CURRENT RESEARCH IN
COMPUTER SCIENCE
ISSN 2788-7707
MAY. 2021 VOL. 1, NO. 1

Analysis and Prediction of Outbreaks Using ML Models

Muchahari McKnight, Sanchez-Rola

Department of Computer Science, College of Computing, Debre Berhan University, Debre Berhan, Ethiopia

ABSTRACT

ML algorithm used dataset time series coronavirus dataset to predict the risk of Covid-19. The forecast of Covid-19 is positive, as there are fears all over the world that the epidemic will impact a lot of people around the world. In this analysis, ML algorithms were used to predict the effects of coronavirus outbreaks in Ethiopia. Anticipation of an outbreak in Ethiopia may help policy makers and governments to take inclusive and necessary action. In this research, SVM and PR models were used to forecast the spread of pandemic trends in Ethiopia. In this research, SVM and PR models were used to forecast the spread of pandemic trends in Ethiopia. Typical datasets contain a time series of actual data and make predictions for the next 30 days using SVM and PR. Shape the investigation outcome to check that SVM performs better in current confirmed cases, recovered cases and reported dataset deaths. According to the forecast model, the confirmed case and death rate will rise over the next 30 days. The PR model shows weak results in all three situations. Because of the complexity of the data set. We conclude that ML prediction depends on current data, which can allow us to know the next state. It was used to take remedial action for the body concerned.

KEYWORDS

Covid-19; Forecasting; Support vector machine; Machine learning; Coronavirus; Polynomial regressing.

1. Introduction

On December 31, 2019, Coronavirus has rapidly become a pandemic disease is named carved-19 and the virus is labeled SARS-CoV-2 [11]. The disease has become a rising communicable disease that encompasses rapid spreading; the virus risking the health of large numbers of people, and thus needs speedy actions to prevent the disease at the public level [3]. Covid-19 spreads very speedily from one person to another person by respiratory droplet [15]. The spread of coronavirus mainly of droplets that produce from infected persons during coughing, sneezing, and direct contact of infected people. These may happen directly in contact or

indirect contact with an infected person. Since December 2019 the first reported case in the coronavirus outbreak was announced in Wuhan, China [3]. The first country outside of china was reported in Thailand in the second week of January 2020[3]. The endemic quickly expanded and spread from Hubei Region to other parts of China due to the great movement of personnel during the Chinese Lunar New Year, while the number of COVID-19 cases in other countries rapidly increased [10].

To prevent the spread of COVID-19 who advice there is a lot of things that the people can do to minimize the spread of COVID-19 just do avoid gathering, stay at home, avoid physical contacting, maintain a safe distance at least two meters apart, wash hands with water and soap frequently and isolate themselves if any sign of disease. This information may not have available to the community to reach accurately. This is difficult for peoples to search consistent information and guidance when people want [3]. The WHO occupied endlessly to identify the most widespread rumors that can believably harm community health. The most exciting aspect of its spread is that a person can hold the virus for many days without showing signs. WHO advises on coronavirus with different social media like Twitter, Facebook, Instagram, LinkedIn, and others on their website.

Basically, the outbreak of coronavirus obviously developing a major international crisis, and learning influences the most important aspect of life. It disturbs the political, social, economic, religious, and financial structure of the globe. As of June 10, 2020, there have been over 7.2 million cases of Covid-19 in the world and more than 411,000 deaths [23] according to the John Hopkins university us on top with 1979411 confirmed cases and 111989 deaths. In Africa confirmed 202,864 cases and 5539 number of deaths within a number of African countries [23]. Corona covers around 210 countries.

On June 20, 2020, Ethiopia has reported 4469 confirmed cases and 72 death cases due to Covid-19 [24]. The cases are reported from the two town administrations (Addis Ababa and Diredawa) and the regional states of the country. The ministry of health and Ethiopian public health institutes has reported the cases. As the Covide-19 continues to spread in Ethiopia, the country will be able to tackle this pandemic. So there is need a strong model that predicts how the virus would be spread across different towns and regional states. The aim of this work is to build a model that analyzes and predicts the spread of the virus in next month's using machine learning algorithms.

Machine Learning (ML) is an application of artificial intelligence that provides the system with the ability to automatically learn and improve the experience without an explicit program [2]. It uses the data to detect patterns in the dataset and regulate program action accordingly [7]. ML focuses on the development of a computer program that can teach them to grow and change when exposed to new data. ML is a method of data analysis and predicts that automates analytical model building. ML is used to analyses and predicts many diseases like Covid-19. ML algorithms like Support Vector Machine, Linear Regression polynomial regression, etc, are used by several scientists in the health care area to predict, detect, and analysis for a variety of epidemics like Coronavirus.

At the time of writing this paper, the sum of confirmed cases recovered cases and death reported are estimated to exceed 4469, 1122, and 72 cases respectively, and reach the peak of this outbreak before 20 June 2020. The rest of the paper is organized as follows: Section 1 brief introduction of the study area about coronaviruses and machine learning. Section 2 related works, assess advanced and recent work done before from the machine learning algorithm on Covid-19. Section 3 methodology, understanding and processing the art dataset, define problem statements and design the proposed work. Section 4 presents the experimental result, Section 5 conclusion of the work, and remark future work are discussed.

2. Related Works

To analyze and predict Covid-19 disease different researchers had implemented different artificial intelligence and machine learning algorithm. R.Sujath, et al.[2] suggested a model of machine learning forecasting for coronavirus disease. Throughout their research, they suggested a useful model for predicting the spread of Covid-19. They performed different machine learning algorithms, such as linear regression, multilayer perceptron, and vector auto regression, and selected as the best predictors a multilayer layer machine-learning algorithm. F. Amira et al.[3] propose an outbreak of Covid-19 data analysis and forecasting of corona trucks. Throughout their research, they have demonstrated the value of disseminating information that can help increase response time to further minimize the risk of Covid-19. The research identified the current pattern of coronavirus disease as visualized in the corona tracker and also forecast using the SEIR model and analytics.

F. Ardabili, et al. [4] developed a machine-learning model to analyze and forecast the Covid-19 outbreak, and investigated a multilayer perceptron, adaptive network-based, fuzzy inference method. On the basis of their research, they proposed machine learning as well as tools for modeling the Covid-19 outbreak. After the experiment, the results of two Machine Learning models (multilayer perceptron adaptive network-based fuzzy inference system) showed a high simplification capability for long-term predictability. Z.zhao et al.[9] predict Covid-19 in different African countries using a maximum-hunting parameter estimation approach and an improved susceptible exposed infection recovered (SEIR) model. The most practical updated model was adopted, which varies from the standard SEIR models. The model was considered to be the quarantine status and intervention steps and a stronger approximation of the latest reported and emerging patterns in the epidemic. Total hassling was introduced on the basis of the spectrum of the previous parameter intervals.

F. Rustam et al. [7] To boost decision-making, a machine-learning algorithm was used to predict potential patients affected by coronavirus. For their research, they used four typical prediction models, such as linear regression, List absolute shrinkage, vector support, and exponential smoothing. Models were projected for the number of new events. The number of fatalities and the number of deaths recovered over the next few days. Among their models, exponential smoothing is ideally suited to forecast the epidemic.In this study, N.Zheng, et al.[14] provided a hybrid artificial intelligence model to predict coronavirus. Based on the enhanced susceptible infected model, the hybrid artificial intelligence model to predict the coronavirus proposed in their research was integrated with the natural language processing module. The model's predicted outcome was strongly consistent with the real case of the epidemic. The proposed hybrid model may therefore be used to examine more specifically the evolution of the transmission law pattern of the epidemic.

3. Methods and Materials

Data were collected from the Github repository data repository by John Hopkins university center of system science and engineering (CSSE) for all countries, including Ethiopia total cases since 1/22/2020 [23]. The dataset consist of daily cases reported. The dataset contains daily time series summary data.

For this work we have taken time series confirmed cases, time series recoveries cases, time series deaths reported tables in CSV format. From these datasets extracted Ethiopian cases. Data samples from the tail of the dataset as shown following table.

Table 1. Time series Covid-19 Ethiopia Cases

DATE	COUNTRY	LAT	LONG	CONFIRMED	RECOVERED	DEATHS
6/17/2020	ETHIOPIA	9.145	40.4897	3630	738	63
6/18/2020	ETHIOPIA	9.145	40.4897	3759	849	65
6/19/2020	ETHIOPIA	9.145	40.4897	3954	934	72
6/20/2020	ETHIOPIA	9.145	40.4897	4070	1029	72

6/21/2020	ETHIOPIA	9.145	40.4897	4469	1122	72

Programming Language

A predictive model is built using python language with Jupiter notebook to data analysis and predicts the daily number of coronaviruses. For data analysis, we used python language with a package NumPy, pandas, matplotlib, seaborn, DateTime, plotly SkLearn.

Machine Learning Models

Machine Learning (ML) is one of the emerging technologies under a subset of Artificial Intelligence (AI). ML is focusing mainly on the designing of a predicting model thereby allowing them to learn and make predictions based on some experience, which is data in the case of the machine. ML uses statistical methods to enable the machine to Improve with experience. ML performs based on the idea extraction of the dataset [7]. ML algorithms are trained on labeled and unlabeled train datasets to predict the use of the model. Machine learning is divided into three main categories, supervised, unsupervised, and reinforcement. Supervised ML is mathematical there is an input variable X and produces an output variable Y and uses an algorithm to learn the mapping function from the input to the output that is Y=f(X). The goal of mapping function input X could predict the output Y [7].

Support Vector Machine (SVM)

For this work, we use a type of supervised algorithm to predict Coronavirus. Support Vector Machine (SVM) is a supervised ML classification as well as regression algorithm [7]. SVM has extremely popular and produces an efficient prediction algorithm nowadays. SVM implemented slightly different from other ML algorithms; it is capable of performing classification, regression, and out layer prediction as well. SVM is a discriminative classifier that is formally designed by a separate hyperplane. It is a representation of examples as points in space that are mapped so that the points of different categories are separated by a gap as wide as possible. The simplest construction of SVM is the linear one, where the hyperplane lies in the space of the input data x. In this case, the hypothesis space is a subset of all hyperplanes of the form. Learning can be regarded as finding the maximum margin separating hyperplane between two classes of points. Suppose that a pair (w,b) defines a hyperplane which has the following equation:

$$f(x) = w\,x + b.$$

When two variables have been curvilinear relationship polynomial model is useful. The polynomial model has been useful in exhausting the representation of the relationship between training and training depth [33]. PR is a general Polynomial regression model for fitting any relationship that is polynomial in the unknown parameters, β, the following polynomial a nth order polynomial in one variable is defined as

$$y = \beta 0 + \beta 1x + \beta 2x\,2 + \cdots + \beta kxn + E$$

Polynomial models are useful in circumstances where the forecaster knows that curvilinear outcomes are present in the actual reaction function. The degree of the polynomial model shall be kept as low as possible. Any transformations can be used to hold the model in the first degree.

For a prediction model, a machine-learning algorithm called a support vector machine (SVM) and a polynomial regression are used. For this work, a time series for casting is performed on the data collected and the model constructed. The number of confirmed cases has been recorded since 1/22/2020. Regular increases

the details of the confirmed cases and their conforming days are given as feedback for the relationship between the two variables. Analysis and predictive models are based on the premise that the pattern of coronavirus pandemics in Ethiopia is close to that of other countries, such as Italy, China, the US and others, over time.

After extracting the time series dataset as shown in Table. the initial data preprocessing step, the dataset has been divided in to training and test set. For this study 75% of data are used for training set and 25% of data are for testing set. We train the model (SVM and PR) by training dataset. SVM and PR have been trained on the daily time series data's such as confirmed cases, recovered cases and death reported. The proposed work method used in the analysis has been shown as a Figure 1.

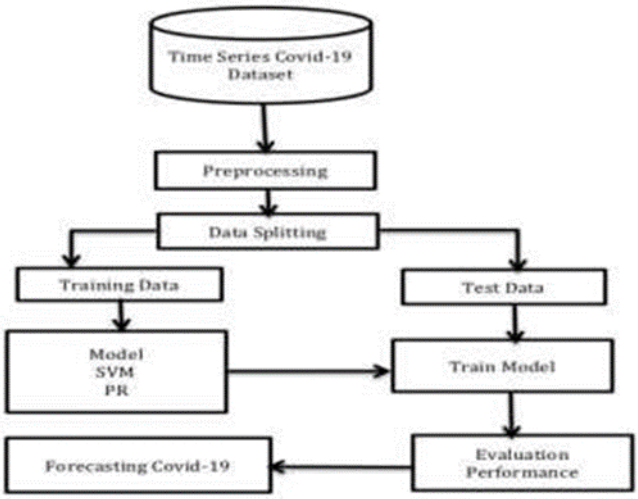

Figure 1. Work flow of the proposed work

4. Experimental Result

Out of a number of confirmed cases until June 21/2020, Ethiopia confirmed 4469 cases, 72 cases of death and 1122 cases of recovery [23]. The statistics for the current data[23] of the first confirmed cases are available in Ethiopia on March 13/2020. Since March 13/2020, the number of confirmed cases of Ethiopia has increased slowly until May/2020. Figure 1 shows the growth of confirmed cases in Ethiopia. Even if Ethiopia has put in place strong control measures, including quarantine, to monitor all cases at home and to test all close contacts, the number of confirmed cases should have increased cases in Ethiopia. From the daily increase in the number of cases, we analyze the number of confirmed cases. The number of confirmed cases is shown in Figure 1 from January 22/2020 to June 21/2020. From this statistic, we infer that the total number of confirmed cases in Ethiopia will increase rapidly after May 2020. As a result of this rise in number, more than ten thousand confirmed cases would be confirmed by the mid of July 2020.

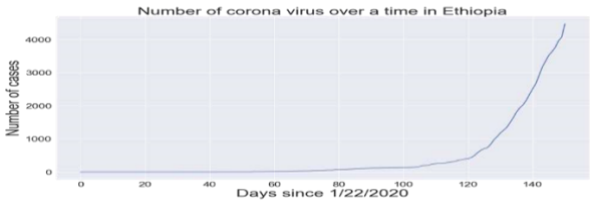

Figure 2. Number of confirmed cases in Ethiopia

Accordingly, the number of cases recovered is shown in Figure 2 from January 22/2020 to June 21/2020. This indicates the average number of coronaviruses recovered from the pandemic. The number of coronavirus cases recovered is expected to increase by the mid of July 2020 as increase number of cases.

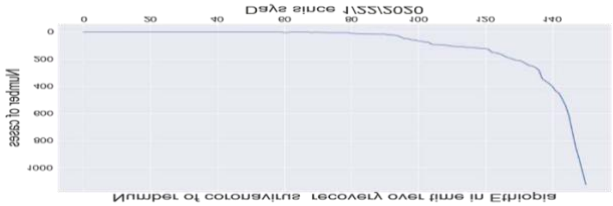

Figure 3. Number of recovered cases in Ethiopia

In the same way, we examine the regular number of deaths in Ethiopia, as shown in Figure 3; the number of deaths is projected to rise after June 2020. By the mid of July, Ethiopia will have more than 10 thousand confirmed cases, and there will be a shortage of hospitals. It may suggest a rise in the mortality rate.

Figure 4. Number of death case in Ethiopia

The overall confirmed case and mortality rate in Ethiopia is likely to be lower than in the rest of the world. As we agree that Ethiopia is on a lower scale compared to other countries. It is feared that resources and an unequal public health system will swamp the patient. Ethiopia should buy time to stock up on test kits and install quarantine places and hospital beds. We needed to expand quickly to restrict the transmission of the community. Figure 4 shows the confirmed cases, recovered cases and death cases in Ethiopia. From all the three cases graphs, the blue line shows confirmed cases, the green line shows the number of recoveries and the red line shows the number of coronavirus deaths. This is actually good number of recovery is more that number of deaths.

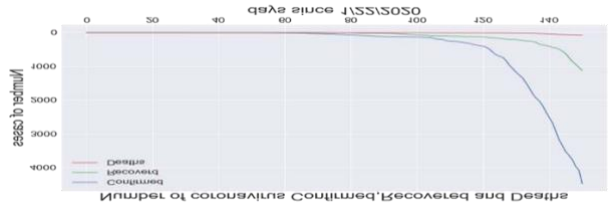

Figure 5. Number confirmed, recovered and deaths of case in Ethiopia

This work seeks to develop a model for future forecasting of the number of confirmed, recovered and fatal cases affected by coronavirus using ML algorithms. As the number of cases increases day by day, the country is in a state of panic. Research lets the country seek to forecast the number of people who might be involved in reported incidents, injuries and recovery incidents for the next 30 days. SVM and PR were used to predict confirmed, reported death and recovery.

Table 2: Performance of model on future prediction of cases

DATASET	MODEL	MAE	MSE
CONFIRMED CASES	PR	1415.92	3170874.90
	SVM	1372.86	1909195.93
RECOVERED CASE	PR	107.24	39316.94

	SVM	98.44	6264.21
DEATH	PR	16.77	701.70
REPORTED	SVM	98.44	26264.20

Figure 5 and 6 shows the performance of SVM and PR models respectively on test confirmed cases data in the form of plot graph. The recent confirmed cases of coronavirus increases day to day. In this work SVM is performs very good in terms of performances where as PR is vey poorly in terms of the evaluation metrics. Figure 5 and 6 show the furcating model of learning algorithms. The green line shows how SVM and PR performs prediction on test data

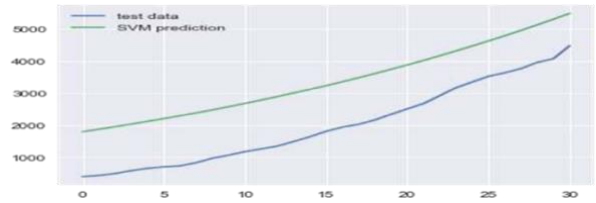

Figure 6. Confirmed case prediction by SVM on test data

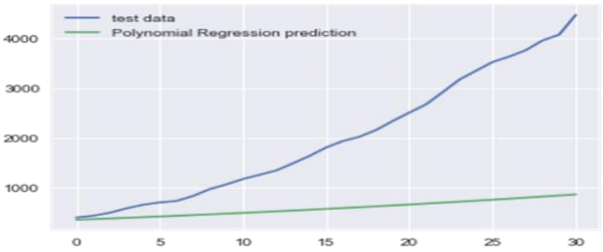

Figure 7. Confirmed case prediction by PR on test data

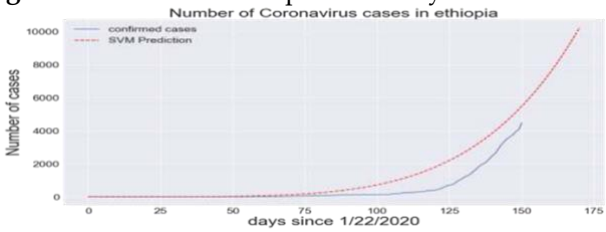

Figure 8. Confirmed case predictions by SVM for the upcoming 30 days

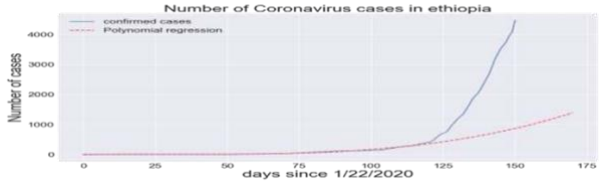

Figure 9. Confirmed case predictions by PR for the upcoming 30 days

Similarly, Figures 9 show the performance of the SVM and PR models on the test case recovered data in the form of a plot graph. In recent reported cases, coronavirus rises daily. In this work, the SVM performs very well in terms of performance where PR is poor in terms of evaluation metrics.

From the over all graphs SVM performs better in all cases such as confirmed cases forecasting recovered case forecasting and death cases forecasting. Bearing in mind the better performance given by SVM model in all the three cases between the two models. So SVM has been used for prediction of Covid-19.

5. Conclusion

For this analysis, the ML algorithm used dataset time series coronavirus dataset to predict the risk of Covid-19. The forecast of Covid-19 is positive, as there are fears all over the world that the epidemic will impact a lot of people around the world. In this analysis, ML algorithms were used to predict the effects of coronavirus outbreaks in Ethiopia. Anticipation of an outbreak in Ethiopia may help policy makers and governments to take inclusive and necessary action. In this research, SVM and PR models were used to forecast the spread of pandemic trends in Ethiopia. In this research, SVM and PR models were used to forecast the spread of pandemic trends in Ethiopia. Typical datasets contain a time series of actual data and make predictions for the next 30 days using SVM and PR. Shape the investigation outcome to check that SVM performs better in current confirmed cases, recovered cases and reported dataset deaths. According to the forecast model, the confirmed case and death rate will rise over the next 30 days. The PR model shows weak results in all three situations. Because of the complexity of the data set. We conclude that ML prediction depends on current data, which can allow us to know the next state. It was used to take remedial action for the body concerned.

In the future, we will be studying furcating using the new dataset and using the most reliable ML models, such as Deep Learning and Multilayer Perceptron. Those MLs that have not been examined in this study may have shown an important trend associated with successful Covid-19 prediction.

References

World Health Organization. (2020). Coronavirus disease 2019 (COVID-19): situation report, 88.

Sujath, R., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. Stochastic Environmental Research and Risk Assessment, 1.

Hamzah, F. B., Lau, C. H., Nazri, H., Ligot, D. V., Lee, G., Tan, C. L., ... & Ong, C. H. (2020). CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. Bull World Health Organ, 1, 32.

Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., ... & Atkinson, P. M. (2020). Covid-19 outbreak prediction with machine learning. Available at SSRN 3580188.

Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020). COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms. medRxiv.

Yadav, D., Maheshwari, H., & Chandra, U. (2020). Outbreak prediction of covid-19 in most susceptible countries. Global Journal of Environmental Science and Management, 6(Special Issue (Covid-19)), 11-20.

Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B., Aslam, W., & Choi, G. S. (2020). COVID-19 Future Forecasting Using Supervised Machine Learning Models. IEEE Access.

Zhu, H., Guo, Q., Li, M., Wang, C., Fang, Z., Wang, P., ... & Xiao, Y. (2020). Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm. BioRxiv.

Zhao, Z., Li, X., Liu, F., Zhu, G., Ma, C., & Wang, L. (2020). Prediction of the COVID-19 spread in African countries and implications for prevention and controls: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. Science of the Total Environment, 138959.

Jia, L., Li, K., Jiang, Y., & Guo, X. (2020). Prediction and analysis of Coronavirus Disease 2019. arXiv preprint arXiv:2003.05447.

Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine, 103792.

Tiwari, S., Kumar, S., & Guleria, K. (2020). Outbreak Trends of Coronavirus Disease-2019 in India: A Prediction. Disaster medicine and public health preparedness, 1-6.

Zheng, N., Du, S., Wang, J., Zhang, H., Cui, W., Kang, Z., ... & Ma, M. (2020). Predicting covid-19 in china using hybrid AI model. IEEE Transactions on Cybernetics.

Yadav, R. S. (2020). Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India. International Journal of Information Technology, 1-10.

Kang, H., Xia, L., Yan, F., Wan, Z., Shi, F., Yuan, H., ... & Shen, D. (2020). Diagnosis of coronavirus disease 2019 (covid-19) with structured latent multi-view representation learning. IEEE transactions on medical imaging.

Nah, K., Otsuki, S., Chowell, G., & Nishiura, H. (2016). Predicting the international spread of Middle East respiratory syndrome (MERS). BMC infectious diseases, 16(1), 1-9.

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.

Allam, Z., Dey, G., & Jones, D. S. (2020). Artificial intelligence (AI) provided early detection of the coronavirus (COVID-19) in China and will influence future Urban health policy internationally. AI, 1(2), 156-165.

Tomar, A., & Gupta, N. (2020). Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Science of The Total Environment, 138762.

Ren, H., Zhao, L., Zhang, A., Song, L., Liao, Y., Lu, W., & Cui, C. (2020). Early forecasting of the potential risk zones of COVID-19 in China's megacities. Science of The Total Environment, 729, 138995.

Yang, Z., Zeng, Z., Wang, K., Wong, S. S., Liang, W., Zanin, M., ... & Liang, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease, 12(3), 165.

Mahalle, P. N., Sable, N. P., Mahalle, N. P., & Shinde, G. R. (2020). Predictive Analytics of COVID-19 Using Information, Communication and Technologies.

Github repository. 2019 Novel Coronavirus COVID-19 (2019-nCoV) DataRepository by Johns Hopkins CSSE. Retrieved June 21, 2020 from https://github.com/CSSEGISandData/COVID-19.

John Hopkins University, "Coronavirus Map," John Hopkins University, 17 March 2020. [Online]. Available: https://coronavirus.jhu.edu/map.html. [Accessed 17 March 2020].

Ethiopian public health institute," Notification Note on COVID-19 Situational Update June10"[Online]. Available: https://covid19.ephi.gov.et/notification-note-on-covid-19-situational-updatejune-10/

Jiang, X., Coffee, M., Bari, A., Wang, J., Jiang, X., Huang, J., ... & Wu, Z. (2020). Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. CMC: Computers, Materials & Continua, 63, 537-51.

Ahire, S., & Prasad, J. (2020). Analytical Insights into COVID19 Pandemic Predictions: A Machine Learning Perspective. Available at SSRN 3604874.

Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., ... & Atkinson, P. M. (2020). Covid-19 outbreak prediction with machine learning. Available at SSRN 3580188.

Zhang, X., Ma, R., & Wang, L. (2020). Predicting turning point, duration and attack rate of COVID19 outbreaks in major Western countries. Chaos, Solitons & Fractals, 109829.

Singh, D., Kumar, V., & Kaur, M. (2020). Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. European Journal of Clinical Microbiology & Infectious Diseases, 1-11.

AlMoammar, A., AlHenaki, L., & Kurdi, H. (2018, September). Selecting accurate classifier models for a MERS-CoV dataset. In Proceedings of SAI Intelligent Systems Conference (pp. 1070-1084). Springer, Cham.

Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. Science of the Total Environment, 138882.

Ostertagová, E. (2012). Modelling using polynomial regression. Procedia Engineering, 48, 500-506.

Copyrights