INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

Corresponding authors.

salkhalidi@kku.edu.sa

prakashcnet@gmail.com

COVID-19: An Efficient Big Data Analytics for SARS-CoV-2 Mutations Prediction: A Machine Learning Approach

Dr. Saeed Q Al-Khalidi Al-Maliki[1]*, **Dr. Prakash Kuppuswamy[2]***,

Dr. Rajan John[3], **Dr. Nithya Rekha Sivakumar[4]**

1 Department of MIS, College of Business, King Khalid University, Abha, KSA
2 Computer Networks Engineering Department, College of CS & IT, Jazan University, Jazan, KSA
3 Department of Computer Science & Engineering, College of CS & IT, Jazan University, Jazan, KSA
4 Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

Abstract

Objectives: This research article aims to identify the level of further mutations of COVID-19 and its vulnerability. Healthcare workers are urged to take precautionary measures to prevent further damage to their community and to avoid further human deaths. At this stage, it is crucial to study, analyze, and understand the COVID-19 variant mutation. **Methods:** The mutation of COVID-19, however, has not yet been validated for the way in which it affects the risk of underlying comorbidities. It may be possible to tackle this challenge by using artificial intelligence, machine learning, and deep learning algorithms. We have used the J48 algorithm and the Linear Regression algorithm. These algorithms were then analyzed based on the accuracy they provided after running them in the output window of WEKA data mining. **Findings:** In this paper, machine learning is used to help clinicians and medical researchers understand COVID-19 variant mutations at various stages. Artificial intelligence-based algorithms provide a better understanding of COVID-19 stages and vulnerability levels. **Novelty:** Artificial intelligence can be a valuable tool for healthcare operations in this pandemic situation. Moreover, healthcare professionals and a data analysis algorithm will probably arrive at the same conclusion based on the data set. However, the use of machine learning-based techniques described in this article will allow for quicker and earlier diagnosis of any type of pandemic situation in the future.

Keywords: COVID19; Machine Learning; Mutation; Variant; Coronavirus; SARSCoV2; Prediction algorithm etc

1 Introduction

The novel coronavirus (SARS-CoV-2) was detected for the first time in China's Hubei province's city of Wuhan in December 2019 [1,2]. This virus causes a severe respiratory disease (COVID-19), which has spread vigorously around the world [3]. The COVID-19 outbreak was declared a pandemic by the WHO (World Health Organization) on March 11, 2020 [4]. Because of the novelty of this disease, its clinical course and treatment are largely unknown [5]. By the end of December 2021, there will be 41359 cases in the European Union Medical Diagnosis Register, of which 6790 are clinical trials conducted with less than 18-year-old people [6–9]. Over 289 million people worldwide have been infected since then, and over 5.44 million people have died from coronavirus disease as of December 2021 [10]. Despite unprecedented efforts by scientists and physicians to sequence, diagnose, treat, and prevent COVID-19, the disease has yet to reveal its lasting effects on individuals after the acute phase. Managing the challenges during the COVID-19 pandemic. As well as city lockdowns, measures such as COVID-19 are imposed when there is a serious threat [10]. COVID-19 is an unmeasurable and unidentified kind of disease that has caused governments worldwide to give warnings to people at high risk. Additionally, it has provided guidelines for observing social distancing and isolation measures since they are more likely to require critical care, including ventilation, and to die if they become ill [11].

However, in real-world situations, considering risk over multiple periods is more realistic and more interesting because we can observe the dynamic nature of the problem. COVID-19 typically spreads as "waves" of infection. It emphasizes the importance of having a multi-period analysis and being myopic is not enough. A multi-period analysis of risk poses technical challenges [12]. Study of similar outbreaks of SARS, MERS, and influenza H1N1. As with previous outbreaks, such as SARS in 2002–2003, clinicians are confronted with new diseases for which there are few effective treatment options [13]. The level of infectious virus needed to make COVID-19 transmissible is not known. However, a carefully conducted epidemiological study that identifies the proportion of pre-symptomatic to post-symptomatic spread is the best way to determine the potential for transmission. As previously stated, previous studies can provide useful information in interpreting epidemiological data [12–16].

New variants or sub-lineages of the SARS-CoV-2 virus could contain mutations that distinguish them from the reference sequences or predominant types already circulating in the population. In contrast with previous or currently circulating viruses, different strains of SARS-CoV-2 can have different characteristics, including the ability to spread more easily, exhibit resistance to existing treatment options, and have no impact compared to previous strains [17,18]. As was the case during earlier pandemic diseases such as SARS, MERS, and H1N1, wide observational clinical studies will be conducted until promising randomized controlled trials become available in order to evaluate potential treatment effects on hospital mortality [17,18]. Because observational studies cannot provide causal conclusions, they cannot replace randomized controlled trials. On the other hand, they can serve as a catalyst for further research into potential treatment options.

Researchers, clinicians, health-care workers, epidemiologists, and decision makers face a formidable challenge as a result of this pandemic. As a part of this global effort, BMC Medical Research Methodology has set up a collection of articles called "Methodologies for COVID-19 research and data analysis". Rapid assessment of a dynamic research area such as COVID-19, where the body of evidence has grown at an impressive rate, requires an approach that is more direct and has a wider scope than the current gold standard methods, such as scoping and systematic reviews. Many research articles have discussed the potential use of machine learning and artificial intelligence in the fight against the COVID-19 crisis [18].

Machine learning (ML) is a part of artificial intelligence, where algorithms analyze large data sets to detect patterns and make predictions. In the healthcare industry, machine learning algorithms have a lot of potential because there is a lot of data generated for each patient. Therefore, it's no surprise that there are multiple successful applications of machine learning in the healthcare industry [19]. An extensive range of tasks can be solved using machine learning techniques in healthcare, as well as many other research applications. In healthcare, machine learning facilitates the following services.

- Classification
- Recommendations
- Clustering
- Prediction
- Anomaly detection
- Automation
- Ranking

Since SARS-CoV-2 mutation rates are high, developing a vaccine to protect against its many subtypes poses a huge challenge to public health worldwide. If an existing vaccine cannot protect against the virus, a new vaccine must be developed [20]. Over the course of the COVID-19 pandemic, the SARS-CoV-2 virus mutated, which resulted in genetic variations in the circulating viral strains. Researchers and healthcare providers are unable to provide an exact medicine for COVID-19 patients. Predicting the mutation of COVID-19 is very important at this point to prevent a major medical problem for the human population. The medical practitioner may be prepared to prevent the high level of risk by knowing the variant's vulnerability.

The article explains the method of creating datasets, the best policy resolutions, how the CORD-19 method has been applied, and how the dataset has been used to build shared tasks. In order to find efficient diagnosis and treatment policies for COVID-19, researchers will continue to bring together the computing community, biomedical experts, and policymakers. Among the many tools that can be focused on contesting the COVID-19 epidemic, automated text analysis stands out as a tool that cannot provide a full solution. Additionally, this work illustrates the value of public access to full-text literature, since it allows users to interact with and discover texts through computational access[21].

COVID-19 can currently only be treated with comprehensive support. In mild cases, patients can often return to full health after appropriate clinical treatment. In order to determine what happened during cancer treatment, this study will analyze patients' blood parameters before and after their complete remission. The paper examines four potential mechanisms leading to lymphocyte deficiency. Initially, the virus may directly attack those cells, killing them. Infection of lymphocytes may be facilitated by the presence of the coronavirus receptor ACE2. Viruses can also attack the lymphatic system directly. There is evidence of acute lymphocyte decline, which may be due to dysfunction in the cells themselves. It might also be due to a direct invasion by coronavirus, which attacks organs like the thymus and spleen[22].

Clinical research companies have been under tremendous pressure since the COVID-19 pandemic broke out. A redistribution of resources and a temporary suspension of face-to-face visits will inevitably restrict research in other therapeutic areas. The clinical research center will screen participants for COVID-19 symptoms, fever, and exposure history before admitting them to the study. Participants can choose to visit remotely, when possible. Masks and physical distance are required for face-to-face interactions, and visitors are not allowed. SARS-CoV-2 infections may cause lower or higher blood pressure in hypertensive patients, worsen diabetes control in diabetics, or accelerate kidney disease progression in chronic kidney disease patients. A delay in treatment due to fear of contracting the virus can result in acute illnesses, hospitalization, and even death. Researchers, coordinators, and clinicians rekindle a sense of perseverance and a mission to use science to solve problems that are important to patients and the public[23].

In this methodical assessment and meta-analysis, we searched for LitCOVID and Embase. Studies published before January 1, 2021 include at least 100 patients. The study sample was collected between the ages of 17 and 87 years old. From observation, it has been estimated that more than 80% of patients infected with SARS-CoV-2 have one or more symptoms, such as fatigue, headache, attention deficit, hair loss, and breathing difficulties[24]. For a better understanding, studies need to be stratified according to masculinity, age, prior comorbidities, severity of COVID-19, and interval of each sign being investigated. From a clinical point of view, a multidisciplinary team is essential to designing anticipatory measures, recovery techniques, and clinical management policies. From the perspective of the overall patient, it aims to solve the care problems after COVID-19[25].

2 Methodology

In this Research article, we have applied these Machine learning techniques to predict the COVID-19 mutation through the classification of the J48 and Linear Regression prediction algorithm. The details about the COVID-19 variant mutation data set from WHO have been included. The classification models J48 and Linear Regression are used to predict further mutation of COVID-19 variants. A significant number of results have been achieved through J48 and REP Tree in areas of representing, utilizing, and learning statistical knowledge. Using trained data, which we have collected from the World Health Organization web portal, is our test option. With this interface, all algorithms could be run at the same time, and the results could be compared. Detailed explanations of the algorithms applied to process COVID-19 variant data sets are given in the section below.

Previously conducted COVID-19 related studies used data that was collected from a variety of sources. The prediction of further mutation of SARS-CoV-2 is not discussed in earlier studies using the Artificial Intelligence tool of Machine Learning. The reason for applying the Weka tool is it has the facility of extending features to diagnose and predict different diseases. In addition, medical practitioners and researchers can expand their research activities with cost-effective and time-saving options. It can also help in solving the problems of clinical research using different applications of Weka. Another advantage of using Weka for the prediction of diseases is that it can easily diagnose any kind of disease with the necessary dataset.

2.1 Data Pool

A dataset is an instance of statistical data in which every attribute of the data represents a variable. Every instance has its own description. In order to compare the accuracy of algorithms with Weka tools for prediction of COVID-19 mutations, we used Mutation data for the prediction and classification of algorithms. For classification and accuracy of further Coronavirus mutation prediction, we used 6 attributes and 98282 instances from 120 countries. We have applied different J48, Rep Tree algorithms with the WEKA data mining tool for our data analysis purpose. This study focused on the disease (COVID-19), rather than the virus, as well as the prediction of a vulnerability to further mutation.

The data pool is collected from the WHO web portal and European Union official website, which is mentioned in Figure 1 Data set. The data set is named the COVID-19 variant. WHO and EU web portals contain a vast and varied number of datasets relating to COVID-19 worldwide datasets. These data are widely used by COVID-19 Researchers and Medical experts to understand the ongoing COVID-19 situation. Various academic papers and researches have been conducted using this COVID-19 data set.
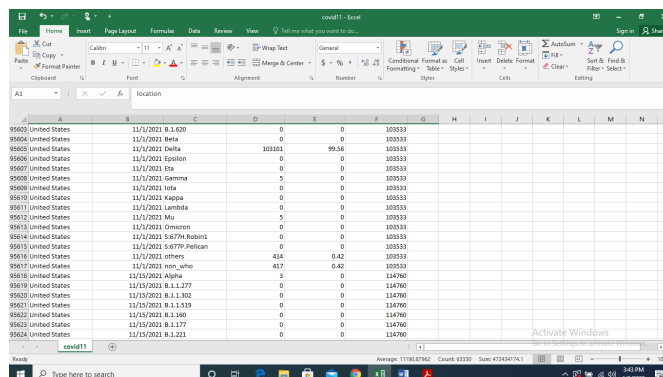
Fig 1. COVID-19 variant Data set (**Source** :https://github.com/owid/COVID-19-data0pt)

2.2 Data Preprocessing

The machine learning algorithm can be used to identify the effects of mutations to determine why mutated variants of SARS-CoV-2 spread more rapidly. By doing this, we could identify mutation variants that are concerning before they spread and notify the reply by medical authorities. Applying machine learning models for classification, the study sought to find prognostic factors for predicting COVID-19 mutation. In preprocessing, data are cleaned, missing data is replaced, data is transformed, and data imbalances are reduced. In this study, 98232 records and 120 nations' mutation data were included, as shown in Figure 2. Preprocessing data plays an influential role in machine learning. For high-quality datasets, imputation and normalization were applied.
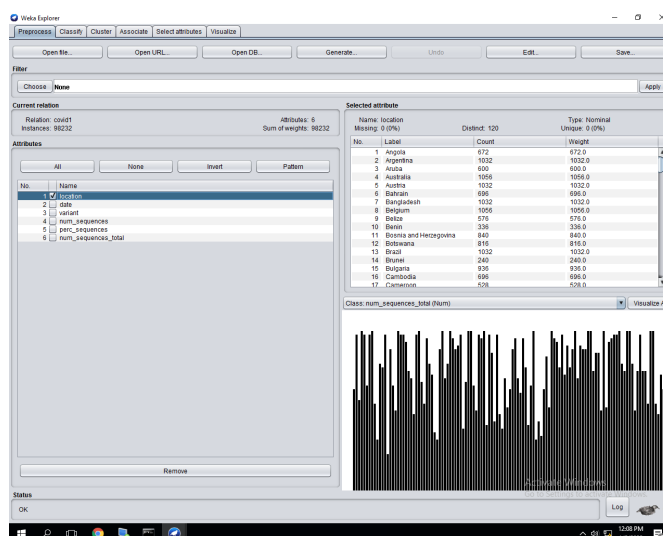
Fig 2. Data preprocessing

2.3 Choosing Research variables

It is a process of selecting the best subset of the relevant variable for use in the prototypical structure. It selects a suitable classification for prediction accuracy. Perfect variable selection helps to evaluate the effectiveness of included variables in the training dataset. The dataset and its description, which we have derived from the WHO portal are shown in Table 1.

Table 1. COVID-19 variant dataset attributes and description

ATTRIBUTES NAME	DESCRIPTION
Location	Name of Country
Date	Mutation date
Variant	COVID-19 Variant
num_sequences	Mutation sequence
Perc_sequences	Percentage of Mutation
num_sequences_total	Total sequence

2.4 Model building

The predictive classifier models were developed to accurately identify COVID-19 variant mutation. The classification model J48 and Linear Regression is used to develop prediction models. We considered these models due to their following characteristics. It is a widely and most commonly used method of empirical analysis in sociology, bio-statistics, clinical medicine, quantitative psychology, econometrics, marketing, and often uses to compare with machine learning studies. It has many advantages including high power and accuracy. The J48 classifier is built independently by applying the general technique of bootstrap aggregating and is a selected sample for the training set. Still this J48 method is dominant model used to the prediction with a degree of certainty. The validation results from j48 classifier models were then combined to provide a measure of the overall performance.

Prediction statistical analysis variables were presented as the mean, standard deviation which is analyzed by the J48 classifier in addition to the Linear Regression. The performance of classification models to predict COVID-19 mutation prediction was calculated by the Receiver Operating Characteristic (ROC). We also calculated the accuracy (AC), Kappa statistic, Mean Absolute Error (MAE) Root mean squared error (RMSE), Relative absolute error (RSE). Weka (V.3.9.5) were implemented to build classifier models. The Weka tools consist a collection of graphical user interface and a visualization for easily performing algorithms.

3 Significance of Research

A disaster has struck with the COVID-19 pandemic. It affects a large number of people. In places where there is a severe pandemic, how food, medicines, masks, and other necessities can be delivered in a timely manner can save lives. As there may be city lockdowns, there may also be obstructions to transportation. Next, it is important to protect the health of service agents, volunteers, and logistics workers. We know that COVID-19 virus is highly infectious and spreads rapidly. Keeping the people who help safe is also crucial. The third challenge is how to ensure that healthcare products, vaccines and drugs are delivered safely to the people in need while operating smoothly through the respective supply chains. As such, designing the logistics system in question is a very important research topic, which should be explored in more depth in the future. The use of different vaccine techniques cannot control COVID-19, due to its variants, which have been identified from 2019 onwards. This study is to use machine learning techniques to analyze how the COVID-19 variants are changing regularly, as well as their effect on humans.

4 Results & Discussion

In the healthcare sector, Machine learning algorithms have been implemented in many applications. For example, it can be used to automate time-consuming and complex tasks within this field. Today, artificial intelligence-based algorithms provide faster processing power and are used in a variety of health care fields, including drug discovery, disease prediction, improved therapeutic diagnosis, etc. We applied machine learning algorithms to predict further mutation situations and vulnerability statistics based on the COVID-19 variant dataset, which received data from 120 nations and 98282 instances. Our goal was to analyze the data with classification models and predict further mutations and their effects. The J48 classifier was chosen

because it performs better with a large number of datasets with a variety of attributes. Moreover, the J48 algorithm stimulated the enactment of the data without any obscuring effects on selected attributes. The performance metrics after attribute selection, parameter tuning, and calibration are used because this is a standard process for evaluating algorithms. The classification results achieved correctly were 8979 instances, i.e. 9.1406%, and the incorrectly classified instances were 89253, i.e. 90.8594%. A total of 98282 instances were involved in the process of classifying the results. According to further statistical analysis, the kappa statistic output was 0.0519, the mean absolute error was 0.0769, the root mean squared error was 0.01961, the relative absolute error was 96.2989%, and the root relative squared error was 98.132%.

The purpose of this work is to predict the chances of further mutation and its vulnerability. We have taken datasets with different performance measures using machine learning. All data were preprocessed and used for test prediction. The highlight of the resulting output is given below.

$$\text{Mean Absolute Error} = 1/n \sum_{j=1}^{n} \left| y_j - y_j \right| = 0.0769$$

$$\text{Root Relative Squared Error} = \sqrt{\frac{\sum_{i=1}^{N} (\theta_i - \theta_i) \cdot 2}{\sum_{i=1}^{N} (\theta - \theta_i) \cdot 2}} = 98/13\%$$

$$\text{Relative Absolute Error} = \frac{\sum_{i=1}^{N} |\theta_i - \theta_i|}{\sum_{i=1}^{N} |\theta_i - \theta_i|} = 96.2989\%$$

$$\text{Root Mean Squared Error} = \sqrt{1/N \sum_{i=1}^{n} (\theta_i - \theta_i)^2} = 0.1961$$

In this study, the accuracy of COVID-19 variants was classified using J48 and Linear Regression models on different data sets, and the results were compared to arrive at a feasible solution and for cross-verification of the predicted result. The above algorithms used by us were applied to a prediction of the COVID-19 mutation situation. The data set details are mentioned in Figure 1. We performed a 10-fold cross-validation test in order to improve accuracy. For each classification, we selected training and testing samples randomly from the base set to train the model and then test it in order to estimate the classification and accuracy measure for each classifier. Figure 2 depicts the COVID-19 attributes related to date, country of origin, and variant types. Next, a linear regression algorithm has been used in this paper to randomly choose the attributes of each variable to allow the estimation of class probabilities. Running the algorithm, we analyze the classifier output with statistics-based output by using 90 percent split training and a 10% reminder test to make a prediction of each instance of a dataset, as shown in Figure 3.
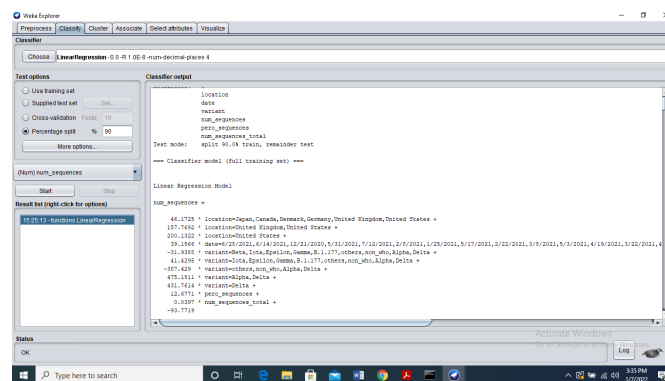

Fig 3. Linear regression classifier Result

The output of the Linear regression classifier result is shown in Figure 4, From the result, we have observed the value of 0.27 is correlation coefficient result, Mean absolute error (MAE)result is 216.809, Root mean squared error (RMSE) result is 1395.4758, Relative absolute error (RAE) value is 165.2355% and Root relative squared error (RRSE) value is 96.4624% from the 10% instances i.e 9823 from 98232 total instances.

We have selected the J48 classifier once again to identify and cross verify with the linear regression to predict any other further mutations. J48 Tree has been used in this paper to decide the target value based on the time to build the model, accurately classified instances and incorrectly classified instances, as well as the accuracy of all classifiers. Figure 5 shows the process results based on various attributes of the dataset used to predict machine learning models and classify their accuracy.

After the J48 prediction classifier, we analyzed results from the obtained classifier. The output gave several statistics based on a percentage split of 90% to make a prediction of each instance of the dataset. Figure 6 shows the classification accuracy achieved from this algorithm i.e. 8.134 % is the correctly classified accuracy for a batch of 9823, incorrectly classified instances
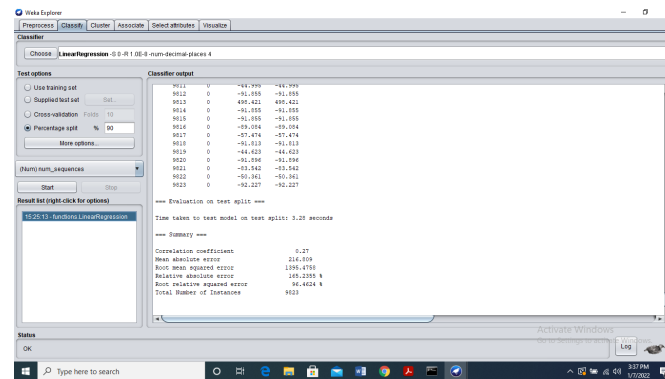
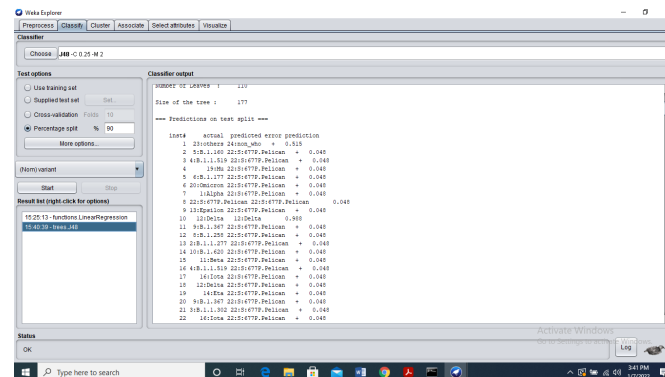**Fig 4.** Linear regression prediction result

Fig 5. J48 Classification Result

9024, Kappa statistics 0.432, Mean absolute error 0.0771, Root mean squared error 0.1966, Relative absolute error 96.5718% Root relative squared error 98.3706%, Coverage of case 96.6609%, and mean region size 83.7053. It has taken for execution duration 3.45 seconds.
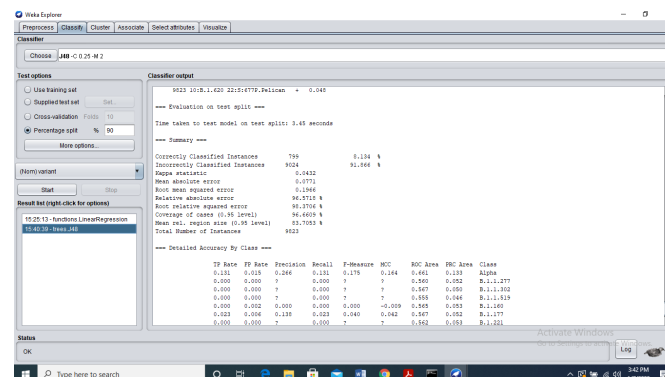

Fig 6. J48 Prediction Result

In Figure 7, we have validated the result in a graphical view. The Y-axis consists of the total number of sequences, and the X-axis consists of the value of COVID-19 mutation data. It shows, that Alpha, Delta and unspecified variants are spreading more than another category of viruses

We performed experiments on different numbers of the selected attributes of the COVID-19 mutation dataset. In addition, we have verified using various preprocess setups such as supervised and unsupervised with various metrics. Next, we have
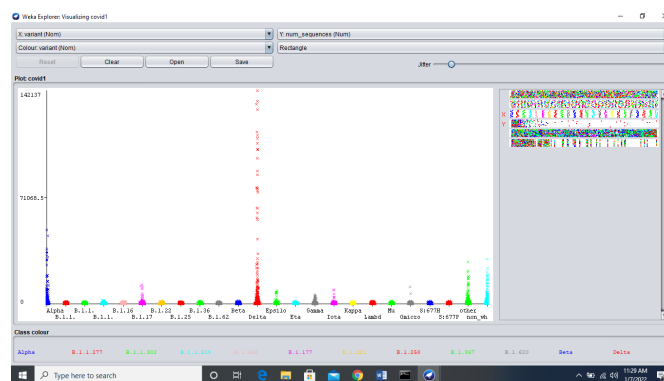
Fig 7. COVID-19 mutation

applied different classifiers using supervised learning, unsupervised learning, cross-validation and percentage split etc.,

5 Research Findings

Healthcare data has expanded rapidly in recent years, and machine learning makes it possible to analyze massive amounts of data quickly. Therefore, it is an opportunity to apply machine learning models to the care of individual patients in medical practice. There is an alarming spread of this new type of coronavirus, which largely evades immunity. Within just two weeks' time, between December 25, 2021 and January 7, 2022, it has affected 24,151,332 people in the world[25]. Coronaviruses can steadily obtain mutations as they spread from person to person. The COVID-19 mutations seem to be appearing in all parts of the world. We found several significant points in the COVID-19 mutation data set derived from the WHO and processed using a machine learning algorithm. In the case of the J48 algorithm, its predicted ROC value is 0.591, not a favorable value. It implies that the given dataset of processed elements has a higher mutation rate and that there are further mutation possibilities. Somehow, the COVID-19 death rate is better than previous COVID-19 variants such as Alpha and Delta. Figure 8 shows the COVID-19 mutation period, and Table 2 shows the death of COVID-19 at intervals.
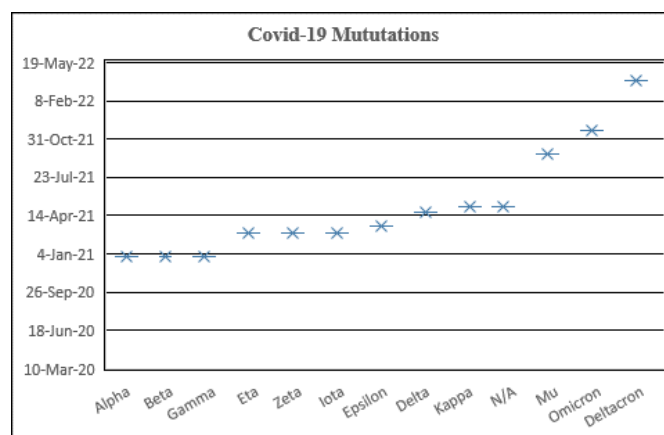

Fig 8. Mutation period

Covid-19 prediction models are rapidly entering the academic literature to support medical decision making at a time when they are urgently needed. Using unreliable predictions to guide clinical decisions could cause more harm than good, which is why methodological guidance should be followed. Predicting COVID-19 is not always as easy as a binary classification. It is important to deal with the complexity of the data. Some reviews excluded some of the data from analysis of earlier studies because participants had neither recovered nor died within that time period.

Table 2. COVID-19 death analysis table

Period	No of Deaths	Total deaths	Average(Monthly)
Up to 31 January 2020	266	266	-
01-2-2020 to 31-6-2020	573125	573391	95565
01-7-2020 to 31-12-2020	1368371	1941762	228061
01-1-2021 to 31-6-2021	2045731	3987493	340955
01-7-2021 to 31-12-2021	1465986	5453479	244331
01-01-2022 to 31-3-2022	708524	6162003	236174

Source: https://www.worldometers.info/coronavirus/worldwide-graphs/#total-deaths

6 Research Constraints

The present study has several limitations that need to be addressed. In the first instance, the information was collected from the WHO with the novelty of different viruses of COVID-19. SARS-CoV-2 is still causing a lot of discussion and confusion. Nevertheless, the World Health Organization itself is unable to identify how and from where the SARS-CoV-2 virus originated. In the next step, a Weka data mining tool was used to predict further mutation and vulnerability, however, the result differed from clinical and laboratory-based data experiments. Scientists, researchers, and clinicians would be able to produce better outcomes due to using physical datasets instead of the available datasets to make therapeutic decisions. The experiment that we have conducted fully relies on computer-based evaluation utilizing J48 and linear regression algorithms. We used a huge dataset of 98232 instances with six attributes, which suggests that a machine learning algorithm may be able to provide the best results. When we have a huge amount of data, it is difficult to validate. We analyzed a huge number of variables that were considered a sample size, even though most of the variables are not statistically significant because they are repeated in the dataset. Due to the large dataset, there is a possibility of producing nearly unbiased estimates of the prediction results. Finally, we used a classification approach for automatic machine learning variable integration, but a deep learning approach would have improved prediction.

7 Conclusion

The purpose of this paper is to predict further mutations of COVID-19 and determine its vulnerability. For the verification and validation of prediction results, we have used the J48 algorithm and the Linear Regression algorithm. These algorithms were then analyzed based on the accuracy they provided after running them in the output window using WEKA data mining. We analyzed the results based on accuracy after running these algorithms. On the basis of a given attribute value, model building time, mean absolute error, and ROC area, J48 and linear regression algorithms were used to analyze the COVID-19 variant data set.

Results show that the J48 algorithm predicted a ROC value of 0.591 and the mean region size was 83.7053. In contrast, linear regression predicted a correlation value of 0.27, and the root mean squared result was more than 1. We have concluded that 73% of further mutations are possible based on the prediction (1-0.27 = 0.73). It is likely to spread to more than 83% of the country (a mean region size of 83.7053) and has some favorable results regarding death rates. According to the WHO report, the total number of deaths as of 31/3/2022 was 6162003. A WHO report shows that the number of dead people is falling. The highest number of deaths was 2045731 between January-2021 and June-2021, the average was 340955. There were 212542 deaths recorded in December 2021 and March 2022 deaths were recorded as 162907, showing a better causality rate than in previous years. Our prediction result is based on the WHO database attribute only, and it has been evaluated by the prediction algorithm. We are not evaluating the data on a clinical basis, and some ambiguity may arise from computer-based data analysis. As a result, the machine learning prediction algorithm predicts that a further mutation is likely to have fewer casualties.

In the future, the research article should lead to more clinical-oriented experiments to optimize the predictive performance of these classifiers for COVID-19 virus-related diagnoses using other feature selection algorithms and optimization techniques. In addition, it may provide physicians with a better understanding of real-world clinical practice, which would let them better identify the vulnerabilities of novel coronavirus diagnoses and preventative measures.

8 Acknowledgement

References

1) Alga A, Eriksson O, Nordberg M. Analysis of Scientific Publications During the Early Phase of the COVID-19 Pandemic: Topic Modeling Study". *Journal of Medical Internet Research*. 2020;22(11). Available from: https://doi.org/10.2196/21559.

2) Alexandra L, Katz R, Lawrence O. The novel coronavirus originating in Wuhan, China: challenges for global health governance". *Pub Med, National Library of Science*. 2020;8. Available from: https://doi.org/10.3390/healthcare10020299.

3) Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*. 2020;382(13):1199–1207. Available from: https://doi.org/10.1056/NEJMoa2001316.

4) Word Health Organization - Regional Office for Europe. Copenhagen, Denmark . . Available from: http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-COVID-19/news/news/2020/3/who-announces-COVID-19-outbreak-a-pandemic.

5) Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*. 2020;28:30566–30569. Available from: https://doi.org/10.1016/S0140-6736(20)30566-3.

6) Thorlund K, Dron L, Park J, Hsu G, Forrest JI, Mills EJ. A real-time dashboard of clinical trials for COVID-19". *Lancet Digital Health*. 2020;6:286–287. Available from: https://doi.org/10.3390/ijerph17114095.

7) Chadaga K, Prabhu S, Vivekananda BK, Niranjana S, Umakanth S. Battling COVID-19 using machine learning: A review. *Cogent Engineering*. 2021;8(1). Available from: https://doi.org/10.1080/23311916.2021.1958666.

8) Global Coronavirus COVID-19 Clinical Trial Tracker. . Available from: https://www.covid-trials.org/.

9) Choi TM. Risk analysis in logistics systems: A research agenda during and after the COVID-19 pandemic. *Transportation Research Part E: Logistics and Transportation Review*. 2021;145:102190–102190. Available from: https://doi.org/10.1016/j.tre.2020.102190.

10) Ivanov D. Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review*. 2020;136:101922–101922. Available from: https://doi.org/10.1016/j.tre.2020.101922.

11) Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, et al. Coronavirus Pandemic (COVID-19). 2020. Available from: https://ourworldindata.org/covid-cases.

12) Chen N, Zhou M, Dong X. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395:30211–30218. Available from: https://doi.org/10.1016/S0140-6736(20)30211-7.

13) Young BE, Ong SWX, Kalimuddin S, Low JG, Tan SY, Loh J, et al. Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. *JAMA*. 2020;323(15):1488–1488. Available from: https://doi.org/doi:10.1001/jama.2020.3204.

14) Huang C, Wang Y, Li X. Clinical features of patients infected with 2019 novel coronavirus in Wuhan. *Lancet*. 2020;(20):30183–30188. Available from: https://doi.org/10.1016/S0140-6736(20)30183-5.

15) Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA*. 2020;323(11):1061–1061. Available from: https://doi.org/10.1001/jama.2020.1585.

16) Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*. 2020;382(18):1708–1720. Available from: https://doi.org/10.1056/NEJMoa2002032.

17) Arabi YM, Shalhoub S, Mandourah Y, Al-Hameed F, Al-Omari A, Qasim EA, et al. Ribavirin and Interferon Therapy for Critically Ill Patients With Middle East Respiratory Syndrome: A Multicenter Observational Study. *Clinical Infectious Diseases*. 2020;70(9):1837–1844. Available from: https://doi.org/10.1093/cid/ciz544.

18) Kumar A, Gupta P, Srivastava A. A review of modern technologies for tackling COVID-19 pandemic, Diabetes and Metabolic Syndrome. *Clinical Research and Reviews*. 2020;14(4). Available from: https://doi.org/10.1016/j.dsx.2020.05.008.

19) Kobayashi T, Jung SM, Linton NM, Kinoshita R, Hayashi K, Miyama T, et al. Communicating the Risk of Death from Novel Coronavirus Disease (COVID-19). *Journal of Clinical Medicine*. 2020;9(2):580–580. Available from: https://doi.org/10.3390/jcm9020580.

20) Page ML. Stopping corona Virus. *New Scientist*. 2021;252:3365–3366. Available from: https://www.sciencedirect.com/journal/new-scientist/vol/252/issue/3365.

21) Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: The COVID-19 Open Research Dataset", PMC Lab, US National Library of Medicine National Institutes of Health. 2020. Available from: https://pubmed.ncbi.nlm.nih.gov/32510522/.

22) Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 Long-term effects of COVID-19: a systematic review and meta-analysis. *medRxiv*. 2021. Available from: https://doi.org/10.1101/2021.01.27.21250617.

23) Bukhari SNH, Jain A, Haq E, Mehbodniya A, Webber J. Machine Learning Techniques for the Prediction of B-Cell and T-Cell Epitopes as Potential Vaccine Targets with a Specific Focus on SARS-CoV-2 Pathogen: A Review. *Pathogens*. 2022;11(2):146–146. Available from: https://doi.org/10.3390/pathogens11020146.

24) Tan L, Wang Q, Zhang D, Qianchuan JD, Huang YQ, Tang Q, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Signal Transduction and Targeted Therapy*. 2020;5(33). Available from: https://doi.org/10.1038/s41392-020-0148-4.

25) Magazine, Centers for Disease control and prevention, United states. . Available from: https://covid.cdc.gov/covid-data-tracker/#variant-proportionsaccessedon9/1/2022.