



## A Survey on Ensemble Learning Approaches for Classification of Real Time COVID-19 Data and Forecasting Severity of the Epidemic

---

Monika Arya and Chaitali Choudhary

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

February 8, 2022

## **A Survey on Ensemble learning Approaches for classification of real time COVID-19 data and forecasting severity of the Epidemic**

Mrs. Monika Arya<sup>2</sup>[0000-0002-2825-2328] and Ms. Chaitali Choudhary<sup>1</sup>[0000-0003-2888-1148]

<sup>1,2</sup> Assistant Professor, Bhilai Institute of Technology, Durg  
chaitali.choudhary@gmail.com, arya.akshara@gmail.com

**Abstract.** Healthcare facility at the side of support of recent technologies like analysis using Artificial intelligence and Machine learning algorithms play a very important role to defend against in prediction, diagnosis and treatment of novel corona virus disease (COVID-19) which emerged in Wuhan City and speedily spreading throughout the globe. In this paper we review the Ensemble learning approaches for classifying and forecasting COVID-19 epidemic which will further help to take preventive steps and necessary actions. For the welfare of human race, in this paper we propose to utilize the benefits of ensemble approach to understand the behavior and pattern of corona virus in order to accurately predict the expected new cases of COVID 19 by using the current and updated streams of information. The information can further be utilized for future prediction of medical requirements. Review of latest related literature including techniques like AI, Machine learning, COVID -19, ensemble approaches etc. The latest information regarding different prediction techniques and analysis methods of COVID 19 data using machine learning, AI and ensemble techniques by different authors were collected and analyzed to identify the role of machine learning and AI for prediction and diagnosis of COVID 19 disease. In this paper, we have used the Ensemble approaches to predict the total active cases, recovered cases, and death cases all over the world. It is implemented using the python library “sklearn”. The prediction can further assist to take the necessary decisions related to lockdown period, medical facilities, government policies and rules etc. In this paper, latest prediction techniques of AI and machine learning for COVID-19 pandemic are identified and concluded that this technology is very significant in detecting the severity of current state and in predicting death cases, forecasting new cases in future and required medical facilities in by collecting and analyzing historical data. We have experimentally compared the ensemble approaches on real time data. Healthcare facilities need to be equipped with proper prediction and decision-making technologies to provide suggestions in real-time. So that better and on-time medical facilities can be provided to handle this virus and help them in to avoid its spread. In this paper we have compared various ensemble learning approaches for prediction of COVID-19. The prediction may further be employed in understanding the pattern of virus and assist in development of a vaccine for COVID-19.

**Keywords:** Corona virus, COVID-19, Artificial intelligence, Machine learning, Ensemble learning, predictive analysis

## 1 INTRODUCTION

COVID-19 was declared as international health threat by WHO (World Health Organization) as a result of up until date no tested medication or remedy is accessible to cure the disease and therefore inflicting high mortality. SARS-CoV-2 is a beta corona virus which is said to be originated from Wuhan. It causes a severe respiratory disorder. SARS-CoV-2 is exceedingly communicable and has quickly spread out all over the globe. The spread of corona virus was alarming and thus was declared as Public Health Emergency of International Concern (PHEIC) by World Health Organization (WHO) as it had spread to 18 countries. WHO named this “COVID-19” and declared this a pandemic as there were more than 1.18 Lakhs cases covering around 110 countries and causing more than 3 thousand deaths till March 11 2020. At this stage fast and accurate assessment of disease is significant. Early clinical assessment can further support in decision making and planning in healthcare systems. Analysis of this disease is very important as it will help in formation of Government action plans to prevent the further spread of this global pandemic. As a contribution towards the well-being of human society, researchers have concentrated their attention in developing models for analyzing and predicting the severity of this epidemic spread using real time and historical data. Several models have upsurged for prediction of COVID-19 and are being officially used round the world for getting current and up-to-date information and taking decisions based on information available. So that control measures can be taken timely. Standard models like simple epidemiological and statistical models are popular in media and have also gained more attention for COVID-19 prediction [5]. The quick and early diagnosis of COVID-19 cases in a large scale is a big challenge for researchers, health industry, government and all health concern. Because not only positive diagnosed patients but also close contacts with active cases need to be confirmed whether they are infected or not as they can become super spreaders. Preventive measures like timely isolation can effectively lessen the spread of the virus. Early detection is also very significant to get more timely treatment at the initial stage of infection. And as up till now there is no proper vaccine or treatment for the disease, the early diagnosis and timely treatment of COVID-19 patients is the only preventive measure. In this article we aim study the ensemble learning approaches which can analyze and help us for prevention and to fight with COVID-19 (Corona virus) and will help us to understand the exponential behavior. Figure 1 depicts the significance of AI and machine learning techniques.

### Role of Machine learning techniques and AI

1. Monitoring the number of currently live COVID19 cases
2. Analyzing and understanding the changing behavior and pattern of COVID19 virus.
3. Analyzing the effect of various factors such as self-quarantine and self-isolation to break the chain.
4. Monitoring the number **of cases treated at home**
5. **Monitoring utilization of hospital beds-** Utilization forecasting uses linear regression models to extrapolate and make predictions based on existing data.
6. Prediction of future cases with such medical diagnosis details.

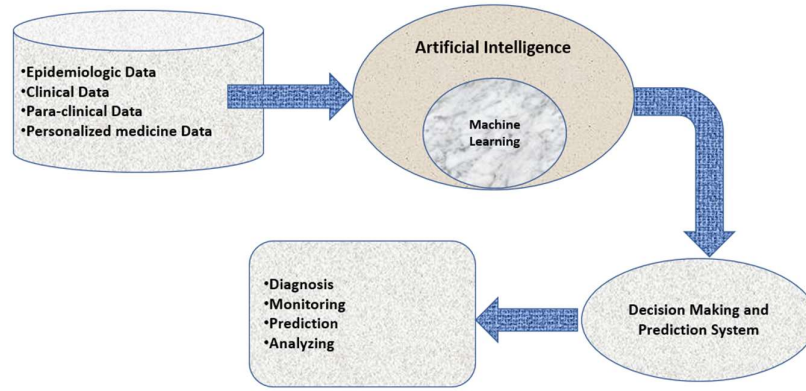


Fig. 1. Role of AI and Machine Learning

## 2 RELATED WORK

Sujath et al. [1] in their work performed linear regression method, MLP and Vector Auto regression method for detecting COVID-2019 cases in India. Yan et al. [2] in their work used a supervised Multi-tree XGBoost classifier as the predictor model and the model has improved accuracy in identification of infected patients. Sekhar et al. [3] in their work suggested Machine intelligence based Generative Network Complex has been used to design drugs based on the experimental structure of SARS-CoV-2 Mpro. The results of this study also rationalize the limited data regarding effectiveness of drugs for COVID-19 therapy, and provide information that can be utilized for choice of candidate drugs for in vitro studies and in vivo studies. Pandey et al. [4] in their work used SEIR model along with Regression model for predictions. The results show that Expected cases may rise in coming weeks of time. The prediction can assist the Government as well as doctors to get prepare for future emergencies. Liu et al. [5] in their work used a cluster based approach which utilized geo-spatial synchronicities of COVID-19 activity across China, and also used data augmentation technique for observations and characteristic of emerging outbreaks. Yang et al. [6] in their work make changes in the original SEIR-equation and the model proved to have improved effectiveness in prediction the COVID-19 cases. Bandyopadhyay et al. [7] suggested a verification method based on Deep-learning Neural Network which will help the doctors for further verification of virus. Wu et al. [8] proposed a more convenient and practical tool called assistant discrimination tool. It assists in antibody detection and this accelerates the identification of patients which may be positive and this can reduce the risk of spread of the disease. They concluded discovered relationship could help to track the progressive behavior of the disease and can guide to make the useful strategies. Pun

et al. [9] in their work suggested models based on machine learning and deep learning. Kassani et al. [10] in their study compared popular deep learning-based feature extraction frameworks for automatic COVID-19 classification.

### 3 EXPERIMENTS AND RESULTS

#### 3.1 Dataset description

All the experiments are performed in COVID-19 dataset which is a collection of the COVID-19 data maintained by Our World in Data. It is updated daily and includes data on confirmed cases, deaths, and testing. The dataset is in CSV format having six attributes. Figure 2 presents the COVID-19 new case and new deaths in India since the 31st Dec 2019 to 29th June 2020. It can be easily observed that the spread growth is exponential and it needs to be controlled.

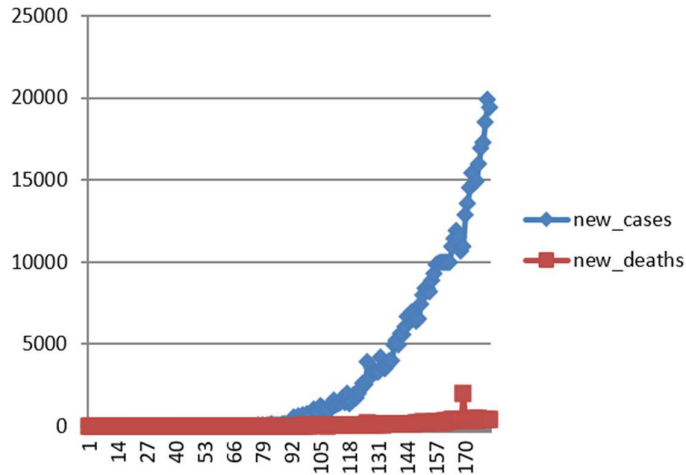


Fig. 2. Exponential growths of COVID-19 Cases

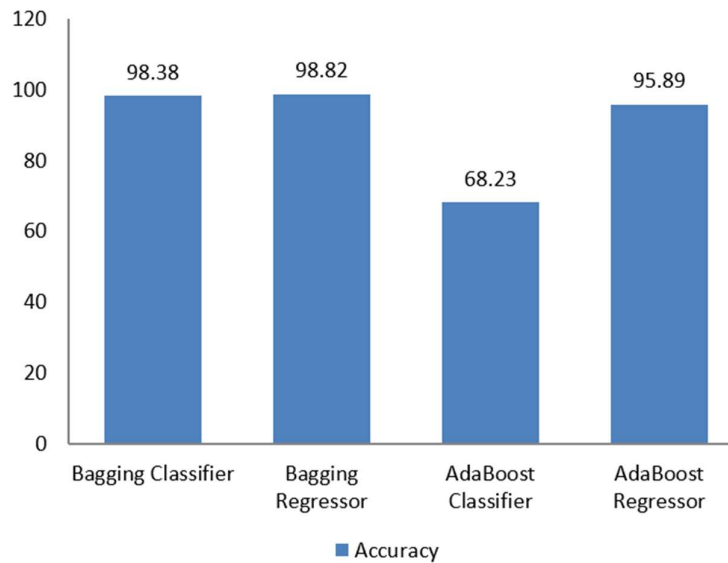
#### 3.2 Ensemble learning predictions accuracies comparison

The COVID-19 spread has caused huge loss of human lives. It is very important to analyze and predict the future possibilities of transmission growth so that the necessary decisions such as expanding the lockdown period, availability of medical facilities, providing the everyday essentials can be taken timely. With this objective, various ensemble models such as decision tree, Adaboost, Random forest, Stochastic gradient Boosting etc are analyzed. For predictions of total death cases worldwide or for

predicting number of infected cases, the Machine learning models are implemented using the python library sklearn and their accuracies are compared in Table 1 and graphically represented in Figure3.

Ensemble Classifiers		Accuracy
<b>Bagging</b>	Bagging Classifier	98.38
	Bagging Regressor	98.82
<b>Boosting</b>	AdaBoost Classifier	68.23
	AdaBoostRegressor	95.89

**Table 1.** Comparison of Accuracies of Ensemble Classifiers



**Fig. 3.** Comparison of accuracies of Ensemble classifiers

#### 4 CONCLUSION AND FUTURE SCOPE

Machine learning techniques have great significance in analysis and prediction of epidemic. The patterns discovered by machine learning techniques can be used for planning remedial and precautionary actions to prevent the growth of virus. In this research Ensemble learning models are employed for analysis and prediction by utilizing the real-time information. These predictions can further be used for forecasting future medical requirements depending upon the expected new cases around the world. However, if the spread follows the predicted exponential behavior then it would result in big life-loss as well as hamper the country financially and economically. Further, this growth

of pandemic can be checked by maintaining zero contacts with vulnerable as well as infected people. This is often realizable by changing into unsocial and strictly following the lockdown rules. In future this study can be further utilized by other machine learning models which can automatically predict the number of upcoming cases for near future. And the predictions can help the government and hospitals to maintain the supply of medical aids.

## 5 References

1. 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.
2. Yan, L., Zhang, H. T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., ...& Huang, X. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence*, 1-6.
3. Sekhar, T. (2020). Virtual Screening based prediction of potential drugs for COVID-19.
4. Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. (2020). SEIR and Regression Model based COVID-19 outbreak predictions in India. *arXiv preprint arXiv:2004.00958*.
5. Liu, D., Clemente, L., Poirier, C., Ding, X., Chinazzi, M., Davis, J. T., ...&Santillana, M. (2020). A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. *arXiv preprint arXiv:2004.04019*.
6. 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.
7. Bandyopadhyay, S. K., &Dutta, S. (2020). Machine learning approach for confirmation of covid-19 cases: Positive, negative, death and release. *medRxiv*.
8. Wu, J., Zhang, P., Zhang, L., Meng, W., Li, J., Tong, C., ...& Zhao, M. (2020). Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. *medRxiv*.
9. Punn, N. S., Sonbhadra, S. K., &Agarwal, S. (2020). COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms. *medRxiv*.
10. Kassani, S. H., Kassani, P. H., Wesolowski, M. J., Schneider, K. A., & Deters, R. (2020). Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning-Based Approach. *arXiv preprint arXiv:2004.10641*.
11. Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, 368(6492), 742-746.
12. Leonenko, V. N., & Ivanov, S. V. (2016). Fitting the SEIR model of seasonal influenza outbreak to the incidence data for Russian cities. *Russian Journal of Numerical Analysis and Mathematical Modelling*, 31(5), 267-279.
13. WHO coronaviruses (COVID-19). Retrieved March 30, 2020 from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
14. Singer, H. M. (2020). Short-term predictions of country-specific COVID-19 infection rates based on power law scaling exponents. *arXiv preprint arXiv:2003.11997*.
15. Dey, S. K., Rahman, M. M., Siddiqi, U. R., & Howlader, A. (2020). Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. *Journal of medical virology*, 92(6), 632-638.