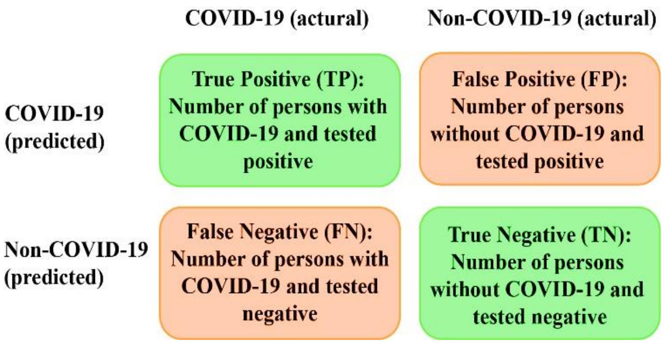
- Convolution layer (extraction function with filtering)
- Strides (displacement of pixels on the input matrix)
- Padding rectified linear unit (ReLU) (introduce non-linearity into the network)
- Padding layer (reduce the number of parameters)
- Fully connected layer (the input to the fully connected layer is the output from the final pooling or convolutional layer)

ResNet uses residual learning instead of trying to learn certain features. Simply put, the residue is the characteristic subtracted from the input that this layer learned. There are some variations of ResNet other than ResNet-50. A shortcut connection that omits one or more levels is the primary concept behind ResNet.

ResNet-50 is a CNN that is 50 layers deep. Figure 3 shows the architecture of ResNet50 with 34 residual layers (excerpted from [16]).

2.3. Evaluation Metrics

ML models cannot be used without measuring the efficiency of their algorithms. To evaluate ML and DL algorithms, certain statistical measures are utilized. These measures are calculated based on true positive (T.P.), true negative (T.N.), false positive (F.P.), and false negative (F.N.) possible outcomes. For example, the four outcomes of the model constituting a 2x2 confusion matrix for COVID-19 tests are displayed in Figure 4.



	COVID-19 (actural)	Non-COVID-19 (actural)
COVID-19 (predicted)	True Positive (TP): Number of persons with COVID-19 and tested positive	False Positive (FP): Number of persons without COVID-19 and tested positive
Non-COVID-19 (predicted)	False Negative (FN): Number of persons with COVID-19 and tested negative	True Negative (TN): Number of persons without COVID-19 and tested negative

Fig. 4 An example of a 2x2 confusion matrix made by a COVID-19 prediction model

The main measures utilized—accuracy, precision, sensitivity (recall), specificity, and an F1 score—are formally defined in Equations 1 to 5.

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

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

$$Sensitivity(Recall) = \frac{TP}{TP+FN} \tag{3}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4}$$

$$F1\ score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \tag{5}$$

To evaluate model performance, an area under the curve (AUC), also known as the area under the ROC curve (AUROC), metric is utilized. It is a number between 0.0 and 1.0 and measures the model's ability to separate positive and negative classes. The higher AUC, the better model's performance at differentiating between the positive and negative classes.

Finally, two popular measures are applied to assess the model's quality: mean squared error (MSE) and root mean square error (RMSE). Both measures are used to panelize the large prediction errors. However, RMSE is more widely used than MSE because it has the same units as the dependent variable Y-axis. The formal definition of MSE is given as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{6}$$

Where:

n is the number of examples in the dataset,
$\hat{y}$ is the predicted value for the i^{th} observation, and
y is the observed value for the i^{th} observation.

RMSE metric is a square root of the MSE and can be defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (7)$$

The following sections explore studies applying ML methods to predict COVID-19 severity, mortality, and vaccine effectiveness.

3. Overview of Works Related to COVID-19 Variants Severity

The COVID-19 pandemic has caused many disorders and upheavals in terms of health, economics, and society. From a scientific standpoint, a single pathology rarely mobilises much effort and resources. As a result, nearly 300,000 articles were indexed in PubMed between 2019 and 2022. This study investigates how innovative solutions relying heavily on digital technologies such as A.I., ML, and DL have emerged successfully.

COVID-19 symptoms, according to WHO, include fever, cough, and fatigue in mild cases and difficulty breathing in moderate cases. In severe cases, patients could suffer from acute pneumonia and organ failure and even face death. Pregnant women, elderly adults, especially those 60 years and older, and persons with medical illnesses, including diabetes, obesity, high blood pressure, cancer, or lung or heart problems, are more prone to experience severe sickness.

A Public Health Ontario [17] study tackles the impact and gravity of the Delta variant for children in the United States. The study shows an increase in pediatric hospitalizations related to COVID-19. Although pediatric hospitalization rates are increasing in the U.S., hospitalizations and COVID-19-related deaths in children remain low relative to clinical severity and COVID-19-related deaths in adults. Some hospitals and regions reported a higher absolute number of hospitalized pediatric COVID-19 cases. A crude analysis comparing pediatric hospitalization rates between Delta-dominant and non-Delta waves with similar case counts indicates a possible increase in hospitalizations due to COVID-19 caused by the Delta variant. The increase in the absolute number of pediatric hospitalizations in the United States may be the result of higher overall transmissibility of the Delta variant and high rates of community-based infection, leading to more pediatric cases and, therefore, pediatric hospitalizations.

To assist in identifying COVID-19 cases and stop its spread, authors in [18] are using online questionnaires to collect data (patients' signs and symptoms) and then using them as input for various prediction models, such as logistic regression (L.R.), SVM, and multi-layer perceptron (MLP). The authors use accuracy, sensitivity, specificity, and precision standard metrics to evaluate the models. Among all evaluated models, MLP shows the best accuracy with a value of 91.62%; therefore, it is the most useful ML model for COVID-19 detection based on signs and symptoms.

[19] developed a prediction model where data consisting of 151 COVID-19 patients from the Tumor Center of Union Hospital in Hubei, China, was collected. After selecting features, the author employed ANN algorithms to build a model that predicted and assessed the severity of COVID-19 in 151 samples. The created model accomplished a good estimation performance, with an F1 score of 96.4% and an AUC of 0.953 (0.889–0.982).

Using ML-based approaches for triage assessment and distinguishing between severe and non-severe COVID-19 is a technique used by [20]. The authors utilized a variety of ML models trained with tested-positive data from six locations in Switzerland. They relied on a set of parameters known as specialized laboratory markers. Sex, hemoglobin, glomerular filtration rate, sodium, C-reactive protein, blood glucose, and leucocytes are the most reliable indicators.

According to the study, the most successful ML model was the SVM, which achieved a median AUC value of 0.96. Using clinical symptoms and laboratory markers for diagnosing seriously ill patients increased the level of model complexity. Authors of [21] also employed the SVM approach to developing a model based on a dataset gathered from the Shanghai Public Health Clinical Center. The dataset is relatively small, containing 336 records. However, the model achieved excellent AUC values of 0.99 and 0.97 in the training and testing datasets, respectively.

To anticipate negative COVID-19 infection-related outcomes, [22] created a predictive score called the COVID-19 acuity score (CoVA) based on data from the Massachusetts General Hospital (MGH) and Emergency Department (E.D.). CoVA was developed to help frontline clinicians in out-patient screenings and used 98 variables, including vital signs, demographics, COVID-19 symptoms, and pre-existing diagnoses. The severity outcome of this model was the ordinal scale of negative events: no event, hospitalization, intensive care unit (ICU) and mechanical ventilation (MV), and death. Using the KNN algorithm, the developed model can be used in forecasting the risk of hospitalization prognosis. However, the collected data (e.g., medical images and vital signs) relied heavily on the hospital or office visit.

Additionally, it only included patients at MGH and E.D. Another predictor developed by [23] collected data from six hospitals in northeastern Pennsylvania. The study evaluated the risk based on age, gender, and historical variables. The developed L.R. model was used to develop the scoring system and performed well in predicting with an AUC value of 0.81.

Forecasting the required number of ICU beds is essential for resource allocation and patient management during the COVID-19 pandemic. A study by [24] created four separate ML classification models, namely random forest (R.F.), extreme gradient boosting (XGBoost), SVM, and L.R., that forecasted the need for hospitalization, admission to ICU, and MV with validation accuracies of 88%, 87%, and 86%, respectively. The prediction model was built based on patient characteristics, clinical symptoms, laboratory findings, and chest X-rays (CXR). While clinical features used in this study produced mixed results, blood laboratory and arterial gas analysis feature generally increased the predictability. Also, to predict ICU admission, authors of [25] developed a model based on classic machine learning classifiers (MLCs), including D.T., SVM, KNN, R.F., and ANN, and an ensemble learning technique, that aggregated the predictions of individual classifiers into a single prediction using a complete analysis data voting system. The results showed that the ANN has the lowest MSE with a value of 0.08 and the highest accuracy with a value of 0.97 over other models.

A survey conducted by [26] investigated different strategies and methods to prevent COVID-19. The study's authors stated that the only way to reduce virus transmission during the lockdown was to applicable government regulations, such as wearing masks, hygiene stations, and maintaining social distancing. They also implemented a real-time A.I. system using Raspberry Pi to monitor public spaces to ensure minimal human interaction. The authors applied an Extremely Randomized Trees Classifier (also known as an Extra Trees Classifier), which achieved an F1 score value of 98.2% and an accuracy value of 99.8%. They deployed a CNN-based face detection model to monitor the violation of wearing face masks. The same idea was applied by [27] on public transport in Morocco to reduce COVID-19 cases. A CNN algorithm with the aid of computer vision was applied to detect passengers with no masks. Detectors were also used to capture drivers' emotions and produce responses that can make drivers more comfortable. The model achieved good results, obtaining 91.23% accuracy with only 300 epochs. The authors used the MobileNetV2 DL model for mask detection, which provided a recognition rate of 97.96% during 50 epochs.

This study has limits. For example, it is difficult to compare the severity of different COVID-19 variants due to differences in transmissibility between strains, as well as variables such as screening practices, time of year when the different variants peaked, public health measures in place, adherence to public health measures at different times during the pandemic, and changes in vaccine coverage and effectiveness. Table 3 summarizes research papers exploiting ML methods to tackle COVID-19 severity. In the next section, we explore ML methods for predicting deaths caused by COVID-19.

Table 3. Summary of works on machine learning methods related to COVID-19 severity

Study	Objective/s	Data	Methodology	Results
E. Fayyoumi, et al., [18] (2020)	Predicting COVID-19 cases based on signs and symptoms.	Online questionnaire on April 2020 containing 105 samples.	Statistical model: LR and ML models: SVM and MLP	MLP achieved the best accuracy (91.62%) and specificity (93.75%), whereas the SVM achieved the best precision (93.75%).
J. Kang, et al., [19] (2021)	Predicting and evaluating the severity of COVID-19 based on clinical data.	The dataset containing 151 samples from the Tumor Center of Union Hospital was collected between Jan-Mar 2020.	ANN algorithms	The model achieved an F1 score value of 96.4% and an AUC value of 0.953 (0.889–0.982).
V. Schöning, et al., [20] (2021)	Building a model for patient triage. The model categorizes a patient's state as non-severe or severe based on a set of selected laboratory markers.	The dataset containing 198 samples (first wave) collected between Feb-Aug 2020 and 459 samples (second wave) from Insel Hospital Group Bern collected between Sep-Nov 2020.	L.R. and ML models: Decision Tree Induction (DTI), RF, KNN, AdaBoost, SVM, and MLP	The SVM was the best, achieving an AUC value of 0.96.
L. Sun, et al., [21] (2021)	Predicting COVID-19 cases.	The dataset contains 336 samples between Dec 2019 and Mar 2020.	SVM algorithm	The model achieved AUC values of 0.99 and 0.97 in the training and testing dataset, respectively.

H. Sun et al., [22] (2021)	Developing an outpatient screening predictive score (CoVA) that classifies the level of acuity based on COVID-19 symptoms and pre-existing diagnoses.	The dataset contains a development cohort (9,381 samples) and a prospective cohort (2,205 samples) collected between Mar-May 2020.	KNN algorithm	Model performance in prospective validation for hospitalization (expected-to-observed event ratio (E/O): 1.01; AUC: 0.76), critical illness (E/O: 1.03; AUC: 0.79), and death (E/O: 1.63; AUC: 0.93).
Z. Chen, et al., [23] (2021)	Determining whether a patient needs to be admitted to the hospital based on age, gender, and historical variables.	The dataset contains 6,485 COVID-19 samples from 6 hospitals in northeastern Pennsylvania.	L.R. algorithm	The model achieved an AUC value of 0.81.
B. Hao, et al., [24] (2020)	Forecasting the level of care needed based on clinical and laboratory data.	The dataset containing 2,566 COVID-19 samples from 5 Massachusetts hospitals was collected in Mar-Apr 2020.	RF, XGBoost, SVM, and L.R. models	The models achieved a validation accuracy of 0.88 for hospitalization, 0.87 for ICU care, and 0.86 for MV.
H. Ghandorh, et al., [25] (2021)	Predicting COVID-19 patients' ICU admission.	The dataset containing 639 records from Saudi National Health Laboratory collected between Mar 1, 2020, and Mar 31, 2020.	MLCs, including D.T., SVM, KNN, R.F., and ANN, and the ensemble learning method	The ANN had the lowest MSE with a value of 0.08 and the highest accuracy with a value of 0.97 over other models.
R. Sanjay, et al., [26] (2021)	Developing a real-time system using Raspberry Pi to monitor social distancing and face masks.	Internet of things devices send the sensors data.	Extremely Randomized Trees Classifier and a CNN-based face detection model	The model achieved an F1 score value of 98.2% and an accuracy value of 99.8%.
T. Karim, et al., [27] (2021)	Developing a CNN-based system to detect mask violations and drivers' emotions.	Emotion dataset containing 31,424 images downloaded from the FER2013 database [28], and the face mask dataset containing 1,485 images.	A CNN algorithm with the aid of computer vision. For face mask detection, MobileNetV2 was used	The model achieved an accuracy rate of 91.23%. The DL model achieved a recognition rate of 97.96%.

4. Overview of Works Related to Methods for COVID-19 Mortality

With the increasing spread of COVID-19, accurately predicting future patient outcomes or at least their vulnerability and risk of death from the disease is critical. A team from the Mount Sinai School of Medicine (New York) describes in [29] a new model for predicting mortality associated with COVID-19. Developed from the most extensive clinical data set to date and based on systematic ML, this prediction model achieved 0.91 accuracies. The ML framework uses L.R., SVM, R.F., and XGBoost for prediction. The model, however, is simply because it is based only on three clinical characteristics regularly collected, namely the patient's age, the minimum oxygen saturation during their admission medical assessment, and the nature of the consultation or initial contact (emergency, hospitalization, outpatient, or teleconsultation). This new evaluation technique could also allow a score to be re-evaluated regularly or even continuously for the patient, whether the latter is "remote monitoring" at home or hospitalized. The assessment could be integrated into the COVID-19 patient's care pathway and their data added to the health record. Clinical teams could also use the outcomes of the prediction model to track peak mortality risks throughout hospitalization. The development dataset consists of 3,841 patients, initially fed into the model for mortality prediction. Applied to other datasets of diagnosed and suspected patients (n = 961 and n = 249), the model validated its level of precision with an AUC value of 0.91. The patient's age, the minimum oxygen saturation on admission, and the type of first medical contact with the patient are sufficient factors to result in the patient's prediction. While predicting mortality in patients with COVID-19 who present a broad spectrum of health issues remains a challenge and hinders disease management, this model for predicting COVID-19 mortality allows a very precise prognostic approach based on simple and accessible clinical characteristics.

The clinical course of COVID-19 infection is very factor, with the overwhelming majority of patients recovering without challenges; however, some individuals cannot recover and eventually pass away. Appropriate and prompt supportive therapy can reduce mortality, and enhancing patient risk classification based on simple clinical data is critical for successful triage when healthcare systems are under strain. The study [30] adopted ML algorithms to create risk stratification and mortality prediction models based on clinical data from 544 COVID-19 patients in New Delhi, India. It compared XGBoost and L.R. algorithms for their predictive performance. The L.R. model generated the most outstanding results in predicting mortality (an F1 score value of 0.71), whereas the XGBoost classifier produced the best results in risk stratification (an F1 score value of 0.81). Biomarkers were developed to predict risk and mortality. The authors also compared the data to a comparable dataset with a Wuhan cohort of 375 people [31] to better understand the causes of India's significantly lower mortality rates. ML methods based on blood test data were proposed in this study to predict COVID-19 mortality risk. Lactate dehydrogenase (LDH), high-sensitivity C-reactive protein (hs-CRP), and lymphocytes form a strong combination.

In summary, the XGBoost classifier accurately predicted mortality with a 96.5% accuracy for each day the sickness persisted and a 90% accuracy for predictions made more than 10 days ahead. The obtained performance metrics are at a high level of confidence in the proposed model. Other prospective characteristics that predict capacity have been identified, but they require data from various sources to demonstrate their relevance and potentially improve the model.

In ML, boosting is a technique where several models are trained sequentially, i.e., each model learns from the mistakes of the previous one. Boosting algorithms can be applied using gradient-boosted decision trees (GBDT). This algorithm was applied by [32] to predict mortality cases caused by COVID-19. The study was conducted in 25 departments in Peru and revealed that Lima, Piura, Huánuco, and Ica departments had the highest death rates. The project was conducted using the Scrum approach, which makes it easy to adhere to the structure and solve the problem. The steps applied in this methodology are identifying the problem; using an analytical approach; gathering and interpreting the data; preparing the data; training, evaluating, and deploying the model; and finally, receiving feedback. The data was obtained from the national open data managed by the ministry of health.

A study using a similar ML method based on blood tests was conducted by [33] to predict mortality risk. The authors examined X-ray features and created a CNN-based DL model to identify COVID-19 patients. A dataset consisting of 5,493 non-COVID images and 3,914 COVID images was used in the proposed model. The model performed exceptionally well, achieving accuracies of 99.76%, 96.10%, and 96% in the training, validation, and test phases, respectively. Their model outperformed MobileNetV2, InceptionV3, and Xception, three state-of-the-art pre-trained models. Analyzing X-ray images to detect COVID-19 infections was also used by [34]. The stand-alone CNN model and hybrid ML models were compared. In this study, the VGG-16 served as a features extractor for the hybrid model, and a typical ML algorithm, e.g., SVM, R.F., or XGBoost, served as a classifier. The study showed no significant improvement when the hybrid model was applied. A high accuracy value of 99.82% and a sensitivity value of 100% were obtained by both VGG-16 and (VGG16+SVM) models. The study used 5 datasets, 4 of which were COVID-19-related datasets, with a total of 1,466 CXR and C.T. images, and 1 non-COVID-19 dataset with 5,863 CXR images.

After being discovered for the first time in the United Kingdom in September 2020, SARS-CoV-2 strain B.1.1.7 has since spread to several countries worldwide [35]. Several investigations have found B.1.1.7 to be more transmissible than previously known variants, although none have found that it increases disease severity. Data analysis from 17,452 COVID-19 deaths and 2,245,263 community SARS-CoV-2 positive tests conducted in England between September 1, 2020, and February 14, 2021, revealed an increase in COVID-19 mortality [36]. Due to mutations in this line that prevent polymerase chain reaction (PCR) amplification of the spike gene target, the presence or absence of B.1.1.7 was identified in 1,146,534 (51%) of these tests (s-gene target failure, SGTF). The authors concluded that the risk of mortality related to SGTF is 55% (95% confidence interval (CI): 39% to 72%) greater after accounting for age, sex, ethnicity, deprivation, nursing facility residency, and local authority of residency based on 4,945 fatalities with known SGTF.

In [13], the authors created a prediction model supporting A.I. and ML algorithms to assess the mortality risk and determine the health risk of COVID-19 patients when the SARS-CoV-2 virus elicited the illness. They employed a dataset of more than 2,670,000 laboratory-confirmed COVID-19 patients, with 307,382 labeled samples in their investigation. The ML model was proposed to help hospitals and other healthcare institutions prioritize patients when the system is overcrowded and determine who needs to be seen first and who has a higher priority for the hospital admission. The results showed an overall accuracy of 89.98% in forecasting death rates. The ML algorithms used were R.F., D.T., SVM, ANN, and L.R. To estimate the death rate in patients, LR and KNN were applied. Another study presented in [37] resorted to R.F. and k-means clustering to predict mortality associated with COVID-19. Due to the virus being widespread in the United States, it was imperative to pinpoint the populations most at risk of mortality due to infection. The study's authors collected data from the Centers for Disease Control and Prevention's (CDC) case surveillance [38] and discovered an unsurprising broad pattern. They found that under-attributed young people and older persons were unaffected by the high-risk group. Their results showed that clustered R.F. performance is better than the boosted R.F. method, with a 0.93 recall and a 0.93 F1 score.

To create and validate predictive models for mechanical requirements and COVID-19 mortality, four current automatic learning algorithms, three data equilibration methodologies, and a set of defining characteristics were used [39]. Complete blood counts (CBC), CXR, demographics, and clinical data for 5,739 patients with confirmed COVID-19 PCR at King Abdulaziz Medical City in Riyad, Saudi Arabia, were collected retrospectively from April 2 to June 18, 2020. However, only 1,508 and 1,513 of these patients met the criteria for admission in the ventilatory assistance program and the mortality criteria, respectively. The following were the outcomes: in a series of independent tests, the model for predicting ventilation needs with the 20 most essential characteristics selected using the ReliefF algorithm from the radiological, laboratory, and clinical data using the SVM technique achieved an AUC value of 86% and a balanced accuracy of 81%. The model's balanced accuracy for the mortality outcome was 80%, and its AUC value was 83%.

Some ML algorithms have been developed to forecast the probability of severe complications and mortality [13, 40-47]. The list of previous works and their characteristics related to COVID-19 mortality can be found in Table 4. This is critical because, during peaks, COVID-19 patients are becoming more common. The allocation and distribution of resources among patients based on their prognosis is a critical problem, and the question of prognosis is crucial. [48] established two models to predict COVID-19 patients' mortality based on clinical and laboratory characteristics.

Table 4. Summary of works on machine learning methods related to COVID-19 mortality

Study	Objective/s	Data	Methodology	Results
A. S. Yadaw, et al., [29] (2020)	Predicting mortality associated with COVID-19 based on clinical characteristics.	The dataset contains 3,841 records collected between Mar-Apr 2020.	L.R., SVM, R.F., and XGBoost approaches	The model achieved 91% accuracy.
S. Alle, et al., [30] (2022)	Developing risk stratification and death prediction models.	The dataset contains 544 records collected between Jun-Oct 2020.	XGboost and L.R. algorithms	XGBoost provided the best risk stratification results with an F1 score value of 0.81. L.R. provided the best results in predicting death with an F1 score value of 0.71.
L. Yan, et al., [31] (2020)	Predicting the mortality based on LDH, lymphocyte, and hs-CRP markers.	The dataset contains 375 blood samples collected between Jan-Feb 2020.	XGBoost algorithm	The model provided predictions with more than 90% accuracy for more than 10 days ahead and 95% accuracy for every day.
E. Ortega-Espinoza, et al., [32] (2022)	Using the Scrum methodology to predict COVID-19 deaths.	Datasets obtained from COVID-19 open data Ministry of Health-MINSA.	GBDT algorithm	Not provided
M. Masud, et al., [33] (2021)	Predicting COVID-19 cases by analyzing C.T. scan images.	A publicly available dataset consists of 5,493 non-COVID images and 3,914 COVID images.	Applied a CNN model and compared it with MobileNetV2, InceptionV3, and Xception models	The model achieved 99.76%, 96.10%, and 96% accuracy in the training, validation, and test phases, respectively.
W. Alawad, et al., [34] (2021)	Analyzing X-ray images to detect COVID-19 infections.	4 COVID-19 datasets with a total of 1,466 CXR and C.T. images (from Github) and a non-COVID-19 dataset containing 5,863 CXR images (from Kaggle).	Stand-alone CNN model and hybrid-ML models SVM, R.F., and XGBoost	The highest accuracy (99.82%) and sensitivity (100%) were obtained by VGG-16 and (VGG16+SVM) models, respectively.

M. Pourhomayoun and M. Shakibi, [13] (2021)	Evaluating the mortality risk and identifying the health risk of COVID-19 patients.	The dataset contains 2,670,000 confirmed COVID-19 cases representing 146 countries.	R.F., D.T., SVM, ANN, L.R., and KNN algorithms	The model achieved 89.98% overall accuracy.
E. Cornelius et al., [37] (2021)	Predicting mortality associated with COVID-19.	The dataset containing over 22 million occurrences of COVID-19 was collected between Jan 2020 to Mar 2021.	R.F., k-means clustering, and AdaBoost methods	Clustered R.F. performance was better than the boosted R.F. method, with a 0.93 recall and a 0.93 F1 score.
A. F. Aljouie, et al., [39] (2021)	Creating and validating predictive models for mechanical requirements and COVID-19 mortality.	The dataset contains 5,739 records of confirmed COVID-19 PCR collected between Apr-Jun 2020.	SVM algorithm	The model achieved an AUC of 86%, a balanced accuracy of 81% for predicting ventilation, an AUC of 83% and a balanced accuracy of 80% for the mortality outcome.
W. Liang, et al., [40] (2020)	Developing a risk score (COVID-GRAM) to predict the degree of patients' suffering.	The dataset contains 1,590 records from 575 hospitals collected from the China National Health Commission between Nov 2019 to Jan 2020.	L.R. algorithm. Statistical analysis conducted with R software	The model achieved a mean AUC of 0.88.
P. Gemmar [41] (2020)	Predicting mortality of COVID-19 patients from their biomarkers.	The same dataset of [30].	SVM and Sugenotype Fuzzy classifier (FIS). An artificial neural network (SOM) [49] was used for biomarker selection.	The model achieved AUC scores up to 97.84%.
M. E. H. Chowdhury et al., [42] (2021)	Predicting mortality risk among COVID-19 patients.	The same dataset of [30].	XGBoost algorithm	The model achieved AUC scores of 0.96 for the derivation cohort and 0.99 for the validation cohort.
S. Bolourani, et al., [43] (2021)	Providing a 48-hour forecast of respiratory failure.	The dataset containing 11,525 records from Northwell Health acute care hospitals was collected between Mar-May 2020.	XGBoost and L.R. algorithms	The XGBoost model achieved the best mean accuracy with a value of 0.919 and an AUC with a value of 0.77.
E. Jimenez-Solem, et al., [44] (2021)	Estimating risks at various management stages.	The dataset containing 5,594 records from the Capital and Zealand Regions, Denmark, was collected between Mar-Jun 2020.	R.F. model	The model predicted hospital admission, ICU admission, ventilator treatment, and death AUC scores of 0.820, 0.802, 0.815, and 0.902, respectively.
K. Ikemura, et al., [45] (2021)	Examining several ML models for predicting a patient's propensity to survive a COVID-19 infection.	The dataset containing 4,376 records from the University Hospital for Albert Einstein Clg. of Medicine was collected in Mar-Jul, 2020.	Used an autoML open-source tool that generated 20 ML models and ranked them	The stacked ensemble model achieved the best area under the precision-recall curve with a value of 0.836.

D. Ji, et al., [46] (2020)	Developing a scoring model named CALL to help clinicians make better decisions on therapeutic strategies.	The dataset containing 208 records from Fuyang Second People's Hospital or the Fifth Medical Center of Chinese People's Liberation Army General Hospital were collected between Jan-Feb 2020.	Multivariate Cox regression model	The model achieved an AUC value of 0.91.
A. Karthikeyan, et al., [47] (2021)	Predicting COVID-19 mortality risk based on blood tests.	The dataset contains 2,779 records from Tongji Hospital in Wuhan.	Neural networks, LR, XGBoost, RF, SVM, and D.T. models	The neural network model achieved the best accuracy value of 96.53% and an F1 score value of 0.969.
K. Wang, et al., [48] (2020)	Developing a mortality prediction model for patients with COVID-19.	The dataset containing 296 records from the First People's Hospital of Jiangxia District in Wuhan was collected between Jan-Feb 2020.	L.R. and XGBoost models	The model achieved an AUC value of 0.88, a sensitivity value of 92.31%, and a specificity value of 77.44%.

5. Overview of Machine Learning's Impact on COVID-19 Vaccines' Efficacy

In this section, we present a comparison between available COVID-19 vaccines. The development of immunizations against the virus responsible for SARS-CoV-2 began as soon as the viral genome sequence was disclosed. This development has progressed at an unprecedented pace, with the first clinical trial conducted shortly after the disease spread in March 2020. A year later, emergency processes were used to approve a dozen vaccines based on various concepts, some of which had not been tested beyond clinical trials. The rapid development of vaccines in response to the pandemic of COVID-19 has been an unprecedented success. Observations made following natural infections or vaccinations have shown that neutralizing antibodies is essential for protection against the disease. The level of anti-virus antibodies necessary for optimal protection is not yet clearly defined, and the persistence over time of these antibodies is not yet known. The messenger ribonucleic acid (mRNA), which represents a first in the design of new vaccines, is effective against symptomatic SARS-CoV-2 infection, as shown by phase III clinical trials and real-life observations. Recent studies show that it also protects against asymptomatic infection and, more generally, against infection. Vaccines' impact on the pandemic depends on the vaccinated population and their effectiveness at providing protection (according to comorbidities, age, pregnancy, or lactation) [50].

In Figure 5, we present the total number of vaccinated people worldwide. At the beginning of vaccinations, the worldwide population was not confident regarding the efficiency of vaccinations. Figure 5 shows that the number of people vaccinated increased after June 2021. It means that people have become more confident about vaccines and their efficiencies in fighting COVID-19 [51].

Several studies have examined the effectiveness of various vaccinations on the transmission and fatality rates of COVID-19 strains. At the beginning of the COVID-19 crisis, vaccinations, BNT162b (Pfizer), mRNA-1273 (Moderna), and ChAdOx1 (AstraZeneca) vaccines were very effective in fighting COVID-19 [69] and reduced the transmission of the virus as well as the number of deaths. The susceptible, infected recovered vaccination1 vaccination2 death (SIRVVD) model was used in the study to establish the target vaccination rate ($P_2^{obj.}$) for reducing Japan's COVID-19 Delta variant (B.1.617) infection levels.

Figure 6 illustrates the total number of fully vaccinated people in the United Kingdom, India, Brazil, and South Africa. As illustrated in this figure, the number of fully vaccinated people in India is unstable. The maximum number of fully vaccinated people in India was obtained in September 2021. Overall, vaccination reduced the number of deaths and mortality rates in several countries.

Efficacy and safety are two significant success factors of COVID-19 vaccines that have been studied. For example, [53] analyzed 123 datasets, including 58,889 individuals receiving a SARS-CoV-2 vaccine and 46,638 receiving a placebo. A placebo, as defined by [54], is an inert material resembling the studied medicine or treatment. A total of 34,041 cases were included in the study's phase II/III randomized clinical trials (RCTs), and the results showed that the mRNA-based COVID-19 vaccines had an efficacy rate of 94.6%.

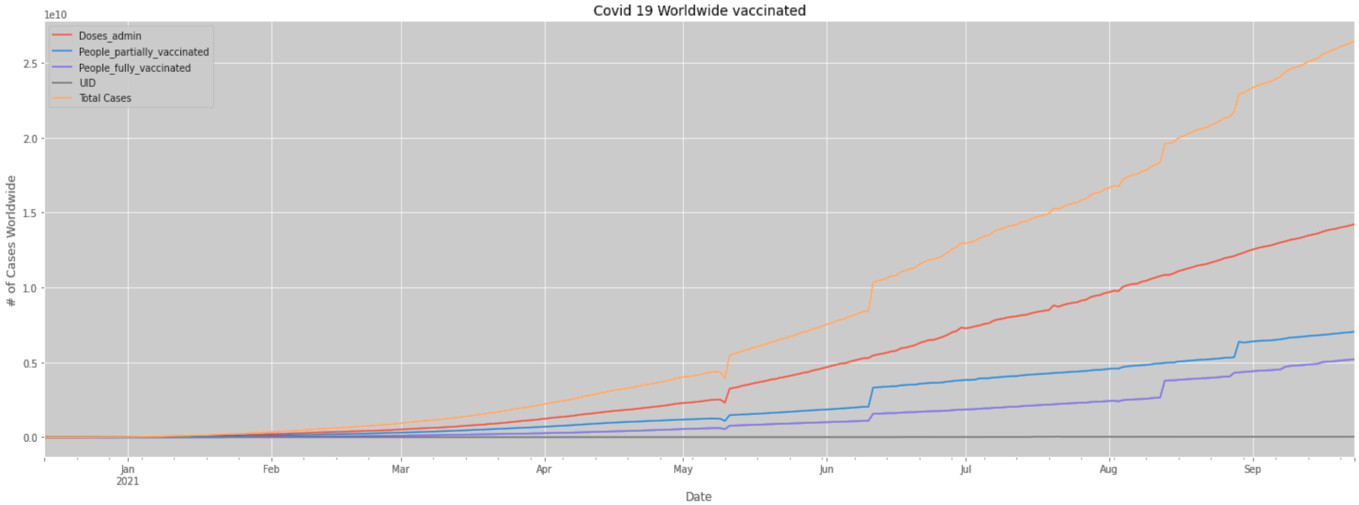


Fig. 5 Total number of vaccinated people worldwide

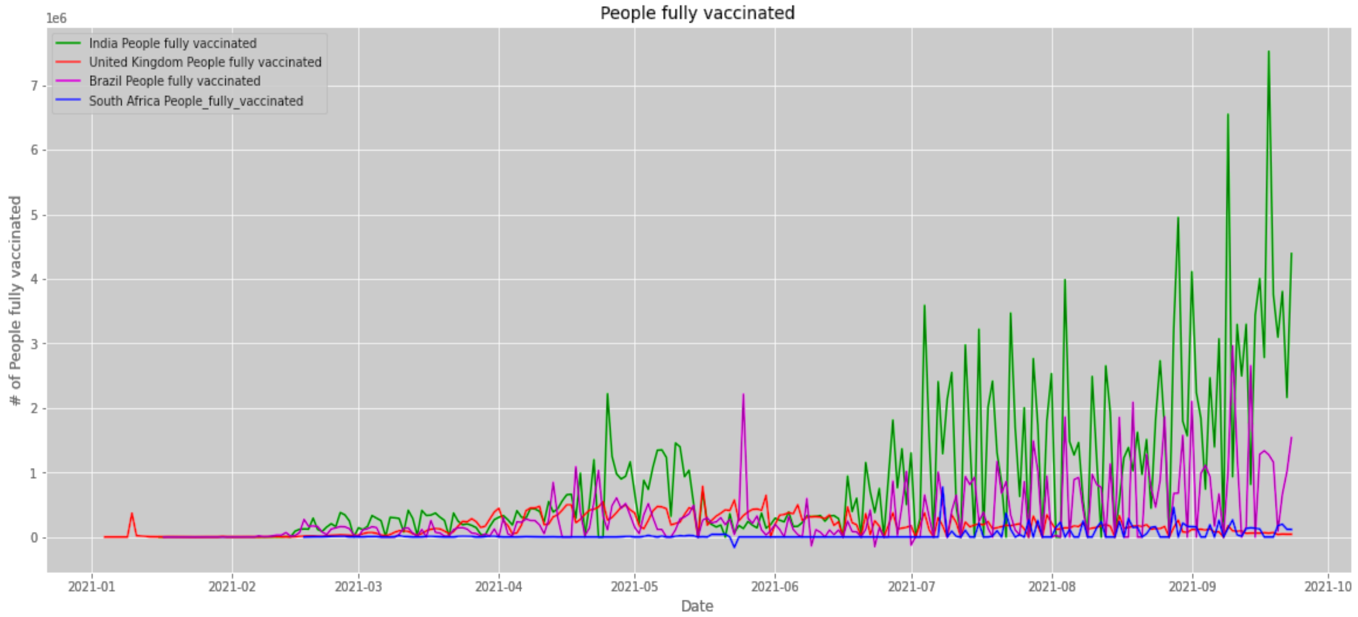


Fig. 6 Total number of fully vaccinated people in the United Kingdom, India, Brazil, and South Africa

It also concluded that except for diarrhea and arthralgia, the mRNA-based vaccinations exhibited the highest level of reported adverse effects (A.E.s). Efficacy and safety were also discussed by [55], in which the authors analyzed both clinical trials and real-world data managed by the CDC in the United States. According to the study, mRNA-based vaccinations had the most remarkable efficacy at 94.29%. It also identified 36 types of adverse drug reactions (e.g., pain, swelling, fever, fatigue, and headache) in clinical trials and real-world data.

A study conducted by [56] used two programs based on ML, OptiVax and EvalVax, that worked in tandem. OptiVax is an open-source implementation of the proposed model, and EvalVax is the vaccine evaluation tool. The OptiVax results suggest that the SRAS-CoV-2 spike protein alone may not be sufficient to provide complete immunization to all racially distinct ethnic groups. According to the research, adding specific peptides to the immune system can improve its response.

In another study's [57] RCT, the Moderna vaccine was reported to have a 94.1% efficacy in avoiding COVID-19 symptomatic infection caused by "wild-type" variants. The genuine effectiveness of this vaccination against the SRAS-CoV-2 pre-infectious variants B.1.1.7 (Alpha) and B.1.351 (Beta) was examined by authors in Qatar, a country with a large working-age population. They found that the vaccine's efficacy was negligible in the first two weeks following delivery but significantly increased in the third and fourth weeks before the delivery of a second dose. Fourteen days after the first dose but before the second dose, the effectiveness against B.1.1.7 infection was 88.1% (95% CI: 83,7% to 91,5%), and fourteen days after the second dose, it was 100% (95% CI: 91.8% to 100.0%). To confirm the main findings, they provided additional sensitivity analyses, including 1) matching by nationality, sex, age, and PCR testing date, 2) L.R. adjusting for a calendar week, and 3) L.R. adjusting for matching factors, i.e., nationality, sex, age, and reason for PCR testing.

Participants in clinical trials of the 2-dose mRNA vaccinations reported local and systemic responses called reactogenicity. Reactions more frequently reported include injection site pain, headache, and fatigue [58]. The study by [59] employed a machine learning-fitted penalized linear regression model to correlate A.E.s with antibody levels. The dataset had 80 subjects, 25 males and 55 females. The study's analytical findings indicate that headache, malaise, and nausea A.E.s demonstrate the highest variable importance and have the most favorable impact on SARS-CoV-2 antibody levels.

[60] highlighted different ML models to predict vaccines' immunogenicity, efficacy, or reactogenicity to fight pandemics such as tuberculosis, malaria, and more recent ones like Zika and SARS-CoV-2. Different metrics, such as accuracy, AUC, or RMSE, for regressions were used to assess the quality of the discussed ML models. However, the study did not focus on ML prediction models for SARS-CoV-2 vaccines.

[61] and [62] also applied linear regression ML to predict vaccine immunogenicity. While [61] discussed the role of innate responses in vaccine immunogenicity in general, the authors in [62] focused on SARS-CoV-2 vaccination in Seoul National University Hospital between March and April 2021, including 42 adults vaccinated with AstraZeneca and 93 adults vaccinated with Pfizer. The multivariate linear regression model was deployed to evaluate the correlation between the anti-S1 IgG O.D. ratio in a post-vaccination serum and local (e.g., injection site discomfort and edema) or systemic (e.g., headache, muscle pain, weariness, and fever) reactogenicity. The study concluded no association between reactogenicity and immunogenicity.

In COVID-19 pharmacology, ML-based algorithms are helpful since they offer a set of tools that enhance the drug research and development process for particular conditions with the use of reliable and high-quality data currently available.

Algorithms such as SVM, ANNs, and DL are employed to develop vaccines. For instance, [63] created automated drug development pipelines for drug discovery, testing, and repurposing. The study focused on sequential learning algorithms and recommender systems and discussed the relativeness of datasets and ML applicability for drug development. Data availability, data quality, and feature selection are key roles in ML model success. The other study's [64] main goal was to identify prospective vaccine candidates.

Several techniques were used, such as reverse vaccinology, immunoinformatics, and DL, specifically LSTM. The labeled dataset, consisting of 100 proteins, was extracted using data mining techniques, and the output of the study was a framework for identifying possible vaccine candidates and producing an epitope-based vaccine against SARS-CoV-2. The authors of [65] described how they applied an "In Silico" analysis to develop an effective multi-epitope peptide vaccination against SARS-CoV-2.

The screening of potential epitopes was done by an online server called CLP Pred that combined quantitative matrix (Q.M.), ANN, and SVM models. According to the authors, combining prediction techniques was more precise and sensitive than utilizing ANN and SVM alone. Works related to COVID-19 vaccines' efficacy are summarized in Table 5.

6. Discussion

This study has potential limitations. It is difficult to try to compare the severity of different COVID-19 variants due to differences in transmissibility between strains, as well as variables such as screening practices, time of year when the different variants peaked, public health measures in place, adherence to public health measures at different times during the pandemic, and changes in vaccine coverage and effectiveness.

From studying techniques to predict severity, mortality, and efficacy, we found that the main problem with the Delta variant was its transmissibility, not its mortality rate. For the Gamma variant, which appeared in November 2020, the mortality rate for the United Kingdom was 3.58%, 2.73% for Brazil and South Africa, and 1.45% for India.

It is also important to mention that this rate could be significant in terms of numbers since the population in India (1.38 billion in 2020) is very big compared to the population in Brazil (212.6 million), the U.K. (67.22 million), and South Africa (59.31 million).

The decrease in the mortality rate could also be explained by the fact that several countries adopted many strategies when fighting COVID-19. As of October 11, 2021, 6.56 billion people have received at least one vaccine dosage, and 2.79 billion were fully immunized, making up 35.8% of the global population [68].

Table 5. Summary of works on machine learning methods related to COVID-19 vaccines' efficacy

Study	Objective/s	Data	Methodology	Results
A. Pormohammad, et al., [53] (2021)	A systematic review to estimate vaccine efficacy, side effects, and immunogenicity.	Clinical trial data contain 58,889 cases collected from 25 publications (123 datasets) as of March 12, 2021.	Analysis was performed using Comprehensive Meta-Analysis Software Version 2.0	The mRNA-based COVID-19 vaccines had 94.6% efficacy in a total of 34,041 cases, and except for diarrhea and arthralgia, the mRNA-based vaccinations exhibited the highest level of reported A.E.s.
C. Cai, et al., [55] (2021)	Analyzing the efficacy and safety of COVID-19 vaccines.	Clinical trial data (194,015 cases) and real-world data (11,936 cases) were managed by the CDC and Food and Drug Administration.	GraphPad Prism was utilized for statistical analyses, and the R Statistical Software was utilized for meta-analysis	mRNA-based vaccines had the highest efficacy of 94.29%.
G. Liu, et al., [56] (2020)	Evaluating vaccine formulations for SARS-CoV-2.	29,403 candidate peptides for MHC class I, and 125,593 candidate peptides for MHC class II.	OptiVax and EvalVax ML models	Adding specific peptides to the immune system improved its response.
H. Chemaitelly et al., [57] (2021)	Evaluating the real-world effectiveness of the mRNA-1273 vaccine (Moderna) against B.1.1.7 (Alpha) and B.1.351 (Beta).	256,037 individuals received at least 1 dose, and 181,304 individuals completed 2 doses between Dec 2020 and May 2021.	Test-negative, case-control study design	The Moderna vaccine had a 94.1% efficacy in preventing symptomatic COVID-19 infection due to infection by "wild-type" variations.
J. Held, et al., [59] (2021)	Examining the relationship between the severity of A.E.s and the anti-SARS-CoV-2 spike protein antibody response.	The dataset contains 80 employees from the Institute for Clinical Microbiology, Immunology and Hygiene of the University Hospital in Erlangen, Germany.	Linear regression model	Results showed that headache, malaise, and nausea A.E.s demonstrate the highest variable importance and positively influence the SARS-CoV-2 antibody levels.
W. Burny, et al., [61] (2017)	Analyzing the role of innate responses in vaccine immunogenicity in general.	A randomized, controlled phase II trial was performed at 14 study centers with 375 participants.	Linear regression model and R Statistical Software	Results revealed correlations between adaptive responses and specific traits of the innate response post-dose 2.
Y. H. Hwang, et al., [62] (2021)	Evaluating the correlation between local or systemic reactogenicity and the anti-S1 IgG O.D. ratio in a post-vaccination serum.	42 vaccinated adults from Seoul National University Hospital and Seoul National University Bundang Hospital between Mar-Apr 2021	Linear regression model	Results revealed no association between reactogenicity and immunogenicity.
C. Réda, et al., [63] (2020)	A survey of drug development data and ML methods.	19 datasets related to drug development and grouped into 5 types of genomic data, interaction data, drug-disease associations, clinical trials, and chemical and drug data.	Sequential learning algorithms and recommender systems	The data availability, data quality, and feature selection played key roles in the ML model's success.

B. A. Abbasi, et al., [64] (2020)	Identifying prospective vaccine candidates against SARS-CoV-2 coronavirus.	Dataset collected from database resources of the National Center for Biotechnology Information (NCBI) and crystal structures of human alleles collected from the Protein Data Bank (PDB).	LSTM as a DL algorithm	The output of the study was a framework for the identification of possible vaccine candidates and the production of an epitope-based vaccine against SARS-CoV-2
Z. Yazdani, et al., [65] (2020)	Creating a multi-epitope peptide vaccine against SARS-CoV-2 coronavirus using an "In Silico" analysis.	The protein sequences were managed by using the NCBI. The epitopes major histocompatibility complex class I (MHC-I), cytotoxic T lymphocytes (CTL), human leukocyte antigen (HLA)-II, and linear B-cell were collected by using the IEDB database [66] and the RANKPEP online server [67].	Combined Q.M., ANN, and SVM models	Predicting vaccine candidates was possible by applying reverse vaccinology. However, the "In Silico" analysis required verification.

Within the scope of this survey, A.I. and ML contributions to this research were inspired by their relativeness and originality. First, the adaptation of ML algorithms for measuring COVID-19 severity and controlling the spread of this disease is discussed.

Diagnosing COVID-19 cases based on a prediction score was essential since resources at the pandemic's peak were scarce. Most research focused on determining the seriousness of the case, i.e., whether it required immediate response by hospital admission, ICU, and MV.

Second, ML algorithms related to predicting the mortality rate of patients with COVID-19 were explored. Most works in this part resorted to XGBoost because it is one of the most popular techniques used in gradient boosting frameworks.

In most studies, it is fast, scalable, and performs better than other ML algorithms, i.e., SVM, L.R., or R.F. The literature also varied in selecting features, e.g., C.T. scan images, biomarkers, demographics, blood tests, or a combination. The methodology of selecting features to train the model also varied.

Last, the investigation was conducted to study different ML proposals to solve vaccine effectiveness and development problems. The primary concern of most studies were side effects (or so-called A.E.s) and immunogenicity.

On the other hand, few works focused on the automation of the drug development process to accelerate the production of vaccine candidates. The datasets in these studies were either clinical trial data, real-world data, or both.

7. Conclusion

This review aimed to monitor COVID-19 variants' impact on transmissibility, disease severity, and vaccine effectiveness. In addition, it presented an overview of the related works on disease severity, mortality predictions, and COVID-19 variant vaccines. In summary, this research aimed to highlight existing models and approaches in handling the COVID-19 pandemic using ML and DL technologies.

Distinct approaches were developed and tested to forecast the mortality or severity of COVID-19. The use of each approach primarily depended on the dataset. For example, a linguistic approach was applied when having an excellent lexicon to rely on it. ML and DL approaches were utilized because they provide more precise results and present a better performance.

Finally, we presented a comparison between different COVID-19 vaccines and concluded by providing a summary of approaches that address the issue of whether vaccines made by Moderna, Pfizer, Novavax, Johnson & Johnson, and AstraZeneca increase immunization rates and decrease COVID-19 variations.

In the future, the following tasks will be conducted. We will present a comparative analysis and evaluation of each variant's impact among the selected countries, and we will detect and compare the phases of spread for each variant among the selected countries. Finally, the obtained results will be used later to avoid the spread in other countries.

Funding Statement

The funding of this research comes from a personal fund.

References

[1] COVID-19 Coronavirus Pandemic, 2022. [Online] Available: https://www.worldometers.info/coronavirus/

[2] World Health Organization, 2022. [Online] Available: https://www.who.int/

[3] B. G. Marin, G. Aghagoli, K. Lavine, L. Yang, E. J. Siff, S. S. Chiang, T. P. Salazar-Mather, L. Dumenco, M. C. Savaria, S. N. Aung, T. Flanigan, and I. C. Michelow, "Predictors of COVID-19 Severity: A Literature Review," *Reviews in Medical Virology*, vol. 31, no. 1, pp. 1-10, 2021. Crossref, http://doi.org/10.1002/rmv.2146

[4] M. T. Huyut, Z. Huyut, F. İlkbahar, and C. Mertoğlu, "What is the Impact and Efficacy of Routine Immunological, Biochemical and Hematological Biomarkers as Predictors Of COVID-19 Mortality?," *International Immunopharmacology*, vol. 105, 2022. Crossref, http://doi.org/10.1016/j.intimp.2022.108542

[5] R. Chen, L. Sang, M. Jiang, Z. Yang, N. Jia, W. Fu, J. Xie, W. Guan, W. Liang, Z. Ni, Y. Hu, L. Liu, H. Shan, C. Lei, Y. Peng, L. Wei, Y. Liu, Y. Hu, P. Peng, J. Wang, J. Liu, Z. Chen, G. Li, Z. Zheng, S. Qiu, J. Luo, C. Ye, S. Zhu, J. Zheng, N. Zhang, Y. Li, J. He, J. Li, S. Li, and N. Zhong, "Longitudinal Hematologic and Immunologic Variations Associated with the Progression of COVID-19 Patients in China," *Journal of Allergy and Clinical Immunology*, vol. 146, no. 1, pp. 89-100, 2020. Crossref, http://doi.org/10.1016/j.jaci.2020.05.003

[6] R. Vaishya, M. Javaid, I.H. Khan, and A. Haleem, "Artificial Intelligence (A.I.) Applications for COVID-19 Pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 337-339, 2020. Crossref, https://doi.org/10.1016/j.dsx.2020.04.012

[7] E. Christodoulou, A. Gregoriades, M. Pampaka, and H. Herodotou, "Combination of Topic Modeling and Decision Tree Classification for Tourist Destination Marketing," *In International Conference on Advanced Information Systems Engineering, Springer*, Cham, pp. 95-108, 2020. Crossref, https://doi.org/10.1007/978-3-030-49165-9_9

[8] W. Xu, N. N. Sun, H. N. Gao, Z. Y. Chen, Y. Yang, B. Ju, and L. L. Tang, "Risk Factors Analysis of COVID-19 Patients with ARDS and Prediction Based on Machine Learning," *Scientific Reports*, vol. 11, no. 2933, pp. 1-12, 2021. Crossref, https://doi.org/10.1038/s41598-021-82492-x

[9] J.R. Quinlan, “Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986. Crossref, https://doi.org/10.1007/BF00116251

[10] J.R. Quinlan, "C4.5: Programs for Machine Learning," *Elsevier*, 2014. Crossref, https://doi.org/10.1007/BF00993309

[11] C5.0: An Informal Tutorial, 2019. [Online]. Available: https://www.rulequest.com/see5-unix.html

[12] V. Vapnik, “The Nature of Statistical Learning Theory,” *Springer Science & Business Media*, 1999.

[13] M. Pourhomayoun, and M. Shakibi, "Predicting Mortality Risk in Patients with COVID-19 Using Machine Learning to Help Medical Decision-Making," *Smart Health*, vol. 20, no. 100190, 2021. Crossref, https://doi.org/10.1016/j.smhl.2020.100178

[14] M. S. Farooq, A. Arooj, R. Alroobaea, A. M. Baqasah, M. Y. Jabarulla, D. Singh, and R. Sardar, "Untangling Computer-Aided Diagnostic System for Screening Diabetic Retinopathy Based on Deep Learning Techniques," *Sensors*, vol. 22, no. 5, 2022. Crossref, https://doi.org/10.3390/s22051803

[15] “Entity Extraction Using Deep Learning Based on Guillaume Genthial Work on NER,” 2018. [Online] Available: https://medium.com/p/8014acac6bb8

[16] An Overview of ResNet and its Variants, 2017. [Online] Available: https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035

[17] COVID-19: Severity of the Delta (B.1.617.2) Variant in Children, 2021. [Online]. Available: https://www.publichealthontario.ca/-/media/documents/ncov/voc/2021/09/covid-19-severity-delta-children.pdf?la=en

[18] E. Fayyoumi, S. Idwan, and H. AboShindi, "Machine Learning and Statistical Modeling for Prediction of Novel COVID-19 Patients Case Study: Jordan," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, 2020. Crossref, https://doi.org/10.14569/IJACSA.2020.0110518

[19] J. Kang, T. Chen, H. Luo, Y. Luo, G. Du, and M. Jiming-Yang, "Machine Learning Predictive Model for Severe COVID-19," *Infection, Genetics and Evolution*, vol. 90, 2021. Crossref, https://doi.org/10.1016/j.meegid.2021.104737

[20] V. Schöning, E. Liakoni, C. Baumgartner, A. K. Exadaktylos, W. E. Hautz, A. Atkinson, and F. Hammann, "Development and Validation of a Prognostic COVID-19 Severity Assessment (COSA) Score and Machine Learning Models for Patient Triage at a Tertiary Hospital," *Journal of Translational Medicine*, vol. 19, no. 1, pp. 1-11, 2021. Crossref, https://doi.org/10.1186/s12967-021-02720-w

[21] L. Sun, F. Song, N. Shi, F. Liu, S. Li, P. Li, W. Zhang, X. Jiang, Y. Zhang, L. Sun, X. Chen, and Y. Shi, "Combination of Four Clinical Indicators Predicts the Severe/Critical Symptom of Patients Infected COVID-19," *Journal of Clinical Virology*, vol. 128, 2020. Crossref, https://doi.org/10.1016/j.jcv.2020.104431

[22] H. Sun, A. Jain, M. J. Leone, H. S. Alabsi, L. N. Brenner, E. Ye, W. Ge, Y. Shao, C. L. Boutros, R. Wang, R. A. Tesh, C. Magdamo, S. I. Collens, W. Ganglberger, I. V. Bassett, J. B. Meigs, J. Kalpathy-Cramer, M. D. Li, J. T. Chu, M. L. Dougan, L. W. Stratton, J. Rosand, B. Fischl, S. Das, S. S. Mukerji, G. K. Robbins, and M. B. Westover, "CoVA: An Acuity Score for Outpatient Screening that Predicts Coronavirus Disease 2019 Prognosis," *The Journal of Infectious Diseases*, vol. 223, no. 1, pp. 38-46, 2021. Crossref, https://doi.org/10.1093/infdis/jiaa663

[23] Z. Chen, N. W. Russo, M. M. Miller, R. X. Murphy, and D. B. Burmeister, "An Observational Study to Develop a Scoring System and Model to Detect Risk of Hospital Admission Due to COVID-19," *Journal of the American College of Emergency Physicians Open*, vol. 2, no. 2, 2021. Crossref, https://doi.org/10.1002/emp2.12406

[24] B. Hao, S. Sotudian, T. Wang, T. Xu, Y. Hu, A. Gaitanidis, K. Breen, G. C. Velmahos, and I. C. Paschalidis, "Early Prediction of Level-of-Care Requirements in Patients with COVID-19," *Elife*, vol. 9, 2020. Crossref, https://doi.org/10.7554/eLife.60519

[25] H. Ghandorh, M. Z. Khan, R. Alsufyani, and M. Khan, "An ICU Admission Predictive Model for COVID-19 Patients in Saudi Arabia," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021. Crossref, https://doi.org/10.14569/IJACSA.2021.0120764

[26] R. Sanjay, S. G. Rajan, C. Supraja, H. Veena, M. Natesh, and N. L. Vidya, "A Survey on Prevention from Emerging Viruses," *International Journal of Computer Trends and Technology*, vol. 69, no. 4, pp. 54-62, 2021. Crossref, https://doi.org/10.14445/22312803/IJCTT-V69I4P110

[27] T. Karim, R. Nihal, B. Sanaa, E. B. Abderrahmane, E. A. Ayoub, Z. Omar, and F. Abdelmajid, "CNN Applied in Public Transport for the Protection Against the COVID-19 Spread," *International Journal of Engineering Trends and Technology*, vol. 69, no. 10, pp. 33-37, 2021. Crossref, https://doi.org/10.14445/22315381/IJETT-V69I10P205

[28] Facial Expression Recognition (FER) Challenge, 2013. [Online]. Available: https://www.kaggle.com/datasets/ashishpatel26/facial-expression-recognitionferchallenge

[29] A. S. Yadaw, Y. Li, S. Bose, R. Iyengar, S. Bunyavanich, and G. Pandey, "Clinical Features of COVID-19 Mortality: Development and Validation of a Clinical Prediction Model," *The Lancet Digital Health*, vol. 2, no. 10, pp. E516-E525, 2020. Crossref, https://doi.org/10.1016/S2589-7500(20)30217-X

[30] S. Alle, A. Kanakan, S. Siddiqui, A. Garg, A. Karthikeyan, P. Mehta, N. Mishra, P. Chattopadhyay, P. Devi, S. Waghdhare, A. Tyagi, B. Tarai, P. P. Hazarik, P. Das, S. Budhiraja, V. Nangia, A. Dewan, R. Sethuraman, C. Subramanian, M. Srivastava, A. Chakravarthi, J. Jacob, M. Namagiri, V. Konala, D. Dash, T. Sethi, S. Jha, A. Agrawal, R. Pandey, P. K. Vinod, and U. D. Priyakumar, "COVID-19 Risk Stratification and Mortality Prediction in Hospitalized Indian Patients: Harnessing Clinical Data for Public Health Benefits," *PLoSOne*, vol. 17, no. 3, 2022. Crossref, https://doi.org/10.1371/journal.pone.0264785

[31] L. Yan, H. T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, X. Huang, Y. Xiao, H. Cao, Y. Chen, T. Ren, F. Wang, Y. Xiao, S. Huang, X. Tan, N. Huang, B. Jiao, C. Cheng, Y. Zhang, A. Luo, L. Mombaerts, J. Jin, Z. Cao, S. Li, H. Xu, and Y. Yuan, "An Interpretable Mortality Prediction Model for COVID-19 Patients," *Nature Machine Intelligence*, vol. 2, no. 5, pp. 283-288, 2020. Crossref, https://doi.org/10.1038/s42256-020-0180-7

[32] E. Ortega-Espinoza, M. Flores-Cruz, D. C. Loayza, J. Velarde-Cuadros, A. Delgado, and E.L. Huamaní, "," *International Journal of Engineering Trends and Technology*, vol. 70, no. 3, pp. 48-53, 2022. Crossref, https://doi.org/10.14445/22315381/IJETT-V70I3P206

[33] M. Masud, M. D. Alshehri, R. Alroobaea, and M. Shorfuzzaman, "Leveraging Convolutional Neural Network for COVID-19 Disease Detection Using C.T. Scan Images," *Intelligent Automation and Soft Computing*, vol. 29, no. 1, pp. 1-13, 2021. Crossref, https://doi.org/10.32604/iasc.2021.016800

[34] W. Alawad, B. Alburaidi, A. Alzahrani, and F. Alflaj, "A Comparative Study of Stand-Alone and Hybrid CNN Models for COVID-19 Detection," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021. Crossref, https://doi.org/10.14569/IJACSA.2021.01206102

[35] N. G. Davies, C. I. Jarvis, W. J. Edmunds, N. P. Jewell, K. Diaz-Ordaz, and R. H. Keogh, "Increased Mortality in Community-Tested Cases of SARS-Cov-2 Lineage B.1.1.7," *Nature*, vol. 593, no. 7858, pp. 270-274, 2021. Crossref, https://doi.org/10.1038/s41586-021-03426-1

[36] S. M. Treibert, "The Severe Acute Respiratory Syndrome Corona Virus Type 2," *In Mathematical Modeling and Nonstandard Schemes for the Corona Virus Pandemic*, Best Masters. Springer Spektrum, Wiesbaden, pp. 9-63, 2021. Crossref, https://doi.org/10.1007/978-3-658-35932-4_2

[37] E. Cornelius, O. Akman, and D. Hrozencik, "COVID-19 Mortality Prediction Using Machine Learning-Integrated Random Forest Algorithm Under Varying Patient Frailty," *Mathematics*, vol. 9, no. 17, 2021. Crossref, https://doi.org/10.3390/math9172043

[38] COVID-19 Case Surveillance Public Use Data website, 2022. [Online]. Available: https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf

[39] A. F. Aljouie, A. Almazroa, Y. Bokhari, M. Alawad, E. Mahmoud, E. Alawad, A. Alsehawi, M. Rashid, L. Alomair, S. Almozaai, B. Albesher, H. Alomaish, R. Daghistani, N. K. Alharbi, M. Alaamery, M. Bosaeed, and H. Alshaalan, "Early Prediction of COVID-19 Ventilation Requirement and Mortality from Routinely Collected Baseline Chest Radiographs, Laboratory, and Clinical Data with Machine Learning," *Journal of Multidisciplinary Healthcare*, vol. 14, pp. 2017-2033, 2021. Crossref, https://doi.org/10.2147/JMDH.S322431

[40] W. Liang, H. Liang, L. Ou, B. Chen, A. Chen, C. Li, Y. Li, W. Guan, L. Sang, J. Lu, Y. Xu, G. Chen, H. Guo, J. Guo, Z. Chen, Y. Zhao, S. Li, N. Zhang, N. Zhong, and J. He, "Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients with COVID-19," *JAMA Internal Medicine*, vol. 180, no. 8, pp. 1081-1089, 2020. Crossref, https://doi.org/10.1001/jamainternmed.2020.2033

[41] P. Gemmar, "An Interpretable Mortality Prediction Model for COVID-19 Patients–Alternative Approach," *MedRxiv*, Cold Spring Harbor Laboratory Press, 2020. Crossref, https://doi.org/10.1101/2020.06.14.20130732

[42] M. E. H. Chowdhury, T. Rahman, A. Khandakar, S. Al-Madeed, S. M. Zughaier, S. A. R. Doi, H. Hassen, and M. T. Islam, "An Early Warning Tool for Predicting Mortality Risk of COVID-19 Patients Using Machine Learning," *Cognitive Computation*, pp. 1-16, 2021. Crossref, https://doi.org/10.1007/s12559-020-09812-7

[43] S. Bolourani, M. Brenner, P. Wang, T. McGinn, J. S. Hirsch, D. Barnaby, T. P. Zanos, and Northwell, "COVID-19 Research Consortium a Machine Learning Prediction Model of Respiratory Failure within 48 Hours of Patient Admission for COVID-19: Model Development and Validation," *Journal of Medical Internet Research*,vol. 23, no. 2, 2021. Crossref, https://doi.org/10.2196/24246

[44] E. Jimenez-Solem, T. S. Petersen, C. Hansen, C. Hansen, C. Lioma, C. Igel, W. Boomsma, O. Krause, S. Lorenzen, R. Selvan, J. Petersen, M. E. Nyeland, M. Z. Ankarfeldt, G. M. Virenfeldt, M. Winther-Jensen, A. Linneberg, M. M. Ghazi, N. Detlefsen, A. D. Lauritzen, A. G. Smith, M. D. Bruijne, B. Ibragimov, J. Petersen, M. Lillholm, J. Middleton, S. H. Mogensen, H. C. Thorsen-Meyer, A. Perner, M. Helleberg, B. S. Kaas-Hansen, M. Bonde, A. Bonde, A. Pai, M. Nielsen, and M. Sillesen, "Developing and Validating COVID-19 Adverse Outcome Risk Prediction Models from a Bi-National European Cohort of 5594 Patients," *Scientific Reports*, vol. 11, no. 1, pp. 1-12, 2021. Crossref, https://doi.org/10.1038/s41598-021-81844-x

[45] K. Ikemura, E. Bellin, Y. Yagi, H. Billett, M. Saada, K. Simone, L. Stahl, J. Szymanski, D. Y. Goldstein, and M. R. Gil, "Using Automated-Machine Learning to Predict COVID-19 Patient Survival: Identify Influential Biomarkers," *Journal of Medical Internet Research*, vol. 23, no. 2, 2021. Crossref, https://doi.org/10.2196/23458

[46] D. Ji, D. Zhang, J. Xu, Z. Chen, T. Yang, P. Zhao, G. Chen, G. Cheng, Y. Wang, J. Bi, L. Tan, G. Lau, and E. Qin, "Prediction for Progression Risk in Patients with COVID-19 Pneumonia: The CALL Score," *Clinical Infectious Diseases*, vol. 71, no. 6, pp. 1393-1399, 2020. Crossref, https://doi.org/10.1093/cid/ciaa414

[47] A. Karthikeyan, A. Garg, P. K. Vinod, and U. D. Priyakumar, "Machine Learning Based Clinical Decision Support System for Early COVID-19 Mortality Prediction," *Frontiers in Public Health*, vol. 9, 2021. Crossref, https://doi.org/10.3389/fpubh.2021.626697

[48] K. Wang, P. Zuo, Y. Liu, M. Zhang, X. Zhao, S. Xie, H. Zhang, X. Chen, and C. Liu, "Clinical and Laboratory Predictors of In-Hospital Mortality in Patients with Coronavirus Disease-2019: A cohort study in Wuhan, China," *Clinical Infectious Diseases*, vol. 71, no. 16, pp. 2079-2088, 2020. Crossref, https://doi.org/10.1093/cid/ciaa538

[49] SOM Toolbox 2.0, 2015. [Online]. Available: http://www.cis.hut.fi/somtoolbox

[50] A. Rottenstreich, G. Zarbiv, E. Oiknine-Djian, R. Zigron, D. G. Wolf, and S. Porat, "Efficient Maternofetal Transplacental Transfer of Anti-SARS-Cov-2 Spike Antibodies After Antenatal SARS-Cov-2 BNT162b2 mRNA Vaccination," *Clinical Infectious Diseases*, vol. 73, no. 10, pp. 1909-1912, 2021. Crossref, https://doi.org/10.1093/cid/ciab266

[51] A. Sittar, D. Major, C. Mello, D. Mladenić, and M. Grobelnik, "Political and Economic Patterns in COVID-19 News: From Lockdown to Vaccination," *IEEE Access*, vol. 10, pp. 40036-40050, 2022. Crossref, https://doi.org/10.1109/ACCESS.2022.3164692

[52] Arshiya Moin, "Artificial Intelligence vs Covid19," *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 5, pp. 5-7, 2020. Crossref, https://doi.org/10.14445/23488387/IJCSE-V7I5P102

[53] A. Pormohammad, M. Zarei, S. Ghorbani, M. Mohammadi, M. H. Razizadeh, D. L. Turner, and R. J. Turner, "Efficacy and Safety of COVID-19 Vaccines: Asystematic Review and Meta-Analysis of Randomized Clinical Trials," *Vaccines*, vol. 9, no. 5, 2021. Crossref, https://doi.org/10.3390/vaccines9050467

[54] National Institutes of Health Website, 2022. [Online]. Available: https://www.nia.nih.gov/health/placebos-clinical-trials

[55] C. Cai, Y. Peng, E. Shen, Q. Huang, Y. Chen, P. Liu, C. Guo, Z. Feng, L. Gao, X. Zhang, Y. Gao, Y. Liu, Y. Han, S. Zeng, and H. Shen, "A Comprehensive Analysis of the Efficacy and Safety of COVID-19 Vaccines," *Molecular Therapy*, vol. 29, no. 9, pp. 2794-2805, 2021. Crossref, https://doi.org/10.1016/j.ymthe.2021.08.001

[56] G. Liu, B. Carter, T. Bricken, S. Jain, M. Viard, M. Carrington, and D. K. Gifford, "Computationally Optimized SARS-Cov-2 MHC Class I and II Vaccine Formulations Predicted to Target Human Haplotype Distributions," *Cell Systems*, vol. 11, no. 2, pp. 131-144, 2020. Crossref, https://doi.org/10.1016/j.cels.2020.06.009

[57] H. Chemaitelly, H. M. Yassine, F. M. Benslimane, H. A. AlKhatib, P. Tang, M. R. Hasan, J. A. Malek, P. Coyle, H. H. Ayoub, Z. AlKanaani, E. AlKuwari, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, H. F. AbdulRahim, G. K. Nasrallah, M. G. AlKuwari, H. E. AlRomaihi, M. H. Al-Thani, A. AlKhal, A. A. Butt, R. Bertollini, and L. J. Abu-Raddad, "mRNA-1273 COVID-19 Vaccine Effectiveness Against the B.1.1.7 and B.1.351 Variants and Severe COVID-19 Disease in Qatar," *Nature Medicine*, vol. 27, pp. 1614-1621, 2021. Crossref, https://doi.org/10.1038/s41591-021-01446-y

[58] J. Chapin-Bardales, J. Gee, T. Myers, "Reactogenicity following Receipt of mRNA-Based COVID-19 Vaccines," *JAMA*, vol. 325, no. 21, pp. 2201-2202, 2021. Crossref, https://doi.org/10.1001/jama.2021.5374

[59] J. Held, J. Esse, K. Tascilar, P. Steininger, K. Schober, P. Irrgang, R. Alsalameh, M. Tenbusch, C. Seggewies, and C. Bogdan, "Reactogenicity Correlates Only Weakly with Humoral Immunogenicity After COVID-19 Vaccination with BNT162b2 mRNA (Comirnaty®)," *Vaccines*, vol. 9, no. 10, 2021. Crossref, https://doi.org/10.3390/vaccines9101063

[60] M. V. Tilbeurgh, K. Lemdani, A.S. Beignon, C. Chapon, N. Tchitchek, L. Cheraitia, E. M. Lopez, Q. Pascal, R. L. Grand, P. Maisonnasse, and C. Manet, "Predictive Markers of Immunogenicity and Efficacy for Human Vaccines," *Vaccines*, vol. 9, no. 6, 2021. Crossref, https://doi.org/10.3390/vaccines9060579

[61] W. Burny, A. Callegaro, V. Bechtold, F. Clement, S. Delhaye, L. Fissette, M. Janssens, G. Leroux-Roels, A. Marchant, R. A. van den Berg, N. Garçon, R. van der Most, and A. M. Didierlaurent, "Different Adjutants Induce Common Innate Pathways That are Associated with Enhanced Adaptive Responses Against a Model Antigen in Humans," *Frontiers in Immunology*, vol. 8, 2017. Crossref, https://doi.org/10.3389/fimmu.2017.00943

[62] Y. H. Hwang, K. H. Song, Y. Choi, S. Go, S. J. Choi, J. Jung, C. K. Kang, P. G. Choe, N. J. Kim, W. B. Park, and M. D. Oh, "Can Reactogenicity Predict Immunogenicity After COVID-19 Vaccination?," *The Korean Journal of Internal Medicine*, vol. 36, no. 6, 2021. Crossref, https://doi.org/10.3904/kjim.2021.210

[63] C. Réda, E. Kaufmann, A. Delahaye-Duriez, "Machine Learning Applications in Drug Development," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 241-252, 2020. Crossref, https://doi.org/10.1016/j.csbj.2019.12.006

[64] B. A. Abbasi, D. Saraf, T. Sharma, R. Sinha, S. Singh, P. Gupta, S. Sood, A. Gupta, and K. Rawal, "Identification of Vaccine Targets & Design of Vaccine Against SARS-Cov-2 Coronavirus using Computational and Deep Learning-Based Approaches," *Bioinformatics and Genomics*, 2020. Crossref, https://doi.org/10.7717/peerj.13380

[65] Z. Yazdani, A. Rafiei, M. Yazdani, and R. Valadan, "Design an Efficient Multi-Epitope Peptide Vaccine Candidate against SARS-CoV-2: An in Silico Analysis," *Infection and Drug Resistance*, vol. 13, 2020. Crossref, https://doi.org/10.2147/IDR.S264573

[66] The Immune Epitope Database (IEDB), 2022. [Online]. Available: http://www.iedb.org

[67] Rankpep: Prediction of Binding Peptides to Class I and Class II MHC Molecules, 2022. [Online] Available: http://imed.med.ucm.es/Tools/rankpep.html

[68] Coronavirus (COVID-19) Vaccinations, 2022. [Online]. Available: https://ourworldindata.org/covid-vaccinations?country=OWID_WRL

[69] Y. Omae, M. Sasaki, J. Toyotani, K. Hara, and H. Takahashi, "Theoretical Analysis of the SIRVVD Model for Insights Into the Target Rate of COVID-19/SARS-Cov-2 Vaccination in Japan," *IEEE Access*, vol. 10, pp. 43044-43054, 2022. Crossref, https://doi.org/10.1109/ACCESS.2022.3168985