Journal of Current Science and Technology, May-August 2023 Vol. 13 No. 2, pp. 221-236
Copyright ©2018-2023, Rangsit University ISSN 2630-0656 (Online)

Cite this article: Sumathi, D., Poongodi, T., Suresh, P., Karthikeyan, S., Chand, N. S. (2023, May). Visualisation and prediction of Covid-19 data using random forest regression. *Journal of Current Science and Technology, 13*(2), 221-236. https://doi.org/10.59796/jcst.V13N2.2023.1738

Journal of Current Science and Technology
Journal homepage: https://jcst.rsu.ac.th

Visualisation and prediction of Covid-19 data using random forest regression

D. Sumathi[1*], T. Poongodi[2], P. Suresh[3], S. Karthikeyan[4], and N. Sree Chand[1]

[1]School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India
[2]School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh 203201, India
[3]School of Mechanical Engineering, Galgotias University, Greater Noida, Uttar Pradesh 203201, India
[4]KPR Institute of Engineering and Technology, Arasur, Tamil Nadu 641407, India

[*]Corresponding author; E-mail: sumathi.research28@gmail.com

Received 25 May 2022; Revised 17 December 2022; Accepted 18 December 2022;
Published online 15 July 2023

Abstract

The outbreak of COVID-19 has spread among several parts of the world. The data pool increases tremendously, which needs excellent attention by researchers of various domains to analyze and determine the measures to handle it. Hence, researchers worldwide are looking into Artificial Intelligence (AI) to resolve the challenges due to this COVID-19. It could be stated that AI can examine huge data mounds so that several new findings can be determined. AI could be deployed in various fields, such as the pharmaceutical industry, the analysis and development of vaccines and antibodies, and drug designing. Due to the impressive progress that AI has made in the latest few years, still, it proves to be the essential quality of the technology and evidence of humans' creativity towards the contribution of developing tools and products which could be statistically and computationally complex. It is observed that AI technology aids in tracing the outburst, patient diagnosis and fastening the procedure of finding a treatment. This work provides an overview of COVID-19, along with the convergence of technologies that could be applied in various sections to handle this pandemic. An extensive exploration of multiple techniques and models has been implemented for the prediction of COVID-19 has been done. Additionally, numerous models have been deployed for predictions of covid states. It has been inferred that XGBoost showed considerable progress in the prophecy.

Keywords: *artificial intelligence; linear regression; machine learning; random forest; visualization; XGBoost*

1. Introduction

Coronaviruses are enveloped, non-segmented RNA viruses which generally spread among humans, mammals, and birds. Most coronaviruses create mild infections like respiratory, enteric, hepatic, and neurologic problems. Generally, there are six coronaviruses species, 229E, OC43, NL63, HKU1, SARS-CoV, and MERS-CoV (Chan, & Chan, 2013), which causes diseases in human. Typical cold symptoms are caused by four prevalent viruses such as 229E, OC43, NL63, and HKU1. However, infections brought on by the other two viruses, such as Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV) (Petrosillo, Viceconte, Ergonul, Ippolito, & Petersen, 2020). SARS-CoV outbreaks in 2002 and 2003 in China, whereas the outburst of MERS-CoV happened in 2012. The evidence provided by (Perlman, & Vijay, 2016) showed that infections caused by these viruses would create an impact in response due to the chronic illness that occurs in the lungs (Peiris, Yuen, Osterhaus, & Stöhr, 2003).

Due to specific attributes such as diversification in genetics, restructuring of human-animal interfaces, and genomes, viruses emerge periodically in humans, as shown in Figure 1. Some patients with pneumonia for unspecified reasons were identified, and they are epidemiologically linked with JK89, the wet animal wholesale market of Wuhan, China, in late December 2019 (Mackenzie, & Smith, 2020). After that, the Chinese Centre for Disease Control and Prevention team investigated with the help of city help authorities and concluded that the cause of pneumonia clusters is the novel coronavirus.

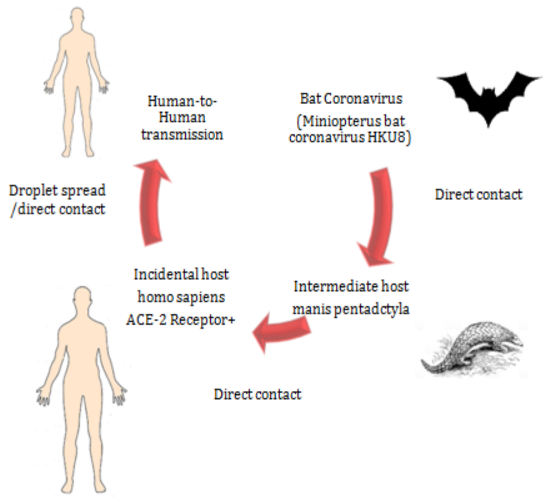

Figure 1 The transmission cycle of COVID-19

The novel coronavirus named SARS-CoV-2 (2019-nCoV) after a deep analysis of lower respiratory tract samples. Commonly, COVID-19 affects the sinuses, nose, and throat of the upper respiratory tract, windpipe, and lungs of the lower respiratory tract. Therefore, indications of COVID-19 could be identified as colds, flu, and allergies. It is shown in Table 1.

There is no specific antiviral treatment or vaccine for COVID-19, yet patients suffering from the severe infection will be provided with oxygen therapy. In addition, mechanical ventilation is given to patients with respiratory failure; however, hemodynamic support is also necessary to manage septic shock. Based on the previous epidemics HCoVs, World Health Organization, on January 28, 2020, summarized the guidelines to prioritize the patients with severe critical respiratory disease, methods to prevent infection; early symptoms and diagnosis; treatment for respiratory failure and septic shock; guidelines for treatments; and steps followed for pregnant patients. The clinical representation of humans with COVID-19 is shown in Figure 2.

Table 1 Symptoms of COVID-19

Symptoms	Cold	Flu	Allergies	COVID-19
Headache	infrequent	severe	infrequent	may occur
General pain	mild	severe	no	may occur
Tiredness	less	intense	rarely	may occur
Fever	infrequent	3-4 days, high (100-102 F)	no	frequent
Runny nose / Sneezing	common	infrequent	can occur	may occur
Sore throat	usual	usual	infrequent	identified
Cough	low to medium	usual, may result in severe	infrequent	usual
Breathing problem	sometimes	sometimes	sometimes	serious infection

1.1 Modes of transmission

The primary cause of the COVID-19 virus is respiratory droplets and physical contact. The contamination spread of COVID-19 occurs primarily due to incessant contact of infected people and unintended contact on surfaces of objects used by the infected persons. In Airborne transmission, a microbe in the droplet is less than 5µm in diameter and transmitted to another person at a distance of more than one meter. Airborne transmission of COVID-19 may occur during the process of intubation, nebulized treatment, ventilator, tracheostomy, and other treatments as per World Health Organization (WHO), 2020. Airway management must be adequately done by the physicians. Physicians with vast experience must be deployed on duty to avoid further viral transmission (Sullivan, Gibson, Berra, Chang, & Bittner, 2020).

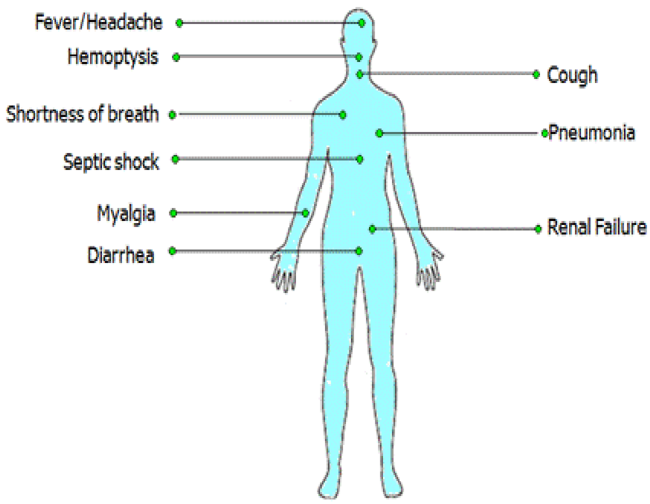

Figure 2 Clinical representation of humans with COVID-19

1.2 Strategies adopted for respiratory failure

The process of insertion of a tube into the body, utilization of ventilator support deprived of a non-natural airway, and other treatments are followed to treat respiratory failure. A proficient operator with all shielding equipment should undergo the rapid sequence intubation process. Lower tidal volumes of 4 to 6 ml/kg of body weight and lower inspiratory pressure of plateau pressure (Pplat) < 28 to 30 cm H2O are maintained during the mechanical ventilation. Generally, the NIV method is used for non-severe respiratory failure, and if the situation worsens, PMV can be recommended (Hui et al., 2019). Clinical studies reported that Remdesivir (GS5734) had performed its actions for the RNA viruses, and illustrations are shown in the research of Amirian, and Levy (2020). The paper's organisation has been described by initiating with the objectives in section 2, methodology in section 3, results and discussion in section 4 and the conclusion mentioned in the last section.

2. Objectives

- To predict COVID-19 by analyzing the data state-wise in India
- To analyze the data pool and determine the measures to handle COVID-19.
- To utilize Artificial Intelligence (AI) to resolve the challenges in handling COVID-19.

3. Methodology

3.1 Integration of digital technology and COVID

The various digital technologies such as Artificial Intelligence (AI) based deep learning, big data analytics, blockchain technology, and the Internet of Things (IoT) play a substantial role in the medical field by providing solutions to different clinical diseases and problems. For example, the proliferation of IoT devices in clinics and hospitals facilitates interconnectivity in the digital ecosystem. Furthermore, it enables real-time data gathering at a large scale, exploiting deep learning techniques to model risks and predict outcomes by understanding healthcare trends. Similarly, using blockchain technology, distributed systems in various organizations can be structured as a Peer-to-Peer (P2P) network, which ensures that data are traceable and more secure.

However, the world is presently fronting a health crisis because of a novel coronavirus named COVID-19, a respiratory disease. The severity of COVID-19 is higher than SARS which happened in 2003, resulting in development due to the relative significance of world trade with China. Many countries have followed the conventional infection control methodologies and health measures that have been utilized for SARS since 2003. However, the quarantine measures followed in 2003 may not be effective enough to handle the large-scale COVID-19 in 2020. Hence, digital technologies such as IoT, big data, blockchain, and AI leverage and augment conventional strategies to tackle COVID-19 in the following potential applications:

i. Monitoring, detecting and preventing COVID-19

ii. Mitigating the impact of COVID-19

3.1.1 Monitoring, detecting and preventing COVID-19

The incorporation of IoT with various technologies offers a platform that facilitates public health organizations to acquire the data that is required for monitoring the COVID-19 pandemic. As per World Health Organization (WHO), 2020, for instance, the "Worldometer" updates the real-time data of COVID-19 on the actual number of confirmed, recovered, and deceased cases worldwide. Moreover, a tracking map has also been introduced by Johns Hopkins University's Centre for Systems Science and Engineering to follow the cases of COVID-19 in real-time all over the world by exploiting the data gathered from various sources (Dong, Du, & Gardner, 2019).

Big data provides various opportunities to perform studies on viral activities and supports healthcare policymakers in different countries to be prepared for the COVID outbreak. The three databases, such as Wuhan Municipal Transportation Management Bureau, the location-based services, and the Official Aviation Guide, are exploited to conduct a study for forecasting COVID-19, which could be utilized by healthcare authorities to plan and control the pandemic worldwide (Wu, Leung, & Leung, 2020). Health regulations defined by WHO, the Self-Assessment tool to generate the annual report state-wise, vulnerability index calculation for infectious diseases, are used to assess the vulnerability and to raise awareness for better preparedness to tackle this viral outbreak in Africa. Singapore's government has joined with WhatsApp to disseminate information accurately about COVID-19 to the public (Radanliev, De Roure, Ani, & Carvalho, 2021). Social media platforms such as Twitter and Facebook are presently utilized to clarify uncertainties and provide real-time updates to the public. Thermal imaging-based facial recognition is adopted to detect people with high temperatures at different screening levels in China.

AI can enrich the diagnosis process of COVID-19 by performing the initial screening for the suspected cases and the confirmed cases; however, providing accurate information with less cost is highly challenging. The datasets of COVID-19-positive patients are more significant in many countries that are more suitable for applying AI and deep learning algorithms (Shuja, Alanazi, Alasmary, & Alashaikh, 2021). Such algorithms can be exploited as a screening tool for the suspected cases by focusing on their travel history and contact for the confirmed cases. Henceforth, patients whose risk factor is high could have proceeded for the confirmatory tests. For mild cases, the common intensive steps are applied by the physicians, such as isolation, Monitoring, and providing treatment for patients.

Moreover, AI algorithms could assist physicians in categorizing patients into three cases: 75% mild, 20% moderate, and 5% severe that, including a high-risk factor of mortality cases. It also supports discovering drugs to combat the present pandemic situation. The authors have analyzed several techniques and protocols to process the ethical features in digital supply chain settings.

3.1.2 Mitigating the consequences of COVID-19

Although the influence of COVID-19 nowadays is profound, managing critical core clinical services are significant. The initial step is to reduce other clinical services and postpone medical appointments to fight against this disease outbreak. However, these strategies will not support for a long time if the situation extends for more than six months. Digital technologies could be preferably utilized in healthcare systems to obtain standard medical care and reduce the physical crowd on hospital premises. For instance, a 'virtual clinic' could be established for telemedicine consultations using imaging data (X-ray). To avoid physical meetings, e-learning platforms can be preferably used in a virtual model. AI-based systems reduce the clinical load potentially for all physicians. For example, the 'chatbot' assists patients in identifying early symptoms; people are alerted to hand hygiene's significance and are informed about the appropriate medical treatment if the symptoms worsen. Also, a patient's data can be recorded to identify the signs that may avoid unnecessary hospital visits. Such kind of data could be used in AI algorithms to detect COVID-19. The role of digital technologies in COVID-19 is summarized in Table 2.

Health centres in China are integrated with blockchain industries and pharmacies to dispatch medications to the patient's doorstep. Moreover, timely delivery could also be ensured with precise tracking. People worldwide trust public-health preventive measures to tackle the COVID-19 outbreak. The extensive utilization of digital technologies could enhance and augment public health-related strategies. This pandemic affords an excessive opportunity for successfully using these digital technologies. Furthermore, it increases the government and general acceptance in the healthcare domain to reduce the impact of such chronic diseases (Ting, Carin, Dzau, & Wong, 2020).

Machine Learning (ML) uses the handling of vast amounts of data to predict the spread of the disease (Tuli et al., 2020). When data processing needs high computation, it is necessary to use cloud computing, and it is used in the research of Gill et al., (2019). The authors, Tuli, Tuli, Tuli, and Gill (2020), suggested a model based on the principle of the Fogbus framework (Tuli, Mahmud, Tuli, & Buyya, 2019). Prediction of the cases, the increase and drop in the number of cases, and the expected dates for the countries to enter into the pandemic. An ML solution based on the Multi-Layer Perceptron (MLP), Artificial Neural Network (ANN), has been recommended by authors (Car et al., 2020) and is deployed for forecasting the various metrics like the death rate based on the location, recovery rate for each site and the count of persons those who are infected in that area. In (Salgotra, Gandomi, & Gandomi, 2020), a new genetic programming-based model (GP) used for the time series forecast of COVID-19 in India has been devised. Certain functions, such as prediction, are needed in all fields to ensure efficient resource utilization and economic improvement (Hu et al., 2018; Singh, & Mohanty, 2017). The study proposed by Ahmed, Al-Hamadani, and Satam (2022) examined the effects of chronic illnesses on the progression of viral symptoms in humans using a dataset of patients from Mexico. The techniques include random forest, decision trees, linear regression, binary search, and k-nearest neighbour. Additionally, the study shows that the virus has the potential to bring about death, particularly in elderly individuals inexorably.

Table 2 Role of digital technologies for COVID-19

	IoT	**Big data**	**Blockchain**	**AI**
Public health strategies	Real-time Monitoring	Modelling disease spread and its potential growth	Manufacturing and supply of vaccines	Prediction of disease based on images and clinical data
	Live tracking	Modelling the situation of vulnerability to fight against disease pandemic	Insurance claims for COVID treatment	Detection of the viral disease using chest images (X-ray)
	Virtual clinics	Modelling the requirement of utilities for clinics and operating theatres	Distribution of medicines at the patient's doorstep	Facilitates "Chat Bots" to handle public inquiries

	Disseminating information via WhatsApp	Modelling pharmaceutical supply	Aggregate data from several sources to support health officials	Automatic diagnosis of COVID-19

3.2 AI to battle against coronavirus

AI is one of the avenues for better understanding viruses to take preventive measures such as

i. Structural biology determines the virus structure that assists in preparing vaccines.

ii. Mathematical models can be exploited for a better understanding of virus transmission.

iii. Computational biology helps in understanding virus evolution.

iv. Dock studies assist in screening inhibitors and drugs.

AI plays a decisive role in preventing and finding a treatment for COVID-19. It promotes the different stages of healthcare, such as rapid analysis tests, syndromic surveillance, and progress in drug development. AI system named "Coronavirus Chest CT Smart Evaluation System" is being deployed to diagnose the infected and suspected cases within a few seconds. AI can also support prioritizing the patients for treatment (Dong et al., 2020; Ishack, & Lipner, 2020; Simonite, 2020; Bai et al., 2020). Based on the mortality rate, it is revealed that older adults are at higher risk than young (healthy) people. Machine learning techniques can be utilized to quickly learn the trained Dataset in terms of predicting higher-risk mortality cases. AI can be incorporated with genetic, environmental, and biological data that ultimately supports finding a cure for COVID-19 (Ai et al., 2020; Shi et al., 2020; Esteva et al. 2017; Shen, Wu, & Suk, 2017).

In this emergency, AI is considered a promising technique that expedites the discovery of drugs for COVID-19. With AI techniques, the current status of drug availability in the market can be identified quickly from the larger datasets that help discover a new drug quickly. Different approaches are followed for developing a vaccine. Scientist usually grows viruses in the cell culture and use them as a vaccine by chemically inactivating them. However, this treatment might induce incorrect antibody responses, which are more harmful than usual (Esteva et al., 2019; Topol, 2019; Ardila et al., 2019). Understanding the objectives of these antibodies is significant to obtain the correct immune responses. It could be accomplished by knowing about the viral protein structure, where the virus will bind and infect cells. The features of the antigenic sites can be determined with the help of solved structures that assist in designing a protein subunit, and later it could be mounted on a platform to be exploited as a vaccine. The study's results (Shah, Mulahuwaish, Ghafoor, & Maghdid, 2021) show that several factors, including social interactions tracked by social networks, will aid in the spread of the virus.

• To battle COVID-19, several technologies are being employed, and they have identified critical applications of AI.

• To stop the COVID-19 virus from spreading, they have identified research problems, directions for prediction techniques, and AI-enabled technology.

Australian Census-based Epidemic Model (ACEMod), which was earlier used for treating the influenza pandemic, has been reformed to treat COVID-19, and it depends on the critical disease transmission parameters. This model has been tuned on various hyperparameters such as prohibitions in travelling, social distancing, banning educational institutions, and isolation of cases (Hu et al., 2020). Later, it is evaluated. From the experimental results, it is decided that the grouping of various approaches is required to alleviate and subdue the COVID-19 outburst. International arrival limitations, isolation of cases, and social distancing for 13 weeks must be combined to produce an amenable level of 80% or above. Maghdid et al., (2021) concentrated on recommending AI technologies that radiologists or other medical practitioners can utilize to quickly and accurately diagnose COVID-19 situations. However, creating such AI tools is difficult because there isn't a publicly accessible X-ray and CT scans Dataset. To achieve this, this study attempts to create an extensive dataset of X-rays and CT scan pictures from various sources and offers a straightforward yet efficient COVID-19 detection method combining deep learning and transfer learning methods. Smartphone sensors were used to detect COVID-19 through the novel framework suggested by Maghded et al., (2020). The research done by Shabbir et al., (2022) compares various machine learning techniques for case severity detection and presents explorative data analysis of symptoms, comorbidities, and other characteristics.

The proposed design (Namburu et al., 2022) emphasizes FPGA deep learning frameworks to detect the corona from X-Ray images. The authors (Kumar Raju, Sumathi, & Chandra, 2021) have used Time Series Analysis and Regression Models to examine the Corona dataset, which has more than 3000 X 5 dimensions.

Additionally, they could forecast the future trend and then suggest a dataset design to understand the pandemic's unresolved issues better. The term "data analytics" is used to describe a variety of different forms of analysis. The integration could be furthered for improved use of electronic medical records (EMR) in healthcare as specified in (Doraikannan, & Selvaraj, 2020) since consumers need more data for a full investigation, which necessitates perception. Particularly for data aggregation, data capture, real-time data streaming, analytics, and other visualisation solutions, some strategies for managing the data and analysis require additional work and ongoing attention. The technologically advanced artificial intelligence (AI) enabled framework interprets the signal data from the smartphone sensors to forecast the grade of pneumonia severity and the outcome of the illness. Several areas in which AI could be deployed to combat the current situation are

- Tracing and forecasting
- Data dashboards
- Treatments
- Raising alert and initial warning
- Analysis and projection
- Social Monitoring

Various implementations have been done, and multiple researchers have done a detailed investigation of COVID-19 cases, as shown in Figure 3. On doing exploratory data analysis, it is observed that the outliers are present in the Dataset, as shown in Figure 4.

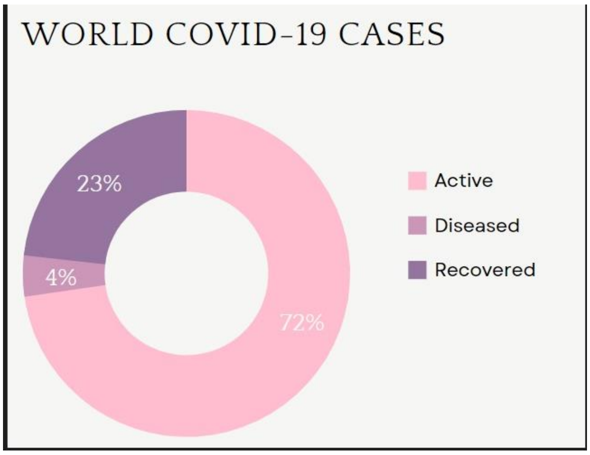

Figure 3 Covid-19 cases worldwide

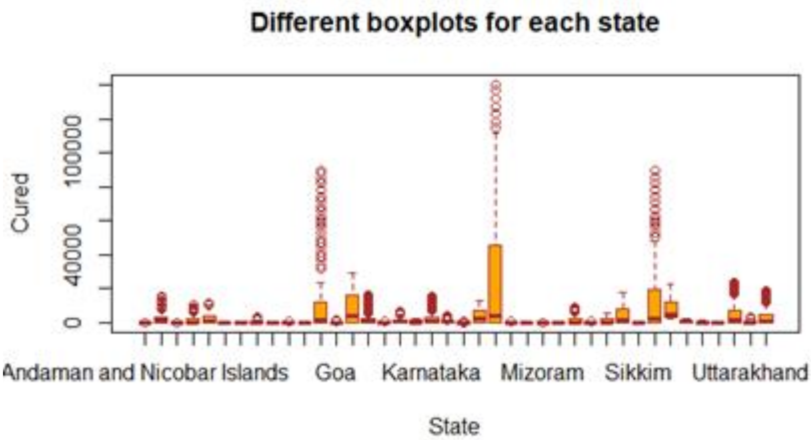

Figure 4 State-wise analysis of cured persons

The remarkable anticipative feature brings awareness to various countries that the world bank has. This is used to determine the facilities regarding healthcare so that abundant precautionary measures must be taken during this pandemic.

3.2.1 Data visualization techniques

The dataset Covid-19 has been taken, and visualization techniques have been applied to determine the correlation and relation between the dependent and independent variables. The Dataset comprises certain information such as the date and Total confirmed cases state-wise for each date, along with the count of deaths and cured. Figure 5 shows the percentage of COVID patients in each state, which is represented in the pie chart.

A detailed analysis of all metrics is required for the action to be taken during this pandemic situation. State-wise distribution of total confirmed cases in our country is given in Figure 6. From this, it has been observed that the state of Maharashtra has the highest spike in cases when compared with the other states. The following form to get mostly affected is Gujarat.

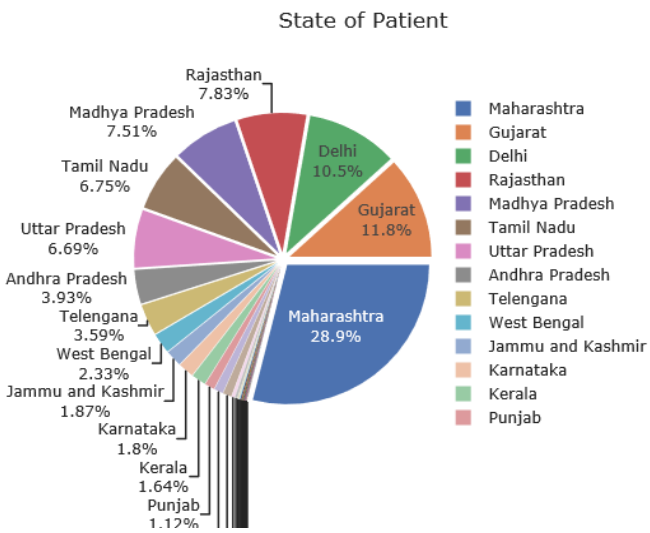

Figure 5 Contribution of states toward COVID-19

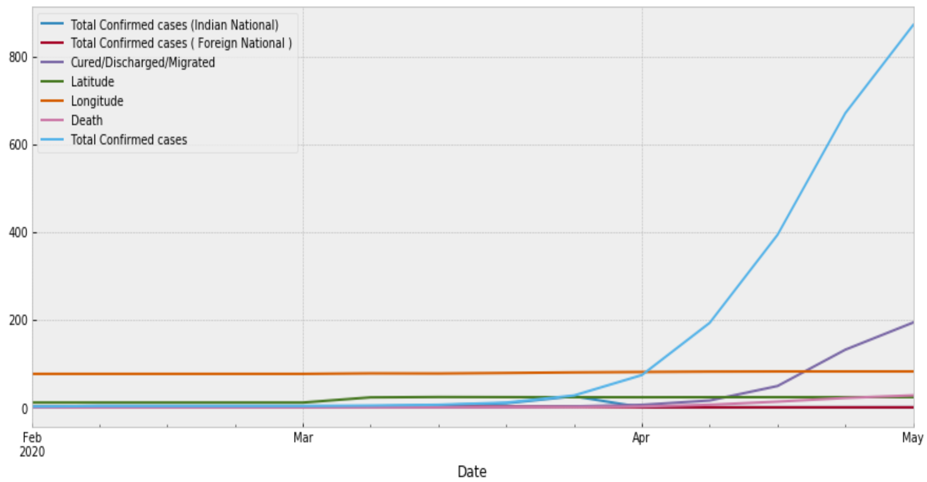

Figure 6 Analysis of metrics and data

Total confirmed cases must be measured every day, which could be plotted based on the month. Exponential growth in the Total confirmed cases of Indian nationals is seen in Figure 6. When comparing the cases of Foreign National, the confirmed cases of Indian National are more. To proceed further, a model has to be deployed for a detailed analysis.

3.2.2 Model deployment

The Dataset of covid-19 is downloaded from Kaggle. Various features have been taken for a detailed analysis, and a few models have been deployed to analyze for further process. The Dataset comprises multiple attributes such as date, time, and state/union territory, along with other information about the count of people who got affected. Data pre-processing is done so that the significant and insignificant attributes can be determined. Further, the Dataset is split into training and test datasets. To proceed further, the authors have done an extensive analysis by considering all states. To initiate the model construction, pre-processing has been done on the Dataset, and we have deployed a few algorithms to develop the model. The architecture of the model is depicted in Figure 7. The Dataset is partitioned into training and test dataset. Then, the models are trained and tested under various models, as shown in Figure 7.

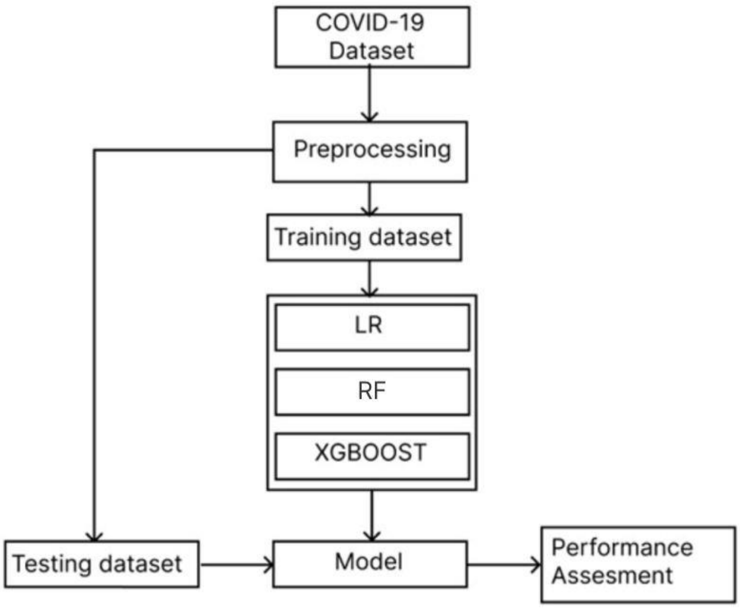

Figure 7 The Architecture of the proposed model

1) Linear regression

Conventionally, regression models are considered supervised learning models that are used for predicting a value that could be either quantitative or discrete type. Hence, as the first step, linear regression is deployed. In this linear regression model, the idea is to construct a hypothesis for determining the estimation of the dependent variable based on the self-determining variable(s).

h"¬0" (x)= θ

Where θ is the intercept and θ1 denotes the total number of cured, θ2 refers to the total number of deaths, and θ3 refers to the state. More focus should be on the cost function during the model construction. Certain metrics used to evaluate the regression model are Mean Squared Error(MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

2) Random forest

Due to the particular characteristic of the maximum voting strategy from each tree, the final value is computed. Then, bootstrapping of the attributes and instances is done haphazardly. The algorithm for the random forest is given below.

1. Initiate the choice of the random samples (α) from the dataset.

2. A decision tree is constructed for each sample. The result will be obtained from the decision trees.

3. A voting scheme is introduced to assess the performance of the predicted result.

4. The one with the highest votes is chosen as the final result.

When applying random forest to all states and considering the features such as the count of cured, death and date, prediction is made based on the number of confirmed cases. To make a prediction, necessary pre-processing techniques like one-hot encoding and handling of missing values have been done on the Dataset.

3) XGBoost regression

New models are created through gradient boosting. These models are used to compute the error that occurs in the earlier model, and later the remaining is included to bring out the final prediction. This regressor is known for its speed and flexibility. The boosting ensemble technique works in three steps:

The model developed initially does the prediction process and will be related to the residual computation.

For example, assuming the target variable is 'γ', the relationship could be represented as (γ-M0). Next, to fit the residuals, a new model is developed, and it is denoted as α1. Next, the initial and new models are joined to achieve the boosted version of the initial model. In that case, the mean squared error obtained from the new model will be less than the old one.

$$M1\ (V) < M0\ (V) + \alpha 1\ (V)$$

The significant feature is that the information could be used to reduce errors. The particular characteristic of XGBoost is that training on it is done with the help of an objective function by performing the minimization of loss.

4. Results and discussion

Figure 8 shows the performance of the models based on the coefficient of determination, also known as R-squared. From the results, it has been identified that XGBoost outperformed the other regression algorithms. Furthermore, experimental results show that the prediction of covid-19. And it has been observed that the projection through XGBoost shows more accurate results.

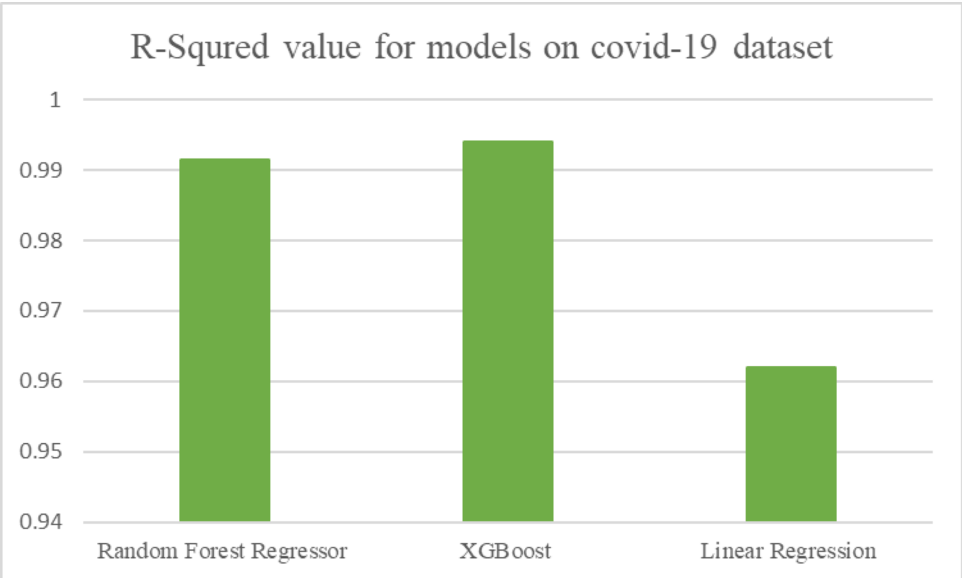

Figure 8 R-Squared comparison for various models on the covid-19 dataset

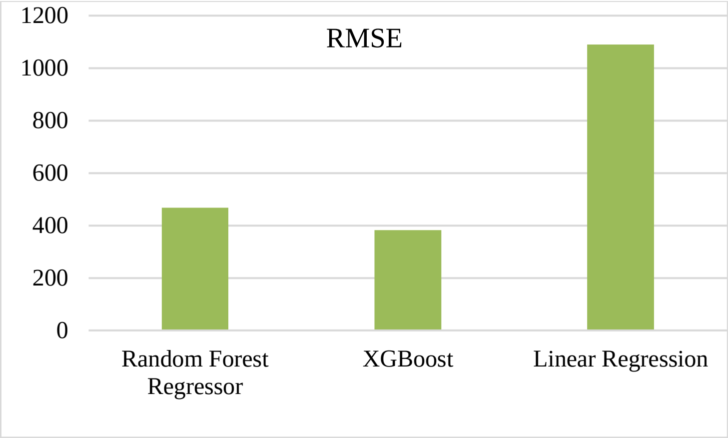

Figure 9 RMSE for various models

RMSE predicts the model's accuracy since it acts as a good metric. However, the significant feature is checking for the fitness of the model. From the results shown in Figure 9, it is well understood that the XGBoost model's RMSE value is less, and it is found to be the best-fit model.

Various hyperparameters used for training the models are described as given below:

- max_Depth: it is used to denote the maximum depth of the tree
- n_Estimators: the number of trees used for boosting
- learning_rate: A factor used to rate the outcome from the tree, whether it could be expanded or contracted.
- Loss function: The loss function is vital in tuning the hyperparameters. Various loss functions have been deployed to check the model and how far it predicts the anticipated outcome of inevitable loss functions like linear square, slightest absolute deviation, Huber and quantile functions. For example, the loss function could be computed in equation 1.

$$\sum_{i=1}^{n} L(y_i, p_i) + \frac{1}{2}\lambda O_v^2 b \quad (1)$$

Where O_v = output value

Figure 10 represents the loss functions on the x-axis, and R-Squared values are plotted. From this, when varying the loss functions, the value of R-Squared differs and is assessed for each model. From the results, it is understood that the Huber loss function outperforms the other loss functions. The loss function for Huber could be represented as:

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & for|y - f(x)| \leq \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2 & otherwise \end{cases} \quad (2)$$

Huber loss is a combination of MSE and MAE. In the above equation, the delta is denoted as the hyperparameter. It is used to determine the

range for both MAE and MSE. It is considered an iterative function to ensure the delta value is appropriate.

MAE: It is used to determine the magnitude of errors on an average from a group of predictions. During this process, directions are not taken into account. The equation and loss functions for MAE are given below in eqn no 3 and Figure 11.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - y_i^p|}{n} \quad (3)$$

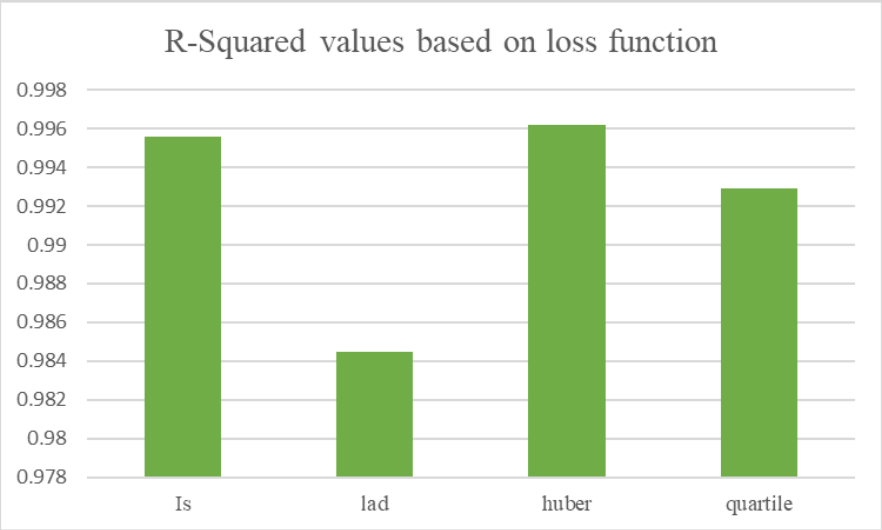

Figure 10 R-Squared values based on the loss functions

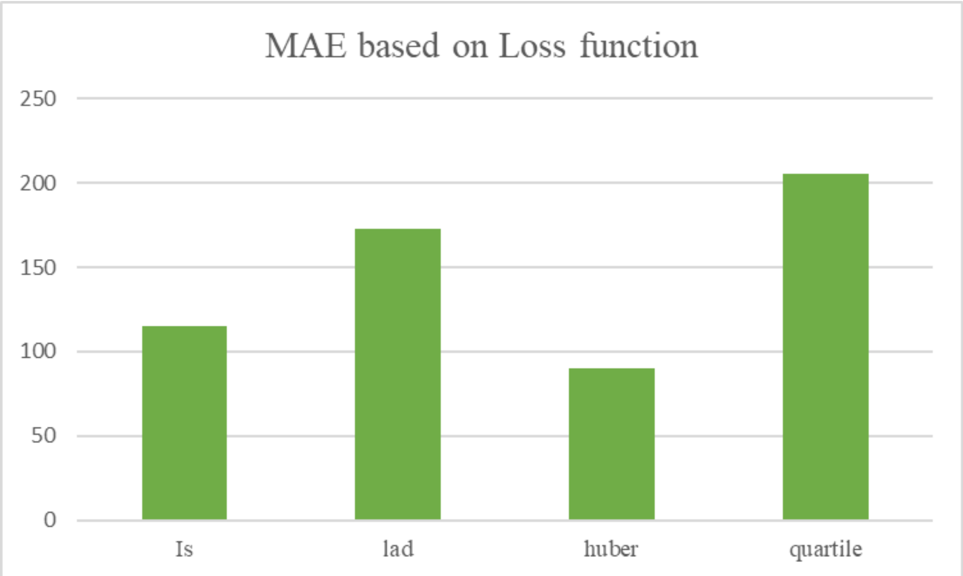

Figure 11 MAE based on loss functions

RMSE is assessed by varying the loss functions to check and validate the model's performance shown in Figure 12. It has been observed from the results that the RMSE computed based on the Huber loss function is found to be low, and it has been proven that the model outperforms based on the Huber loss function.

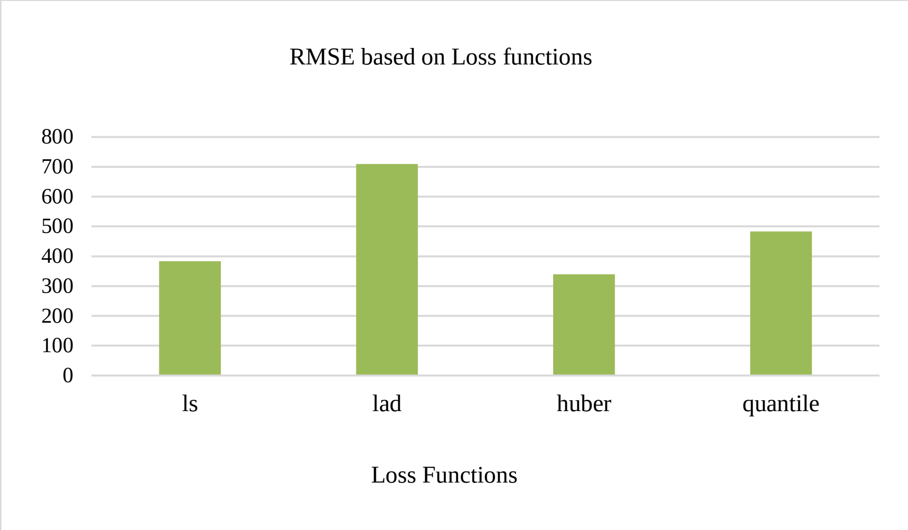

Figure 12 RMSE based on loss functions

5. Conclusion

The deployment of various techniques of AI with several successful stories is managing our daily activities. Currently, AI has a footprint in dealing with COVID-19 in a diversified manner. It has been observed from various inferences that there are applications based on AI techniques in analyzing chest radiology images, time-series data, mining, and Natural Language Processing. With the help of AI techniques, several measures could be taken to eradicate this pandemic in various ways. Also, it helps to create awareness among people. This work has discussed different AI techniques that have been implemented to combat the virus. Extensive data analysis is done to illustrate the association of various attributes explicitly. A detailed analysis of infected cases state-wise has been presented. Multiple models have been deployed and assessed for their performance. Among the models deployed, XGBoost has shown a remarkable improvement in prediction. It has been tuned based on the loss functions. The R-Squared value obtained through the Huber loss function through the XGBoost model is 99.65%. This model outperforms the other models. Apart from that, the fatality rate has been discussed. Other aspects such as social distancing, increase in population and adapting to the other healthcare protocols imposed might impact the actual cases of patients that are encountered. Further, the authors plan to analyze the time series data and accelerate the prediction process through other deep learning models.

Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of Interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Availability of data and material

Not applicable

Code availability

Not applicable

Author contributions

The corresponding author claims the major contribution of the paper including formulation, analysis and editing. The co-authors provide guidance to verify the analysis result and manuscript editing.

Compliance with ethical standards

This article is a completely original work of its authors; it has not been published before and will not be sent to other publications until the journal's editorial board decides not to accept it for publication

6. References

Ahmed, A. H., Al-Hamadani, M. N., & Satam, I. A. (2022). Prediction of COVID-19 disease severity using machine learning techniques. *Bulletin of Electrical*

Engineering and Informatics, *11*(2), 1069-1074. https://doi.org/10.11591/eei.v11i2.3272

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., ... & Xia, L. (2020). Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*, *296*(2), E32-E40. https://doi.org/10.1148/radiol.2020200642

Amirian, E. S., & Levy, J. K. (2020). Current knowledge about the antivirals remdesivir (GS-5734) and GS-441524 as therapeutic options for coronaviruses. *One health*, *9*, 100128.https://doi.org/10.1016/j.onehlt.2020.100128

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., ... & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, *25*(6), 954-961. https://doi.org/10.1038/s41591-019-0447-x

Bai, H. X., Hsieh, B., Xiong, Z., Halsey, K., Choi, J. W., Tran, T. M. L., ... & Liao, W. H. (2020). Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. *Radiology*, *296*(2), E46-E54. https://doi.org/10.1148/radiol.2020200823

Car, Z., Baressi Šegota, S., Anđelić, N., Lorencin, I., & Mrzljak, V. (2020). Modeling the spread of COVID-19 infection using a multilayer perceptron. *Computational and mathematical methods in medicine*, *2020*. https://doi.org/10.1155/2020/5714714

Chan, P. K., & Chan, M. C. (2013). Tracing the SARS-coronavirus. *Journal of thoracic disease*, *5*(Suppl 2), S118. https://doi.org/10.3978/j.issn.2072-1439.2013.06.19

Doraikannan, S., & Selvaraj, P. (2020). Python for healthcare analytics made simple. *Blockchain and Machine Learning for e-Healthcare Systems*, 367. https://doi.org/10.1049/PBHE029E_ch15

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*(5), 533–534. https://doi.org/10.1016/s1473-3099(20)30120-1

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Correction: Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *546*(7660), 686-686. https://doi.org/10.1038/nature22985

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature medicine*, *25*(1), 24-29. https://doi.org/10.1038/s41591-018-0316-z

Gill, S. S., Tuli, S., Xu, M., Singh, I., Singh, K. V., Lindsay, D., ... & Garraghan, P. (2019). Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, *8*, 100118. https://doi.org/10.1016/j.iot.2019.100118

Hu, S., Liu, M., Fong, S., Song, W., Dey, N., & Wong, R. (2018). Forecasting China future MNP by deep learning. In *Behavior engineering and applications* (pp. 169-210). Springer, Cham. https://doi.org/10.1007/978-3-319-76430-6_8

Hu, Z., Ge, Q., Li, S., Jin, L., & Xiong, M. (2020). Artificial intelligence forecasting of covid-19 in china. *arXiv preprint arXiv:2002.07112*. https://doi.org/10.18562/IJEE.054

Hui, D. S., Chow, B. K., Lo, T., Tsang, O. T., Ko, F. W., Ng, S. S., ... & Chan, M. T. (2019). Exhaled air dispersion during high-flow nasal cannula therapy versus CPAP via different masks. *European Respiratory Journal*, *53*(4), 1802339. https://doi.org/10.1183/13993003.02339-2018

Ishack, S., & Lipner, S. R. (2020). Applications of 3D Printing Technology to Address COVID-19-Related Supply Shortages. *The American journal of medicine, 133*(7), 771–773. https://doi.org/10.1016/j.amjmed.2020.04.002

Kumar Raju, B. K. S. P., Sumathi, D., & Chandra, B. (2021). A Machine Learning Approach to Analyze COVID 2019. In *Proceedings of International Conference on Computational Intelligence and Data Engineering* (pp. 237-248). Springer, Singapore. https://doi.org/10.1007/978-981-15-8767-2_21

Mackenzie, J. S., & Smith, D. W. (2020). COVID-19: a novel zoonotic disease caused by a coronavirus from China: what we know and what we don't. *Microbiology Australia*, *41*(1), 45-50. https://doi.org/10.1071/MA20013

Maghded, H. S., Ghafoor, K. Z., Sadiq, A. S., Curran, K., Rawat, D. B., & Rabie, K. (2020, August). A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: design study. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 180-187). IEEE. https://doi.org/10.1109/IRI49571.2020.00033

Maghdid, H. S., Asaad, A. T., Ghafoor, K. Z., Sadiq, A. S., Mirjalili, S., & Khan, M. K. (2021, April). Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. In *Multimodal image exploitation and learning 2021* (Vol. 11734, pp. 99-110). SPIE. https://doi.org/10.1117/12.2588672

Namburu, A., Sumathi, D., Raut, R., Jhaveri, R. H., Dhanaraj, R. K., Subbulakshmi, N., & Balusamy, B. (2022). FPGA-based deep learning models for analysing corona using chest X-ray images. *Mobile Information Systems*, *2022*. Article ID 2110785, 1-14. https://doi.org/10.1155/2022/2110785.

Peiris, J. S., Yuen, K. Y., Osterhaus, A. D., & Stöhr, K. (2003). The severe acute respiratory syndrome. *New England Journal of Medicine*, *349*(25), 2431-2441. https://doi.org/10.1056/NEJMra032498

Perlman, S., & Vijay, R. (2016). Middle East respiratory syndrome vaccines. *International Journal of Infectious Diseases*, *47*, 23-28. https://doi.org/10.1016/j.ijid.2016.04.008

Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020). COVID-19, SARS and MERS: are they closely related?. *Clinical microbiology and infection*, *26*(6), 729-734. https://doi.org/10.1016/j.cmi.2020.03.026

Radanliev, P., De Roure, D., Ani, U., & Carvalho, G. (2021). The ethics of shared Covid-19 risks: an epistemological framework for ethical health technology assessment of risk in vaccine supply chain infrastructures. *Health and Technology*, *11*(5), 1083-1091. https://doi.org/10.1007/s12553-021-00565-3

Salgotra, R., Gandomi, M., & Gandomi, A. H. (2020). Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos, Solitons & Fractals*, *138*, 109945. https://doi.org/10.1016/j.chaos.2020.109945

Shabbir, A., Shabbir, M., Javed, A. R., Rizwan, M., Iwendi, C., & Chakraborty, C. (2022). Exploratory data analysis, classification, comparative analysis, case severity detection, and internet of things in COVID-19 telemonitoring for smart hospitals. *Journal of Experimental & Theoretical Artificial Intelligence*, 1-28. https://doi.org/10.1080/0952813X.2021.1960634

Shah, S., Mulahuwaish, A., Ghafoor, K. Z., & Maghdid, H. S. (2021). Prediction of global spread of COVID-19 pandemic: a review and research challenges. *Artificial Intelligence Review*, 1-22. https://doi.org/10.36227/techrxiv.12824378.v1

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, *19*, 221. https://doi.org/10.1146/annurev-bioeng-071516-044442

Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O., Gu, J., ... & Zheng, C. (2020). Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *The Lancet infectious diseases*, *20*(4), 425-434. https://doi.org/10.1016/S1473-3099(20)30086-4

Shuja, J., Alanazi, E., Alasmary, W., & Alashaikh, A. (2021). COVID-19 open source data sets: a comprehensive survey. *Applied Intelligence*, *51*(3), 1296-1325.

https://doi.org/10.1101/2020.05.19.20107532.

Simonite, T. (2020). Chinese hospitals deploy ai to help diagnose covid-19. *Wired*. Retrived form https://www.wired.com/story/chinese-hospitals-deploy-ai-help-diagnose-covid-19/

Singh, N., & Mohanty, S. R. (2017). Short term price forecasting using adaptive generalized neuron model. In *Advances in Computer and Computational Sciences* (pp. 419-428). Springer, Singapore. https://doi.org/10.1007/978-981-10-3770-2_39

Sullivan, E. H., Gibson, L. E., Berra, L., Chang, M. G., & Bittner, E. A. (2020). In-hospital airway management of COVID-19 patients. *Critical Care*, *24*(1), 1-8. https://doi.org/10.1186/s13054-020-03018-x

Ting, D. S. W., Carin, L., Dzau, V., & Wong, T. Y. (2020). Digital technology and COVID-19. *Nature medicine*, *26*(4), 459-461. https://doi.org/10.1038/s41591-020-0824-5

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, *25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Tuli, S., Basumatary, N., Gill, S. S., Kahani, M., Arya, R. C., Wander, G. S., & Buyya, R. (2020). HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments. *Future Generation Computer Systems*, *104*, 187-200. https://doi.org/10.1016/j.future.2019.10.043

Tuli, S., Mahmud, R., Tuli, S., & Buyya, R. (2019). Fogbus: A blockchain-based lightweight framework for edge and fog computing. *Journal of Systems and Software*, *154*, 22-36. https://doi.org/10.1016/j.jss.2019.04.050

Tuli, S., Tuli, S., Tuli, R., & Gill, S. S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*, *11*, 100222. https://doi.org/10.1016/j.iot.2020.100222

World Health Organization (WHO). (2020). Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations. https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations

Worldometer. (2023). "Worldometer" updates the real-time data of COVID-19 on the actual number of confirmed, recovered, and deceased cases orldwide. https://www.worldometers.info/

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, *395*(10225), 689-697. https://doi.org/10.1016/S0140-6736(20)30260-9