Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

PREDICTING FUTURE POSSIBLE COVID 19 OUTBREAKS USING MACHINE LEARNING

SHAWWAL RASHED
Lecturer Computer Science, Department of Software Engineering, Superior University, Lahore, Pakistan.

SHEERAZ AKRAM
Associate Professor Computer Science, Department of Software Engineering, Superior University, Lahore, Pakistan.

SHAHARYAR RAFIQ
Department of Software Engineering, Superior University, Lahore, Pakistan.

FAWAD NASIM
Assistant Professor Computer Science, Department of Computer Science, Superior University, Lahore, Pakistan.

ARFAN JAFFAR
Dean, Department of Computer Science and Information Technology, Superior University, Lahore, Pakistan.

MUHAMMAD IMRAN TARIQ
Assistant Professor Computer Science, Department of Computer Science, Superior University, Lahore, Pakistan.

Abstract

By June 2020, almost 9 million confirmed verified Covid-19 cases had been confirmed, with above 468 thousand deaths. Contact with tainted objects can transmit the viruses and touch your mouth, eyes, or nose to an infected person. Since how much corona diseases are spreading quickly, it is hard to test because of time and cost factors. For a long time, ML has become reliable in medical fields. Usage of ML to estimate COVID-19 in patients will reduce the delays in the outcome of the clinical trial and develop a delay in providing proper medical treatment to the patients. The approach is based on Covid-19 dataset cases from WHO. The dataset includes data on each country's case, death, and recovery totals. The approach uses a Support Vector Machine (SVM) and polynomial regression (PR) to learn a model from the data. The model's outcomes demonstrate that the strategy can precisely forecast each country's case, mortality, and recovery rates. The study delivers an in-depth exploration of the global pandemic, its impact on various countries, and the future trends of the pandemic. This study highlights the need for an active response from citizens and governments of countries to global pandemic diseases. The study also provides recommendations about the steps that citizens and governments of countries need to take to reduce the effects of pandemic diseases.

Index Terms: Machine Learning; Classification Techniques; COVID-19; Supervised Learning

1. INTRODUCTION

A novel coronavirus that causes the contagious respiratory disease COVID-19 was first discovered in Wuhan, China, in December 2019. Since the virus's global distribution to more over 200 nations and regions, there has been a pandemic [1]. On January 30, 2020, the WHO classified COVID-19 a public health crisis of global concern. Worldwide, and

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

over 160 million cases and much more than 3.3 million fatalities have been recorded as of May 2021 [2]. Nations all over the world noticed this respiratory infection right away as examples of Covid-19 were quickly spreading. Since the moment it was recognized in China [5], a growing number of people was tainted with Covid-19. Since it was announced as an epidemic disease, the WHO has distributed directions for all nations regarding this infection, which includes how people can make a difference. The disease is stained, how to not be affected by infection, and what safety precautions need to be taken. When to see a clinic, the degree of contaminated situations, and the impact of this infection following a thorough examination of contaminated individuals [2]. The period from the virus's exposure and the start of symptoms is known as the COVID-19 incubation period. It is predicted to last between 2 to 14 days, with 5 days serving as the median [3].It can also be spread through contact with contaminated surfaces or objects, and close contact with an infected person. To reduce the risk of transmission, it is important to practice social distancing, wear a face covering, wash your hands often, and avoid touching your face [4]. As such, the side effects of this infection do not show the power fields with the framework, so if we interact with the infected person, but with a solid un-sensing framework then we can be stained [5]. COVID-19 Pandemic in the most affected nations has had a devastating effect on their economies. Many countries have seen their GDP shrink and unemployment rise as businesses have been forced to close due to the pandemic [6]. In addition, many of the most affected nations have had to implement strict social distancing measures and lockdowns, which have further limited economic activity. Governments have had to provide financial aid and support to individuals and businesses to help them cope with the economic effect of the epidemic [7]. In some cases, governments had to take extreme measures such as printing more money or issuing debt to finance their efforts. In the long term, it’s expected that the economic effects of the epidemic will be felt for years to come, with some countries struggling to recover from the economic damage caused by the pandemic [8]. Medical applications of machine learning are becoming increasingly popular as the technology continues to improve. Huge volumes of data can be examined by ML algorithms to spot trends and patterns that might not be visible to the naked eye. This has the ability to identify illnesses, forecast results, and suggest therapies. The battle against COVID-19 has been significantly aided by machine learning. It is being utilized to create predictive algorithms that foresee the transmission of the virus and identify possible hotspots. It is also being used to develop AI-based diagnostic tools to help diagnose patients quickly and accurately. Machine learning is also being used to develop AI-based drug discovery and development tools to identify potential treatments and vaccines. The branch of study known as machine learning (ML) enables the program to learn without explicitly modifying it [9]. AI's subset of machine learning (ML) has been moved out of design recognition, where data can be organized for consumer understanding [10]. For a long time, several applications based on ML have been established in a variety of industries, including medical services, banking, military equipment, space exploration, and so on. So far, machine learning is a fast-moving and constant creature sector. Machine learning algorithms can be trained on large datasets to learn patterns and make predictions [11]. These predictions can be used

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

to help businesses make decisions and automate tasks. Covid-19, a brand-new Coronavirus, was identified for the first time in 2019. This is comparable to the virus known as SARS-Cov. Only a small number of cases were found in the Middle East. By June 2019, every patient had healed. However, the potential for this virus to cause a global outbreak is high, and research is needed to understand its transmission, severity, and treatment. The outbreak Research in this field is necessary because of the Covid-19 outbreak. As a result, many researchers offer their opinions and recommendations for addressing this epidemic, including China utilising a SEIR / AI model that was successful in forecasting Covid-19 infections [15]. Predicts the likelihood of a serious illness for patients with Covid-19 who are hospitalised and determines their likelihood of survival. As a result, the (Linear regressions) logistic regression models used. The score calculates the likelihood of a serious illness for patients with Covid-19 who are hospitalised [16].To reduce medical burden and potentially reduce the Covid-19 rate of mortality ML: X Gboost used Survival as a result accuracy was 100% to predict high-risk patients and potentially reduce mortality. The accuracy of the prediction was 81% [17]. To predict skin detection of high-risk cases increasing in serious Covid-19, statistical type of analysis was used as a results it helped to identify the Covid-19 patient for effective management [18]. The development of a stochastic mathematical model to forecast Covid-19 cases, susceptible-exposed-infectious-recovered aiding in medical care and resource distribution. According to the model, when 55% to 65% of the population is impacted, the flock can be exempted [19]. Utilizing LSTM-based approaches, the prognosis for new Covid-19 cases is realistic and consistent with original data [20]. the development of a forecasting model to assess clinical needs Supply and medical support will be tested, and R Naught's projections will aid in policymakers' planning [21].COVID-19 Positives to develop a model to predict the risk for patients, ML models Compartmental Models in epidemiology, regression model can directly help to prepare health care [22].Machine learning is use to help fight against COVID-19 pandemic in a variety of ways. It can used to forecast spread of the virus, identify high-threat areas, and develop strategies to contain the virus. Machine learning can also be used to identify potential treatments and develop new vaccines. Additionally, machine learning can use to examine huge amounts of data to better understand the virus and its effects [12].Time series prediction is the process of using historical data to predict future values of a given variable. This type of prediction is used in many areas such as stock market forecasting, weather forecasting, and economic forecasting [13]. Time series prediction uses a variety of methods, including linear regression, autoregressive models, and artificial neural networks. The accuracy of the predictions depends on the quality of the data and the chosen model. The ML-based practices that were demanding forecasting global effects and trends of the Covid-19 epidemic presented in this section. SVM, Polynomial Regression, and Naive Bayes are effective machine-learning algorithms for disease diagnosis [14] here, we will briefly explore the predictive algorithms. The literature work demonstrates the effectiveness of these three techniques to forecast various diseases, which encourages the use of these methods for the prediction of novel Coronaviruses.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

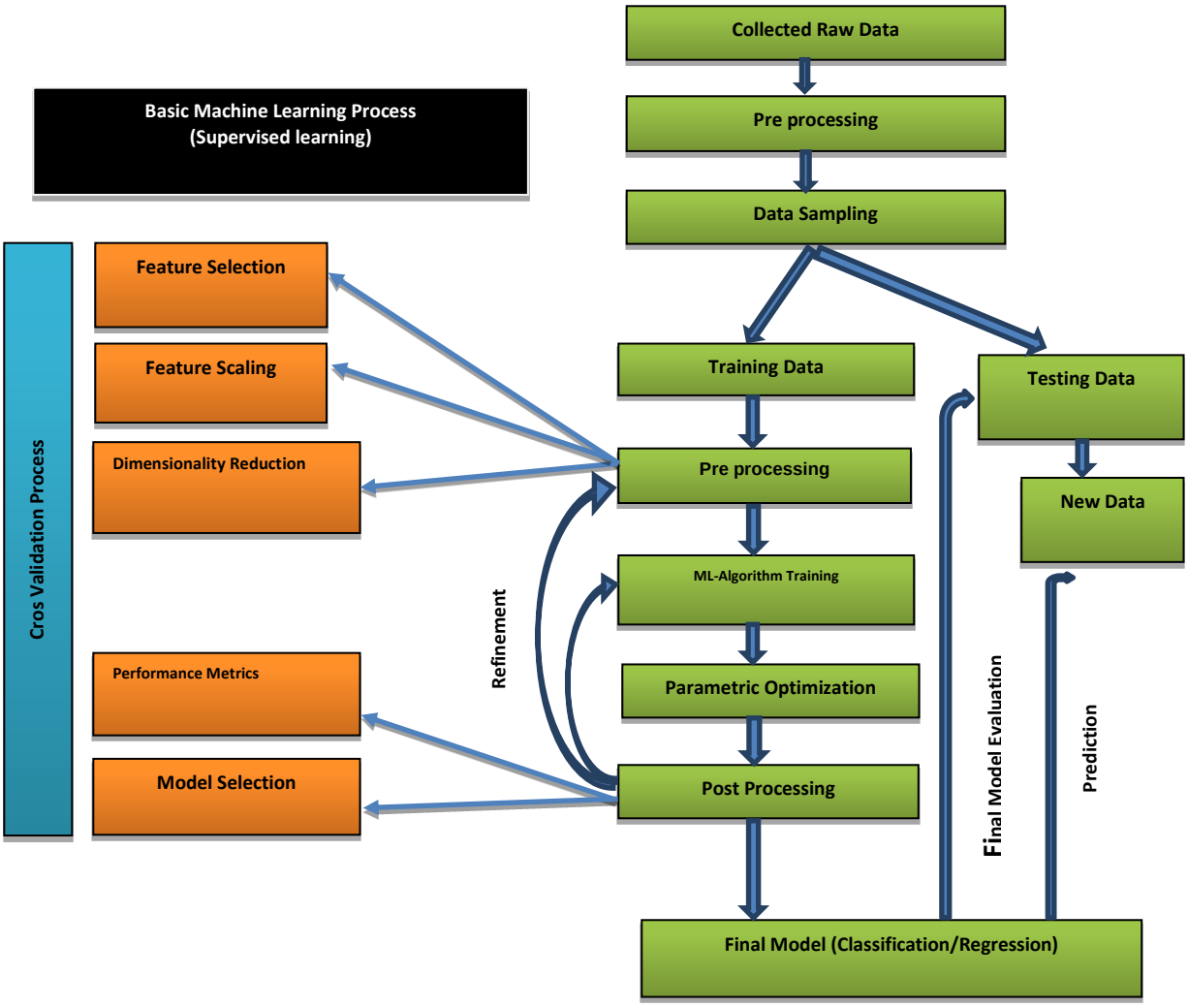

Figure 1: Machine Learning Flowchart process

2. METHODOLOGY

Machine learning algorithms provide the steps and ideas that a system should consider before resolving a specific problem. These algorithms analyse and simulation data to predict results within a specified range. Additionally, some of these methods recognize previous performances in outcome prediction and adjust, optimise, and enhance when new data is added. Simply said, machine learning algorithms tend to become "smarter" over time.

Depending on the type of algorithm employed, machine learning models assess data and produce precise findings using a range of parameters, such as the gamma factor, max depth, n neighbours, and others. Those parameters are the outcome of a test dataset,

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

which is a smaller dataset overall. The Covid-19 pandemic's cumulative count and time-series dataset have been the subject of experiments using Python's Jupiter Notebook. Our goal is to assess and forecast future Covid cases using ML algorithms based on historical trends. SVM, and Polynomial Regression techniques—among the most effective predicting techniques—have all been used and comparison tested for this purpose. A supervised approach to machine learning called "Support Vector Machine" (SVM) can be applied to classification or regression problems. However, it is most typically used in classification-related concerns. When employing a Svm classifier, every parameter is represented as a single, n-dimensional dimension (n being the number of features), with the value of each feature being the value of a particular coordinate. We next carry out computations by locating the hyper-plane that successfully separates the two classes. Simply stated, support vectors are the coordinates of a single observation. The SVM classifier is a frontier that best distinguishes between the two classes (hyper-plane/line). A relationship between a single dependent variable (y) and a single independent variable (x) is described as an nth degree polynomial in the regression technique known as "Polynomial Regression." As we observed with the Simple Regression Model, a linear regression works well when applied to a linear dataset. However, if we use the same model, unchanged, to a non-linear test set, the results would be drastically different. The loss function will grow as a result, leading to a high mistake rate and declining accuracy. We therefore require the Polynomial Predictive model for such situations when the data points are organised non-linearly. The architecture, which is shown in Figure 2, depicts the workflow and steps involved in applying a prediction model to the Covid-19 pandemic dataset. The process starts with domain understanding, in which the problem is assessed and the goal of the problem is stated. The understanding of data structure is the second phase, which must be completed before any problem can be implemented. The third feature, selection, is a crucial stage in which it must be determined which aspect of the data will be used to make future predictions and analyze which attributes will be directly connected to prediction. Next, the realistic Covid-19 dataset has been used to test the prediction methods Polynomial Regression and SVM. The comparative analysis of the algorithms used to the anticipate global spread of Covid-19 is represented by the final phase.

The gathering of data is the first step in the forecasting process. For reliable predicting results, an accurate dataset is necessary. To forecast the global effects and trends, the Covid-19 outbreak's actual time-series dataset was used. The dataset was gathered from GitHub and Kaggle.com, a well-known service that offers helpful datasets. The experiments on Covid-19 prediction have been carried out using various datasets. In the training set referred to as the learning phase, the model has been trained. Once the machine has gained knowledge of the characteristics and traits of the data, it applies it to a test set to make predictions in the future. To obtain more precise predictions and findings for the Covid-19 epidemic, 75% of the data was utilized for testing and 25 percent respectively for learning. Instead of the smaller testing set, the larger testing set assures greater precision of forecasts. The performance of AI algorithms must be improved by

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

hyper parameter tuning. The SVM, and Polynomial Regression algorithms each have a set of chosen hyper parameters.

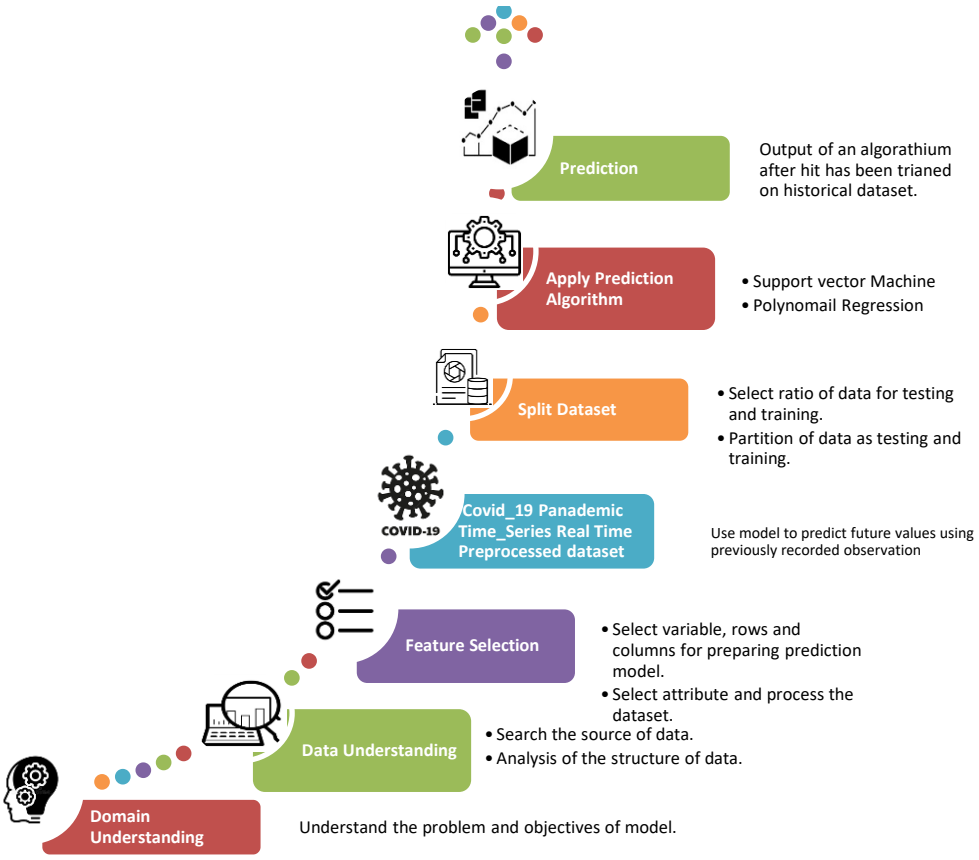

Figure 2: Analytical flow procedure of COVID-19 using Machine learning techniques

3. RESULTS

3.1 Comparative Discussion

The forecasts for Covid-19 that have been generated using various AI-based techniques are displayed in the sections at the top. While all of the tactics improve accuracy and reduce false positives, they each have advantages and limitations. Table 1 shows the estimated MAE and MSE for Polynomial regression and SVM. SVM utilizes support vectors and hyperplanes, where the support vector is made up of information values that seem to be nearer to the hyperplane and have an impact on its location and orientation. The primary goal of this method is to identify the optimum hyperplane with the largest margin. Gamma, epsilon, shrinkage, and degrees were among the top hyper parameters which we evaluated this approach with on the Covid-19 pandemic dataset. The prediction technique was more accurate when these hyper-parameters were configured to their maximum potential. The SVM analysis of Covid-19 time series data produces MAE =

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

265489.6 and MSE = 1397118600180.2, which are better than the Polynomail regression MAE and MSE. The gap between the testing dataset and SVM forecasts increases when MAE and MSE are high, indicating that the SVM methodology is much less effective.

Figure 3: Graph between the Test dataset and the SVM model

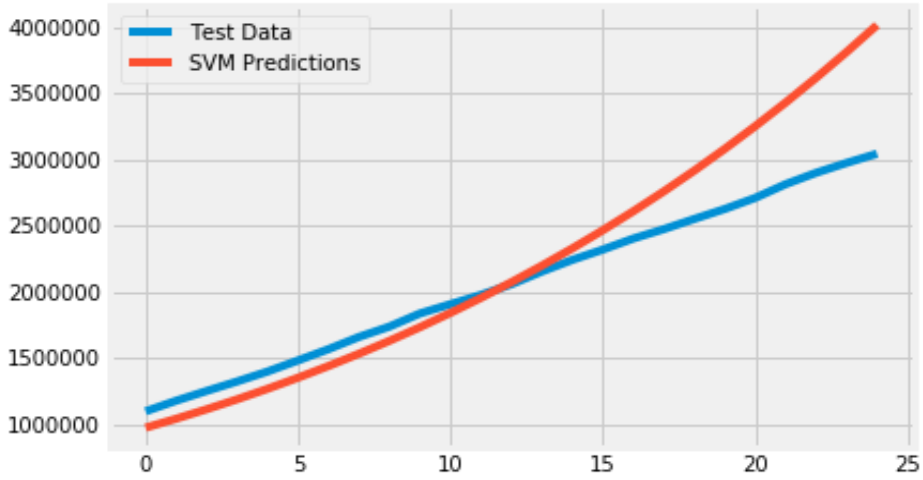

Figure 4: Graph between the Test dataset and PR model

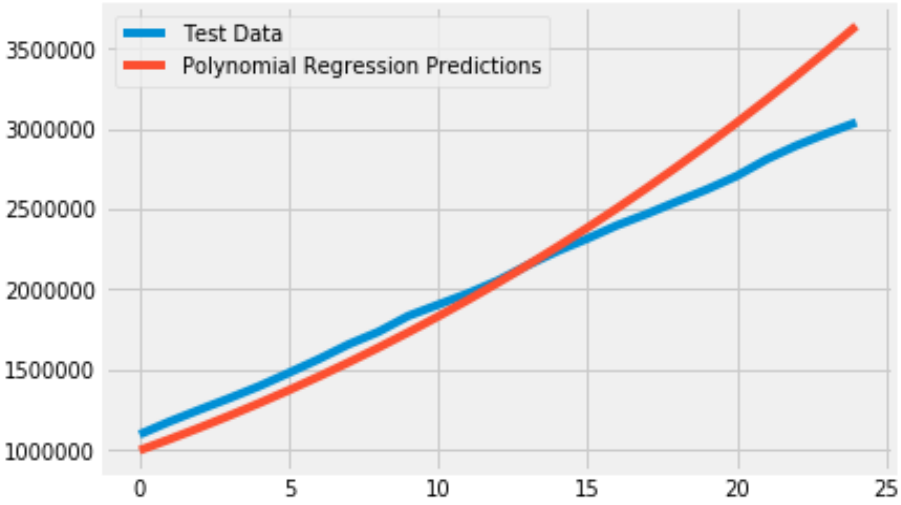

Table 1: MAE & MSE Techniques score

Names	MAE	MSE
SVM	265489.6	1397118600180.2
PR	172343.8	54097652301.4

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

4. CONCLUSION AND FUTURE WORK

Since the Covid-19 epidemic, scientists and healthcare groups from all over the world have advocated for the development of fresh forecasts and quick screening techniques for predictions. The application of artificial intelligence and machine learning (AI) in healthcare companies is widespread. To foresee the outbreak and its effects, we employed machine-learning techniques including SVM and regression registration. Polynomial regression has lower MSE and MAE values and performs better predictions. The framework of this study of Covid-19 trends prediction is also covered in this report, illustrating that AI and machine learning are becoming more and more common in forecasting, screening, drug discovery, and contact tracing. AI is helpful in the treatment of Covid-19 patients, but it also helps the government make the best decisions. Although most AI strategies cannot be used in a real-world setting, they are nevertheless impressive when dealing with outbreaks.

Acknowledgment

Thanks to our teachers, friends and family for providing moral support and encouragement throughout the research activity.

Funding Statement

The authors are pleased to announce that this research is sponsored by The Superior University, Lahore, Pakistan.

Conflicts Of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

1) Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y. ... & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China a descriptive.

2) Pokhrel, P., Hu, C., & Mao, H. (2020). Detecting the coronavirus (COVID-19). ACS sensors, 5(8), 2283-2296.

3) Zaki, N., & Mohamed, E. A. (2021). The estimations of the COVID-19 incubation period a scoping reviews of the literature. Journal of infection and public health, 14(5), 638-646.

4) Rai, B., Shukla, A., & Dwivedi, L. K. (2021). Incubation period for COVID-19 a systematic review and meta-analysis. Journal of Public Health, 1-8.

5) Elias, C., Sekri, A., Leblanc, P., Cucherat, M., & Vanhems, P. (2021). The incubation period of COVID-19 a meta-analysis. International Journal of Infectious Diseases, 104, 708-710.

6) Wang, H. Y., Li, X. L., Yan, Z. R., Sun, X. P., Han, J., & Zhang, B. W. (2020). Potential neurological symptoms of COVID-19. Therapeutic advances in neurological disorders, 13, 1756286420917830.

7) Her, M. (2020). How is COVID-19 affecting South Korea what is our current strategy. Disaster medicine and public health preparedness, 14(5), 684-686.

8) Burkov, A. (2019). The hundred-page machine learning book (Vol. 1, p. 32). Quebec City, QC, Canada Andriy Burkov.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 56 Issue: 02: 2023
DOI10.17605/OSF.IO/YP9FN

9) Gupta, D., Julka, A., Jain, S., Aggarwal, T., Khanna, A., Arunkumar, N., & de Albuquerque, V. H. C. (2018). Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. Cognitive systems research.

10) Al-Turaiki, I., Alshahrani, M., & Almutairi, T. (2016). Building predictive models for MERS-CoV infections using data mining techniques. Journal of Infection and Public Health, 9(6), 744-748.

11) Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons & Fractals, 135, 109864.

12) Gong J, Ou J, Qiu X, et al. A tool for early prediction of severe coronavirus disease 2019 (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020; 71:833−40.

13) Qi X, Jiang Z, Yu Q, et al. Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study [Preprint]. Posted 2020 Mar 3. medRxiv 2020.02.29.20029603. https://doi.org/10.1101/20 20.02.29.20029603.

14) Ghosal S, Sengupta S, Majumder M, et al. Polynomial Regression Analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases - March 14th 2020). Diabetes Metab Syndr 2020; 14:311−5.

15) Salgotra R, Gandomi M, Gandomi AH. Evolutionary modelling of the COVID-19 pandemic in fifteen most affected countries. Chaos Solitons Fractals 2020; 140:110118.

16) Dutta S, Bandyopadhyay SK. Machine learning approach for confirmation of COVID-19 cases: positive, negative, death and release [Preprint]. Posted 2020 Mar 30. medRxiv 2020.03.25.20043505. https://doi.org/10.1101/2020.03.25.20043505.

17) Zhao Z, Chen A, Hou W, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. PLoS One 2020;15:e0236618.

18) Hernandez-Matamoros A, Fujita H, Hayashi T, et al. Forecasting of COVID19 per regions using ARIMA models and polynomial functions. Appl Soft Comput 2020; 96:106610.

19) Chatterjee K, Chatterjee K, Kumar A, et al. Healthcare impact of COVID-19 epidemic in India: a stochastic mathematical model. Med J Armed Forces India 2020; 76:147−55.

20) Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Sci Total Environ 2020; 728:138762.

21) Pandey G, Chaudhary P, Gupta R, et al. SEIR and Regression Model based COVID-19 outbreak predictions in India [Preprint]. Posted 2020 Apr 1. arXiv:2004.00958. https://arxiv.org/abs/2004.00958.

22) Jehi L, Ji X, Milinovich A, et al. Individualizing risk prediction for positive coronavirus disease 2019 testing: results from 11,672 patients. Chest 2020; 158:1364−75.