A PROPOSAL ON

PREDICTING INCIDENCE OF COVID-19 INFECTIONS IN NIGERIA USING POLYNOMIAL REGRESSION MACHINE LEARNING TOOL.

**INTRODUCTION**

Commonly known as Corona virus (SARS-COV-2), is a very infectious disease which has spread all over the globe. This virus belongs to the Coronavirdae family. Cough, fever, shortness of breath, headache, muscle ache, loss of smell and taste can all be symptoms of COVID-19 (Razai, 2020). The origin of this virus is still not known, but studies have concluded that it is linked to Beta-coV genera, a virus which infects rodents and bats (Cascella ., 2020). Wuhan city, China reported the first case in December 2019. 182,276,267 cases have been confirmed and 3,947,630 deaths have been reported as of 29 June 2021. Various health industries are looking at new techniques and methods to tackle this pandemic. Data Science is currently one of the hottest trends in the modern era. It covers artificial intelligence, machine learning, polynomial regression, deep learning, algorithms, modeling, statistics and simulation. It has been argued that Artificial Intelligence (AI) will play a crucial role to help fight against this highly infectious virus. It also provides great support to academic and clinical studies (Browning, 2020). ML has a lot of scope in a variety of fields like engineering, interdisciplinary science, psychology, social analysis, earth observations, hazard mitigations, urbanized locations etc. Various implementations of Machine Learning (ML) ideas were used in applications like tracking people using facial recognition, disinfecting areas using drones (Estrada, 2020), medicines and food delivery by automated robots, detection of COVID-19, drug discovery etc. AI’s subset is ML.

Referring to available literatures, the applicability of ML and hybridized models successfully in various field of engineering and natural disasters (Abbaszadeh Shahri, Asheghi ., 2021; Abbaszadeh Shahri, Kheiri ., 2021; Abbaszadeh Shahri & Maghsoudi Moud, 2020, 2021), psychometric analysis (Orrù, 2020; Rosenbusch .2021), medical and pharmaceutics (Kan, 2017; Réda ., 2020; Vamathevan ., 2019), incorporating with graph theory (Abderrahim ., 2014) as well as social sciences (N. C. Chen ., 2018; Grimmer .2021; Hindman, 2015) have been approved. ML uses statistical models (Gale, 1988) with a minimum amount of knowledge to solve complicated problems. The algorithms include Linear Regression, Logistic Regression (LR), Decision tree (DT), Random Forest (RF), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), K-means clustering, Naïve Bayes (NB) model, etc. Deep Learning (DL) is a subcategory of ML and it concentrates on building Neural Network (NN) models which use feed-forward and back propagation to learn various data trends. They are playing an important role in battling COVID-19 as well. They aid in the diagnosis of the viral disease and also predict the severity of the same. In the pharmaceutical industry, these models are used to observe the genetics and mutations of COVID-19 to enhance drug prediction and vaccination. AI learns various disease transmission models which are used in predicting outbreaks, public monitoring, epidemic detection and patient tracking. We first analyze a few papers which have surveyed and reviewed machine learning methods to tackle COVID-19

**STATEMENT OF PROBLEM**

By developing a system that uses regression polynomial, a supervised machine learning tool to predict the number of incidence of COVID-19 infections and deaths, this study solves the following problems:

In the past, diagnosis has been plague by several factors based on the current manual process. Some of these issues include:

1. Wrong/ uninformed decission due to improper prediction as a result of wrong algorithms
2. improper or inadequate allocation of resources to isolation location and centers
3. Drugs prescription is hardly ever done with the right age of the patient in mind.

**AIM AND OBJECTIVES OF THE STUDY**

The aim of the study is to develop a system, which will provide informed decisionand aslo capable of predicting the number of incidence of COVID-19 infections and deaths in any part of world/Nigeria. This will help various health authorities and individuals to know how to adopt the various COVID-19 guidelines, with the aim of curtailing the spread of the virus. Its objectives include the following:

1. Collection of relevant dataset on the number of incidence of COVID-19 deaths and infections
2. Development of regression polynomial machine learning algorithm, which will be used to predict/forecast the number of incidence of COVID-19 deaths and infections,
3. Performance evaluation of the predicted result.

**SIGNIFICANCE OF THE STUDY**

Regional Health Authorities:

The spread of the virus can be effectively be controlled by the various health authorities because they will be well equipped to fight the virus through the predicted number of infections and deaths in any particular region. This means that they will know how and when to enforce the various COVID-19 protocols/guidelines.

**SCOPE OF THE STUDY**

The adopted polynomial regression machine learning tool based algorithm it is designed to predict and interpret incidence of covid-19 infectious

**LIMITATIONS OF THE STUDY**

1. I encountered many difficulties in the pursuit of material and resources that help for the successful completion of this project work.ii. It was also difficult to collect data for mine case study due to unavailability of personnel it gives me adequate information in my area of study.iii. Unavailability of academic materials.
2. Lack of financial support

**RELATED WORK**

A number of researchers have already tried to review the applications of machine learning to tackle the pandemic. Chamola. (Chamola., [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) reviewed the applications of IoT, block chain, drones, AI and 5 G to fight COVID-19. Clinical features, diagnostic procedures, various transmission mechanisms, managing and limiting the damage from this virus using the above technologies were discussed in the paper. Pandey , (Panday, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), surveyed automatic CT scans along with image processing and DL to diagnose COVID-19. She explored 74 papers that included specified preprocessing techniques, classification methods and experimental results. Mohamadou. (Mohamadou., [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), used a mathematical approach along with AI to explore various datasets and also predict the pandemics that might occur in the future. It was also found that most of the AI implementations on X-ray and CT images were based on CNN models. Modern approaches to tackle COVID-19 were discussed in (A. Kumar .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Effective tracking and patient quarantine were effectively done using various applications in this paper. The application of ML to tackle COVID-19 was surveyed in (J. Chen  [2020b](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Disease diagnosis, drug and vaccine development, pathogenesis and virologic studies with the help of ML were looked into. Deep learning techniques to manage the pandemic were discussed in (Sufian ., [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Systematic study of Deep Transfer Learning (DTL) and Edge Computing (EC) along with DL was presented along with a DTL pipeline to mitigate future COVID-19 outbreaks. An overview of speech and language processing to tackle COVID-19 was discussed in (Deshpande & Schuller, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The paper further explains non-speech modalities that may complement further audio and speech analysis. A review regarding the use of ML applications in battling pandemics was discussed in (Lalmuanawma., [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The paper targeted medication, screening, treatment, forecasting, prediction, contact tracing and vaccination procedures for the pandemic. The disadvantages of AI in battling this deadly disease were discussed in (Naudé, [2020a](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The paper discussed about the lack of data available in clinical research and also about data privacy and rigorous AI-human interaction. The involvement of big data along with AI to fight the pandemic was reviewed in (Pham .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The research emphasized the use of the above techniques in preventing the severe effects caused by the pandemic. Chest x-ray images along with AI for detection of COVID-19 were reviewed in (Ilyas ., [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Tsikala Vafea , [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Deep learning architecture models like Inception net, Resnet, Google net etc., were discussed here. Automatic COVID-19 detection is very important to prevent the spread of the virus through contact. A short review regarding the applications of AI for COVID-19 were seen in (Ahuja, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Vaishya , [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Proper screening, predicting future patients along with early diagnosis and detection of this infectious virus were reviewed in the paper. The reduction of workload among health care workers with the help of ML was also discussed. Applications of AI in developed countries to combat COVID-19 were discussed in (Naudé, [2020b](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Unberath  [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The study claimed that AI is a very powerful tool to reduce the economic and health impacts of the pandemic. Social control and drug/vaccine development can be enhanced using the applications of AI. In another study (Nguyen  [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), applications of AI along with blockchain to combat COVID-19 were discussed. Blockchain helped in managing the pandemic by protecting user privacy and enabling early outbreak detection, whereas AI was used to identify various treatments and played a vital role in drug manufacturing according to the study. Image processing along with machine learning to process medical images for COVID-19 was reviewed in (Ulhaq , [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). This survey paper provides an initial review on the available articles on computer vision and COVID-19. Management of COVID-19 using AI was discussed in (Shaikh ., [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Epidemic models were surveyed in (Shinde , [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The paper explained forecasting techniques that was divided into two types, mathematical models with stochastic theory and machine learning techniques. This helped in analyzing various trends which led to the spread of the disease. Sharma .(Sharma & Gupta, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), surveyed ML applications related to COVID-19 diagnosis. The study was based on the Indian population. Future diagnosis with the help of AI to make medical decisions was discussed. Review on the applications of big data and AI to manage the pandemic were discussed in (Barragán & Manero, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The above technologies were used to handle large volumes of data obtained from epidemic outbreak monitoring, public health surveillance, regular situation briefing, trend forecasting from health facilities and government institutions. Forecasting the number of cases using machine learning was explored in (Ahmad .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Various researches about the prediction of the number of confirmed cases using ML was reviewed in this paper. A brief overview of recent studies to tackle the viral disease using ML was performed in (Bullock .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Kannan .(Kannan .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), used ML to find and discover new drugs that could potentially cure COVID-19. In this paper, vaccine development using ML techniques was discussed in depth. Critical patients were managed using AI in (Rahmatizadeh .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). ML was used with other diagnostic factors to predict whether a patient would require an ICU facility or not. Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans with the aid of ML were used in the diagnosis of COVID-19 (Dong .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The paper indicated that imaging characteristics and their changes help in the diagnosis of this viral disease. ML methods were desperately needed to bring out the best of imaging, claimed the study. Chest X-ray (XSR) images along with AI were used to recognize COVID-19 patients in (Goodwin .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Even though there are so many researches, cooperation must always exist between data scientists when it comes to exchanging data, tools and code (Luengo-Oroz .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Peiffer-Smadja .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The main challenges in battling COVID-19 were studied in (Van der Schaar .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) and they also discussed the use of ML to solve those problems independently. Shi .(Shi .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) used machine learning techniques for data acquisition, diagnosis and prognosis of infected patients. Biological data mining was used to perform COVID-19 detection in (Dasgupta .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Another paper (Tayarani N, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), discussed the uses of AI in battling COVID-19. In ([Chamola.,](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), disaster management using ML approach was discussed. Alafif .(Alafif .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), surveyed AI applications in COVID-19 diagnosis. In (Latif .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), a review of DL techniques to combat COVID-19 was discussed in detail. Table 1 summarizes the existing literature that uses ML applications to tackle COVID-19.

Table 1. Existing literature of the application of artificial intelligence and machine learning in battling COVID-19

RT-PCR (Reverse Transcription Polymerase Chain Reaction) test is the golden standard to diagnose COVID-19. However, it is also prone to false negatives and inaccurate results. It might also not be able to detect newer strains of COVID-19 in the future. Hence the use of CT-Scans, X-rays, blood tests and sound analysis should also be used in COVID-19 diagnosis. They can also be used in events, such as a pandemic peak, since the RT-PCR test takes a considerable amount of time to give results. These methods can also be used along with RT-PCR tests to increase the sensitivity.

Machine learning applications are constantly used in health care industries. They can be trained with a large amount of data. Once trained, they are known to produce accurate outputs. They can assist the health care workers, since these algorithms are extremely fast and efficient. The above studies have proved that the use of Machine Learning and Deep Learning in the diagnosis of COVID-19 are extremely helpful. A huge amount of research is happening each day; hence it is very important to classify and group the literature using the process of systematic literature review. This helps in bridging the gap between the researchers and the technologies involved. This paper aims to answer the following questions:

* How does the use of cutting-edge machine learning techniques in COVID-19 improve system accuracy?
* Which of the various COVID-19 diagnostic techniques are better when compared to each other?
* What are the current challenges and the possible future research trends?

**Machine learning overview**

This section presents a brief recapitulation of the ML algorithms that help in tackling COVID-19. The application of ML in tackling COVID-19 is very important for the current situation. Supervised learning, unsupervised learning and reinforcement learning are the main types of ML algorithms.

**Supervised learning**

The training data is labeled in supervised learning ML algorithms. The set of output/labels are also given. These algorithms learn from the dataset and after training a certain amount of data, predict accurate outputs.

**Classification**

Trained data is grouped into one of the predefined classes (Soofi & Awan, [2017](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). One of the disadvantages of these algorithms is their inability to handle missing data. But fortunately, these missing data can be replaced with existing data (Kantardzic, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). There are many classification algorithms which are being used to analyse COVID-19 trends. The following are some of them.

* **K Nearest Neighbours (KNN)—**Classification is based on the location of the data points. Similar objects are grouped together based on the number of neighbors. The classification is non-parametric. No assumptions are made about the data. It is also called a lazy learner algorithm since its learning procedure is slow and the action is performed only at the time of classification. A user-defined k is set as the minimum number of neighbours required. Generally, when the value of k is higher, the predictions are more trustworthy. Weights are attached to the neighbours to increase accuracy during classification. KNN was used in (Ginantra .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) to detect respiratory infections. Yin .(Yin .[2018](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), used KNN to detect serious influenza. Locations of the infected users were tracked in (Cho, [2016](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) using KNN.
* **Support Vector Machine (SVM)—**A hyperplane is used to classify the data into different groups. SVM finds the maximum marginal distance to find the minimum gap among data points. Hyperplanes act as boundaries and separate the data points into classes. A single line exists when there are only two attributes. A 2D plane is generated when the input size is increased to three. When the number of features increases, it gets very difficult to picture the hyperplane. Mori .(Mori .[2016](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), used SVM to predict the outbreak of disaster for a specific location. SVM was used with IoT (Internet of Things) and CNN (Convolutional Neural Networks) for the diagnosis of COVID-19 (Le .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Naïve Bayes—**Bayes’ theorem can also be used to classify the dataset. All the features which get classified are assumed to be completely independent of each other, hence the name naïve. Splitting of data is done into two components, feature matrix and response vector. In the feature matrix, all data is stored in the form of rows, and in response vector, the outcome class is defined. In (Sadhukhan .[2018](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) and (Assery .[2019](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), Naïve Bayes classifier has been used to group tweets. It helped manage social networking problems during the period of the pandemic.
* **Logistic Regression—**It is used to predict the categorical variable (dependant variable) based on one or more independent variables using a sigmoid curve as a cost function. Sigmoid function (Logistic function) is a S-shaped curve that classifies the data into one of the classes. Logistic regression can be used in binomial, multinomial and ordinal classification. Mohammadi .(Mohammadi .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) used ANN and LR to classify COVID-19 patients in Iran. In (Bhandari .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), Logistic Regression was used to predict the mortality rate from various blood and clinical parameters. Forecasting of COVID-19 in Kuwait was done using LR in (Almeshal .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Decision Trees**—Leaf nodes predict the outcome, decisions are taken by the branches and the features represent internal nodes of a decision tree. Decision trees distribute instances by assembling them down the tree from the root to leaf nodes, which provides the classification of the instance. In (Elhoseny, [2019](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), the location of the users has been detected using decision trees during pandemics. In (Chen & Liu, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), the prediction of different factors which led to psychological distress during the pandemic was discussed with the help of decision trees. A hybrid face mask detection application was developed by using decision trees in (Loey .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Random Forests**—Classification using a decision tree can lead to overfitting when the data is large. These limitations can be fixed by random forests. Different decision trees are classified and then merged together to improve accuracy. This method is called “bagging”. COVID-19 health prediction was made using the RF algorithm in (Iwendi .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The random forest machine learning approach was used to estimate the spatial-temporal distribution of COVID-19 daily instances around the world (Yeşilkanat, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Artificial Neural Network (ANN)**—They are a completely connected network and are multi- layered. The layers include an input layer, an output layer and many hidden layers. The nodes in one layer are attached to the previous and next layers. Inputs pass through an activation function. The output of one layer becomes the input for the next layer. ANN along with IoT was used to detect the location of the user (Luoh, [2013](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Here the data was limited, yet there was a high accuracy. In (Polese .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), ANN was used to detect the exact group of people present in an area.
* **Deep Neural Network (DNN)**—DNN is an adaptive model and also offers creativity since neural networks are generally monotonous. The total number of nodes to analyse results are very large. DNN consists of three layers- Input, output and many masked levels. Progressive learning and the ability to modify their results are the reasons DNN is used. Chhikara .(Chhikara .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), used DNN for emergency crowd evacuation. It is important to transfer healthy people to non-infectious zones.
* **Convolutional Neural Network (CNN)—**CNN is a better algorithm to classify images since it has a

Better accuracy and capturing orientation. They have many different layers called pooling layer, convolution layer, fully connected layer, etc. A CNN based model (More .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) was used with IoHT (Internet of Health Things) for diagnosing people with COVID-19.

**REGRESSION**

The regression technique is a predictive learning feature that maps a data object to a predictive variable with real meaning (Gutierrez .[2016](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Regression uses methods of statistics to establish a relationship between the target and predictor variable. Regression also helps us to model the association between a dependent and independent variable. Variables like age, temperature, salary, etc. can be easily forecasted after training a certain amount of data.

* **Linear Regression** linear regression is a simple algorithm that is used for predictive analysis. It shows the linear relationship between dependent and independent variables. If only one input exists, such a model is called simple linear regression and if more than one input exists, the model is termed a multiple linear regression model. A multivariate linear regression model was used to predict new active cases of coronavirus disease (COVID-19) (Rath .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Predictive modelling of COVID-19 positive samples in Nigeria was made using a linear regression model (Ogundokun .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Polynomial Regression**—Non-linear dataset can be modelled using polynomial regression. A non-linear curve is set for the dependent and independent variables. Polynomial features of a given degree are obtained from original features and then the model is developed using linear regression. COVID-19 epidemic estimates in India based on the polynomial regression model were analysed in (Pandey .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The use of hierarchical polynomial regression models to forecast COVID-19 spread at the worldwide scale was used in (Ekum & Ogunsanya, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Support Vector Regression**—Support Vector machines can be used for classification as well as regression. A hyperplane with maximum marginal distance is determined to fix the maximum number of data points near the margin. Maximum data points must be within the boundary lines and the maximum number of data points must be included in the hyperplane. It was used in the short-term prediction of COVID-19 confirmed cases in Brazil (Ribeiro .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Ridge Regression—**Ridge regression includes a small amount of bias which helps in getting better long-term predictions. This bias is called the ridge regression penalty. The penalty can be calculated by the individual features. High collinearity between independent variables is the use case of the algorithm, since both linear and polynomial regression would fail in the above condition. The model can also be used if there are more samples than parameters. It also reduces the complexity of models; hence it is also a regularization technique. Forecasting the development of the COVID-19 epidemic using a hybrid polynomial-Bayesian ridge regression model was discussed in (Saqib, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).

**Unsupervised learning**

The data is unlabelled in unsupervised learning. The user does not know the pattern of the data. The algorithms use distinct techniques to analyse the information structure. The most common algorithm which uses unsupervised learning is clustering.

* **K-Means—**Data is divided into different clusters. Initially, total number of clusters are assigned and random centroids are chosen among the dataset. In each iteration, the centroids are updated. Either Euclidean distance or cosine distance is used to calculate the similarity score. Projection of the number of reported COVID-19 cases using k-means was discussed in (Vadyala .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The prediction of COVID-19 cases using the K-Means algorithm in the United States of America was discussed in (Zhang & Lin, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **K-Medoids**—It is used when there are a lot of outliers in the data. Data points are chosen randomly as a median. The rest of the data points are placed among these medoids depending upon the minimum distance. After certain iterations, we get medoids which successfully cluster the dataset. Deployment of Data Mining Techniques which included k-medoids algorithm in National Food Security in Indonesia during the COVID-19 pandemic was discussed in (Elsi .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Fuzzy C-Means—**Centroids are randomly assigned and data is initialized randomly to these clusters. This algorithm operates by assigning membership to each data point corresponding to each cluster centre based on the distance between the cluster centre and the data point. More the data is near the cluster centre, more is its membership towards the particular cluster centre. A development of an intelligent model based on the Lévy slime mould algorithm and adaptive fuzzy C-means for detecting COVID-19 infection in chest X-rays has been developed in (Anter .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). ROI extraction in CT lung images of COVID-19 using Fast Fuzzy C means clustering was discussed in (S. N. Kumar .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). This helped the radiologists and other medical professionals.

**Reinforcement learning**

Maximizing the reward by taking suitable action in a particular situation is called reinforcement learning. Various machines and software try the best path or behaviour for a given situation in this method. In supervised learning, the answer key is available to the data during the training phase, whereas, in reinforcement learning, the reinforcement agent determines the course of action and no answer key is available to guide the model. The models learn from experience even in the absence of training data set.

* **Q-learning**—Sequential decision making is used in Q-learning. For each input, the best solution is calculated and the corresponding input always depends upon the previous output. Objective function Fπ can be maximized or minimized based on various scenarios. π corresponds to the policy and P defines a state. Its optimal policy comes from its interactions with the environment. A strategy for the prevention of COVID-19 affected patients using multi-robot collaboration and the Q-learning approach was presented in (Sahu .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). A Novel Methodology Based on Deep Q-learning/Genetic Algorithms for Optimizing Covid-19 pandemic government actions was discussed in (Miralles-Pechuán .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).
* **Markov’s decision process**—Learning from interactions to achieve a specific goal is the fundamental technique in this algorithm. The environment and the reinforcement agent continually interact with each other, the agent selects the action and the response to these actions

are given by the environment, which in turn presents new scenarios to the agent. The epidemiological Markov model was used to analyse the COVID-19 pandemic and clinical risk factors of patients in (W. Zhang .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).

**COVID-19 case data**

COVID-19 patients and their corresponding locations can help keep track of the pandemic growth in many areas. Many countries are supporting data collection and sharing information related to COVID-19. John Hopkins University collects the data regarding the positive cases which occur each day. It also collects information regarding the death rate and also the number of patients cured (CSSEGISandData, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Kaggle also updates the COVID-19 data regularly (sudalairajkumar Data, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Here, other details of patients, including location, reporting date, etc. are shared. nCOV2019 (nCoV2019Data, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) contains important information like national health reports of the COVID-19 victims. Details like date of confirmation, travel history, list of symptoms and geo-locations are shared. New York Times gives state wise details which consist of positive cases and casualties each day (Nytimes, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The above datasets are given by the government but community datasets are also available. People also report new infections using their social network (B .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Applications which thrive on data visualization can use the dataset (Koubaa, [2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The limitation of these datasets is that they belong to a particular country most of the time.

**Textual data**

Online sources are used to collect a huge amount of data which can be used to understand the nature of the virus. An important source is social media. Discussions made about COVID-19 are already available. Tweet ID’s (E. Chen .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) and tweet text data (Smith, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) are being used by various communities. These datasets can be used with NLP (Natural Language Processing) to monitor the spread of this deadly virus (Aramaki .[2011](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Culotta, [2010](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Lampos .[2010](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Sharma . (Sharma .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) has also separated COVID-19 data after summarising 5 million tweets. Zarei et al. (Zarei .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) have used Instagram to analyse 5300 posts related to COVID-19. Academic publications have also created a lot of textual information. Clinical studies are used to extract information (Zhao .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). NLP models are used for this purpose (Manning, [2016](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Allen Institute has contributed 45,000 articles which relate to COVID-19 (Allen-Institute, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). NCBI (National Centre for Biotechnology Information) is also contributing with data (NCBI, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). WHO has collected recent research publications and has dedicated an entire database to it (WHO, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Wikipedia has also collected substantial information about COVID-19 which can be easily downloaded for analysis ([2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)).

**Medical data**

Medical data is used for this diagnosis and treatment of COVID-19. X-rays and pathology reports can be used for prognosis. There are a few open-source X-ray scans of COVID-19 patients (Wang & Wong, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). They can be used to assess and diagnose infections using computer vision (Mery, [2015](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). There are other X-ray datasets as well, such as ([2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Cohen .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Mvd, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). These datasets cannot be easily understood by a non-biological student. Thus, we require help from radiologists and clinicians to properly label the data. The datasets are small and deep learning techniques might not be efficient on them. Along with X-rays, CT scans are also used for COVID-19 diagnosis. A lot of CT scan datasets are already available. In (MegSeg, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), CT scans of 60 patients were taken. A larger dataset which includes 288 CT scans is present in (J. Zhao .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The United Kingdom also provides COVID-19 chest X-rays and CT scans of various patients who were infected (Nhsx, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Leeds University Institute[119], UK NHSx [120] and Ellis Alicante (E. A. Foundation, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) also provide images of infected lungs of COVID-19 patients. Genomic sequencing datasets are also available. The datasets RCSB (Research Collaboratory for Structural Bioinformatics) data (Goodsell .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), And GHDDI (Global Health Drug Discovery Institute) (Zhavoronkov .[2020a](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) analyse the structure of RNA, the structure of spike protein, etc.

**COVID-19 data from developing countries**

Countries like Algeria, Kenya, Egypt, South Africa, Nigeria and Tanzania are sharing data. Zhao et al. (Z. Zhao .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) studied the COVID-19 analysis for African countries like Senegal, South Africa and Kenya. The dataset provides the number of cases occurring each day, deaths and the recovered patient details. West Africa has also provided the dataset (Martinez-Alvarez .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). State-wise analysis along with other demographics are provided by these datasets.

**Other datasets**

There have been other useful datasets which have become prominent during the time of the pandemic. For example, air quality improved during lockdown since there were very few vehicles and the industries had shut down. Air quality of six major cities had improved rapidly due to the lock down (Cadotte, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The data is available to the public (W. Inc., [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Mobility trace data is available in [128]. In (Wurtzer .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), samples of wastewater were taken. It was found that 6 samples tested positive for COVID-19.

**Medical diagnosis OF COVID-19 using ML**

COVID-19 cases are increasing every day. Medical images have started to play an important role in the diagnosis of COVID-19. CT scan and X-ray of patients using AI have been used to facilitate health care personnel. These images are tested and classified based on geographic infection classification, organ recognition and disease classification. The prediction of diagnosis becomes faster using these methods. The prognosis of the disease can also be easily monitored.

**COVID-19 detection using lung CT scan**

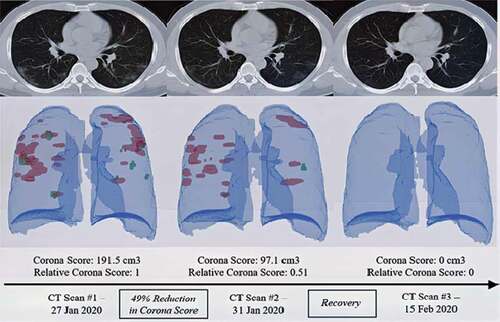
CT scans are used to diagnose COVID-19 along with RT-PCR (Real-Time Polymerase Chain Reaction) tests. It has been found that the infection generates ground-glass opacities and long periphery consolidations. These can also occur before the lab test. Hence, health care professionals can use CT scans to diagnose the disease when a lot of people are being tested. The scan also shows the severity of the affected lungs and also how the illness can progress. The doctors can make medical decisions based on it. Lung defects because of COVID-19 are unique and can be easily diagnosed. The CT scan can take up to 15 minutes, but the AI can validate the extent of severity within seconds [143]. The function of CT Scan can improve using Advanced Image Processing and ANN by training a large amount of CT scan images of the COVID-19 patients. CT imaging tests with AI normally have the following steps, division of Region of Interest (ROI), pulmonary tissue removal, infected region removal and finally, classification of COVID-19. The images of lung organs and ROIs combined with AI-based testing turns out to give favourable results.

ROIs can be used to find ulcers, bronco pulmonary segments, lobes of the lungs infected, etc. The CT images are classified using various types of DL networks. E.g.: V-Net, U-Net, VB-Net, RPN have been used. After the primary CT scan is taken, deep CNN is used to evaluate the characteristics of the image concerning SARS-CoV-2 victims. The CNN based algorithm obtained an Area Under Curve (AUC) of 92%. Detection of patients using AI systems enhanced and it correctly classified 17 patients out of 25(65%) and the radiology department confirmed that they were negative (Mei .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The CNN model predicted 13 out of 25 (52% of the images), as positive. The clinical model classified 16 out of 25(64%), as COVID-19 positive. The combination of the models classified 17 out of 25 as COVID-19 positive. A research team of the Tianjin Medical University Cancer Institute and Hospital had obtained CT scans of 180 patients who had severe pneumonia even before the pandemic (PhysicsWorld, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). CT scans of 79 positive COVID-19 patients were taken in (Wang .[2021a](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). A CNN based model was trained using these images. An accuracy of 89.5 % was obtained after testing the images. The radiologists who identified the images were only 55% correct. RADLogics algorithm (Scudellari, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) was used to monitor the condition of diagnosed COVID-19 patients using CT-scan imaging. A study (K. Zhang .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) used DL to train CT scan images to search for COVID-19 infection among patients admitted to the hospital. A few researchers trained 532,000 CT images to train their AI-based model. The feature they were concentrating on was “tell-tale lesions” which was unique for COVID-19 patients. The AI model had an accuracy of 85% and successfully differentiated COVID-19 patients from pneumonia. For radiologists, it is very difficult to distinguish between regular pneumonia and the disease which led to the pandemic. Hence a company, VIDA Diagnostics (VIDA. LungPrint Clinical Solutions, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), created a LungPrint device which uses AI to analyse CT scans which correctly identified all respiratory conditions, including COVID-19. In (Analytics, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), companies NLH and NVIDIA created a DL model to diagnose COVID-19 using CT scans. The datasets came from Italy, China and Japan. 2724 samples were taken from 2619 patients. Two models were developed (Full 3D, Hybrid 3D). These models were trained to predict COVID-19 by using lung CT Scan images. The hybrid model had a better accuracy of 92%, whereas the full 3D model had 91.7% accuracy. Chen (Chen .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) designed U-Net++ which used a deep learning structure. 46,096 images were collected for model creation and validation. It achieved an accuracy of 95.24%, specificity of 93.5% and 100% sensitivity. Huang et al. (L. Huang .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) used a method called InferRead TM (Thermo Meter) CT which measures the infection severity in the lungs. Pulmonary and lobe extraction, categorization of pneumonia and medical metrics were the three modules incorporated in the tool. Mild, moderate, severe and fatal stages were the four modalities. A deep learning software was trained with 126 images, six were mild, 94 were moderate, 20 severe and 6 fatal cases. CT Scans of 52 COVID-19 patients were analysed by Qi .(Qi .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). A methodology called pyradiomics was used to find 1218 traits from each image. CT radiomics models used two algorithms, LR and RF. Pneumonia lesion extracts were used as classifiers during training. They successfully predicted the severity of COVID-19 among patients. The duration of time a person had to stay in the hospital was differentiated. If the person stayed for less than 10 days, it was a short term stay, while the long-term stay was more than 10 days. LR model achieved a sensitivity of 100% and a specificity of 89 %. The RF model showed a sensitivity of 75 % and a specificity of 100%.

**COVID-19 detection using chest X-rays**

Chest X-ray (CXR) images are an exceptionally useful method to diagnose COVID-19 patients. [Figure 2](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666" \l "f0002) shows the architecture of DL-based CT image classification for COVID-19 diagnosis (Scudellari, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). X-ray images are easier to understand and can also be easily acquired unlike CT images. The studies (Basu & Mitra, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Pandit & Banday, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) analyse X-ray images for detection of COVID-19. AI based CXR image testing involves correction of data, training and finally testing of the images. Deep learning algorithms such as CNN, U-Net++, etc. are used for faster detection of COVID-19 from CXR images. CXR machines are less costly than CT scan machines and the images are processed faster. Normally, the X-rays are interpreted manually by radiologists. When it comes to COVID-19 cases, radiologists can detect only 69% of X-rays due to lack of technology (Cristobal.baeza, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Trained ML models can detect COVID-19 accurately and quickly.

Figure 2. **The RADLogics method calculates the amount of recovery using a specific score based on different computed tomography (CT) scans (Scudellari, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)**)

[](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)

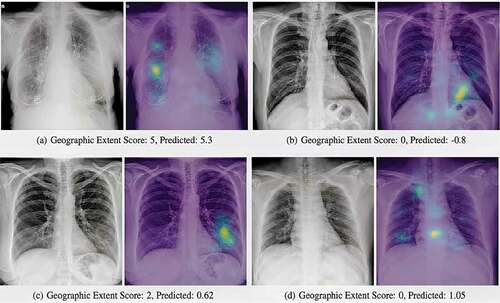
In (Pandit & Banday, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), a dataset of positive COVID-19 cases, typical pneumonia and non-infectious cases were compared and 1428 CXR scans were taken. A trained model of VCG-16 was used to perform the classification. The results had an accuracy of 96 percent. Many DL algorithms exist which provide accurate results using CXR images in detecting COVID-19. Basu and Mitra (Basu & Mitra, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) worked on a concept called domain transfer learning to detect COVID-19. A collection of 20,000 CXRs was trained and tested by the Gradient Class Activation Map to detect the viral infection. The CXR dataset was divided into 4 groups, i.e., no disease, other viral diseases, bacterial pneumonia and the dreaded COVID-19. A fivefold cross validation technique was used for disease diagnosis. 100% accuracy was achieved for COVID-19 data and the overall accuracy was 95.33 %. Students of Cranfield University developed an AI-based model to detect COVID-19 using CXR images (Cranfield University, [2020a](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666), [2020b](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). To obtain various characteristics of CXR images, various ML and DL techniques were used. The model would detect COVID-19 which would not appear generally to the naked

eye. There were 2 models, the first model differentiated normal patients and patients with pneumonia. The second model distinguished between pneumonia and COVID-19. Using chest X-ray scans for COVID-19 detection, another model (Ozturk .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) using the Darknet-19 model was developed. The sensitivity achieved was 85.35%, specificity 92.18% and f1-score of 87.37%. Researchers of King’s College London General Hospital and a tech company called “Zoe” used AI-based detection to identify COVID-19 pathogens. Symptoms were assessed along with the routine RT-PCR COVID-19 tests (Cranfield University, [2020c](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). 100% sensitivity and 90.5% accuracy were achieved by a model called COVID-19AID (Mangal .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). It used a DNN model which was trained on CXR images that were publicly available. In (Das .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), a CNN model named Truncated Inception Net accurately distinguished between COVID-19 positive and negative cases. The dataset was divided into four categories- Tuberculosis positive, pneumonia positive, COVID-19 positive and healthy patients. A 99.96% accuracy was obtained when detecting COVID-19 from normal cases and pneumonia. While comparing COVID-19 patients with others, the accuracy achieved was 99.92%. Transfer learning was used on CXR images to predict COVID-19 patients by Minaee et al. (Minaee .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Four CNN models, DenseNet-161, Squeeze Net, ResNet, ResNet50 were used. They used datasets of both healthy and COVID-19 patients. The average specific rate was 90%, sensitivity was in the range of 97.5% for these models. A decision tree classifier was used to classify CXRs in (Yoo .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). There were three decision trees which were trained by CNN. The first tree separated normal X-rays from abnormal ones, the second decision tree separated tuberculosis images from normal images and the third tree separated COVID-19 from tuberculosis images. An accuracy of 98%, 80% and 90% was obtained by the decision trees respectively. GAN (Generative Adversarial Network) was used in a study (Loey .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) along with deep transfer learning to identify COVID-19 patients. There were 4 classes mainly normal, COVID-19, viral pneumonia and bacterial pneumonia. The three transfer models GoogleNet, AlexNet and ResNet18 were used for testing. In the first scenario, all 4 data classes were used, while three classes were used in the second scenario and only two classes were used in the last scenario. COVID-19 class was present in all scenarios. 80.6% accuracy was obtained in the first scenario using GoogleNet. In the second scenario, AlexNet obtained an accuracy of 85.62%. In the third scenario, 100% accuracy was obtained by GoogleNet. To classify COVID-19, 24,678 CXR images were used in (RadBoudumc, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). CAD4COVID-19-XRay, a deep learning model was trained. Segmentation of lungs was obtained using U-Net and CNN. An AUC of 81% was obtained by the model. The author concluded that it could be useful in low resource systems where there are no proper diagnostics equipment.

**Classifying the extent of severity of COVID-19 with ML**

Disease progression using CXR images of lungs can be monitored efficiently with the help of ML and DL techniques. In (Cohen .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), studies are being made to determine a severity score based on CXR images. A DenseNet neural network model is used to predict disease progression. 94 images of COVID-19 patients were used for disease progression classification. Two kinds of scores are being used as in Table 4. The lung participation and ambiguity are being shown in a single X-ray image. The severity of patients using the DenseNet model can be observed in [Figure 3](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666" \l "f0003)

Figure 3. ***DenseNet model to predict severity of COVID-19 from X-rays (***Cohen .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666))

[](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)

For each particular lung, the lung involvement level was calculated by using a degree of unification and ground glass ambiguity. An increase or decrease of supervision, tracking of the effectiveness of treatment in the ICU, can be found out using the score. CSNN (Convolutionary Siamese Neural Network) was used in (Ridley, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) to generate a score which would predict whether the patient required intubation or not. In (Qi .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), 71 CT scans of 52 COVID-19 patients were obtained. 1218 attributes were extracted from CT images using the methodology pyradiomics. Models using LR and RF were built based on pneumonia lesions. 97% and 92% AUC were obtained during testing of LR and RF models respectively. LR model obtained a sensitivity of 100% along with 89% specificity. RF model obtained 75 % sensitivity and 100% specificity. If the number of days a patient stayed in the hospital was less than 10 days, then it was classified as short term stay, or else it was classified as long term stay. In (Health EUROPA, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); New York University, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), a mobile application was developed which used CXR images of 160 patients. The biomarkers would diagnose the severity of COVID-19 from mild (level 0) to extreme levels (level 100).

**COVID-19 detection using sound analysis**

Coughing can be a symptom for more than 30 diseases other than COVID-19. So COVID-19 detection using cough alone is difficult. Various sounds like breathing, moaning, heartbeat, etc. can be used for diagnosis. Auscultation was used in the hospitals to collect such signals. Table 5 gives a brief overview of researches related to COVID-19 detection using sound analysis. Research has already started on how sounds can be recorded and trained to diagnosis healthy and COVID-19 patients. A phone app (Imran .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) used the dataset of 48 COVID-19, 131 pertussis, 102 bronchitis and 76 healthy cough recordings to train the model (Stevens .[1937](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). KAs (HospiMedica, [2020a](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), a mobile application was developed by Zensar Technologies. They used AI to monitor patient health by using various cough signatures which were disease-specific. The patients had to answer 15 questions and the cough sound was recorded by the company’s AI platform. Based on the answers and cough sounds, a rating from 1 to 10 was given where 1 meant minimum risk and 10 meant maximum risk. A person’s score was monitored for many days to diagnose COVID-19 accurately. Computer Audition (CA) was combined with AI to study the cough sounds of various COVID-19 patients (Schuller .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). However, there is not much documentation about the effectiveness of this method. A respiratory simulation called BI-AT-GRU (Y. Wang .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) was developed using neural networks. It differentiated 5 types of respiratory sounds such as biots, Cheyne-Stokes, Central-Apnea, Eupnea and Tachypnea. Sputum testing was used in COVID-19 diagnosis. The project was called Coswara (Sharma .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) and it used various biomarkers like cough, breath and other speech sounds. Nine different vowel sounds were recorded. The classification mechanism was still under development when this paper was written. Flusense (Al Hossain .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), a portable device, was developed by the University of Massachusetts, Amherst. It is based on a neural network model which identifies cough and properly detects diseases including COVID-19. The developers claim that this device is being used in many medical centres. Ravelo et al. (Ravelo, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), recently launched an AI-based system which used cough to diagnose COVID-19. The user had to provide information such as geographic location, symptoms, cough sound, etc. Individuals also had to mandatorily upload the results of their COVID-19 test too. Using these, a DL based algorithm was used to determine COVID-19 infection using the patient cough sounds and other data. Iqbal et al. (Iqbal & Faiz, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) used a mobile application which recognised suspicious cough sounds to determine the respiratory health of a person. A system which used both coughing and breathing was created by a research group at the University of Cambridge (Brown .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The breath sounds of COVID-19 patients, healthy people and asthma patients were separated. 7000 users were tested with a minimum of 200 registered COVID-19 patients. LR, Gradient Boosting Trees and SVM models were used. AUC curve along with other parameters were used to find the efficiency of the algorithm classified. The average AUC was around 60%. 367 breath sounds were taken to develop a low-cost cell phone model (Bagad .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Analysis was done extensively on the breath sound. The model detected pneumonia from non-pneumonia successively.

**COVID-19 prediction using blood tests**

Identification of COVID-19 infection is generally done using RT-PCR tests. However, the test takes a considerable amount of time to give results (4–5 hours under ideal conditions). It also needs trained personnel to handle the samples. Its sensitivity is also an issue (unable to avoid false negatives). CT scans and X-rays have been used with the help of ML to provide accurate results. CT scans can be very expensive and might not be affordable for everyone, especially those who live in underdeveloped countries. They can also cause cancer if used excessively. However, X-rays are affordable, but they are also prone to false-negative results (Echenique, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Blood test results are available within 15 minutes. They can be used to detect COVID-19 in low resource countries or in events such as a pandemic peak. It can also be used together with RT-PCR test, to increase the sensitivity. Two ML models were used to predict COVID-19 in San Raffaele Hospital (Brinati .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). White blood cells (WBC), platelets, C-reactive protein (CRP), Aspartate Aminotransferase (AST), Alanine Transaminase (ALT), Gamma-glutamyl transferase (GGT), Alkaline phosphatase (ALP), Lactate dehydrogenase (LDH) of 279 patients were taken from routine blood tests which were used to train the ML models. One model was used to predict positive cases and the other to predict negative cases. The accuracy obtained was 82% and 86% respectively. 879 confirmed COVID-19 cases from NHS Trust hospital, England were studied in (Heldt .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Anonymised demographic data, laboratory results and physiological clinical variables were extracted from EHRs (Electronic Health Report). Multivariate logistic regression, extreme gradient boosted trees and random forest algorithms were used for classification. The severity of COVID-19 cases was predicted based on the blood samples. Blood lactate and creatinine levels played a pivotal role in predicting COVID-19 patient trajectories. AUC scores ranged from 0.76 to 0.87. 72 features from the blood sample of 1624 patients were used for COVID-19 detection with the help of ML models in (Federico .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). AUC obtained by the algorithms were in the range of 0.83 to 0.90. EHR reports of 1357 patients from different European countries were collected to predict COVID-19 in (Tschoellitsch .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Blood count, electrolytes, creatinine, liver enzymes, bilirubin, c-reactive protein were all included in the final model. An accuracy of 86% percent and an AUC of 0.74 was obtained by this model. In (Kang .[2021](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), blood reports of 151 COVID-19 patients from Tongji Medical College, China was collected. ANN algorithms were used to develop the models which predicted the severity in COVID-19 patients. The average AUC obtained was 0.953. 253 samples from 169 patients were collected from four different hospitals in China (Wu .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). 24 haematological parameters and 25 biochemical parameters were selected for analysis. Accuracy of 0.9795 and 0.9697 was obtained from the training and testing set respectively. Prediction of Accurate Respiratory Distress Syndrome (ARDS) was researched in ([Xu .,](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Details of 659 COVID-19 patients were taken from 11 different regions in China. The studies concluded that male patients with BMI > 25 were more prone to develop ARDS. Abnormal count of lymphocytes, Creatine kinase, NLR (Neutrophil to lymphocyte ratio), LDH and CRP led to ARDS. DT algorithm achieved the best results. A Gradient Boosting Decision Tree (GBDT) model was used to detect COVID-19 in (Yang .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). 3356 patients (1402 positive, 1954 negative) were observed and 27 blood parameters were used to know the COVID-19 infection status. An AUC of 0.854 was obtained. The external dataset result had an AUC of 0.838. Non-severe COVID-19 patients showed high values for lymphocytes, platelets and eosinophils and had low neutrophils, delta neutrophil index (DNI), NLR and platelet-to-lymphocyte ratio (PLR) in the study (Kazancioglu .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). However, severe patients showed an increasing trend for PLR, platelets and eosinophils. Laboratory data of 120 patients were chosen for this research. An AUC of 0.819, 0.817 and 0.716 was achieved. The research concluded that NLR and PLR values could be used to predict the severity of COVID-19 using different models in ML. Various ML models and ANN was used to identify COVID-19 from full blood counts in (Banerjee .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The data was used from Israelita Albert Einstein hospital, Brazil. RF, flexible ANN and shallow learning models were used for prediction. The study said that COVID-19 patients tend to fewer less platelets, eosinophils, leukocytes, lymphocytes and basophils. However, an increase in monocytes was noted. An AUC of 94–95% was achieved by the models. Table 6 gives an overview of other researches which use ML to identify COVID-19 by using routine blood samples. Table 7 compares all the above models which were discussed.

**Machine learning for vaccine and drug development**

In combination with a huge quantity of data, the capacity for automatic abstract component learning has had a great impact on the efficacy of ML. Vaccination and drug discovery are fields of extreme importance. ML is used to provide behaviour prediction, integrated property prediction and ligand-protein interactions. The vaccines which are used for the treatment of COVID-19 focus on genomics and proteomics study. ML has contributed to intelligent medicine, using which the development of new medicines has taken place.

**Vaccine development using ml for COVID-19**

ML and DL help by observing the viral protein structure that helps in the vaccine components dissemination, and also helps medical and biological researchers in reviewing large research papers at a fast pace. Vaccines are mainly available in three types: Vaccines for pathogens (e.g.: Common flu), subunit vaccines (e.g.: shingles, pertussis) that use only a small part of the virus such as a protein, and vaccines which use nucleic acid that inject a small number of viral genes, so as to make the body immune to the virus (Decario, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). ML is used to speed up the growth of nucleic acids and subunit vaccine types. The structure of protein composition has to be analyzed and researchers develop new drugs that can work for different protein structures. However, finding this unique 3D structure takes a lot of time. Evaluating genetic sequences and protein structure can be simplified with the help of ML and DL systems. AlphaFold was introduced by Google DeepMind (HospiMedica, 2020; Institute for Protein Design, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), an advanced ML application that predicts the function of 3D proteins using their genomic sequences. The system can be used for COVID-19 as well. It also helped the scientific community to understand the virus by publishing a protein prediction for all the proteins that are related to SARS-CoV-2. The first 3D atomic map to understand the spike protein component of the virus was developed by the researchers at the University of Texas. AlphaFold has provided accurate predictions for the protein spike structure. 3D atom models for the SARS-CoV-2 protein scale that were simulated (Rees, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) was used by the researchers of the University of Washington. New proteins were created to emulate Coronavirus. These proteins would embed with the healthy cells to prevent viral attacks. AI was combined with cloud computing to prevent the spike protein from binding to healthy human cells. A vaccine could be produced using this technique (TABIP, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). The researchers of Flinders University studied the structure of Coronavirus and then designed a vaccine using the data collected (Flinders University, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). It was called Covax-19 as shown in [Figure 4](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666" \l "f0004).

Figure 4. **Covax-19™ is an Australian-designed COVID-19 vaccine modelled by AI-based technologies** (Flinders University, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666))

[](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)

The team used AI and cloud technologies to speed up the production of COVID-19 vaccines in (Flinders University, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Herst .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Institute for Protein Design, [2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). Computer- based models of S protein along with its human receptor, the enzyme converting angiotensin 2 (ACE2) was used to understand how the COVID-19 virus was harming healthy human cells. Herst et al. (Ong .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)), used MSA (Multiple Sequence Alignment) algorithm for the detected GenBank’s SARS-CoV-2 protein sequence to model the nucleocapsid phosphoprotein sequence in the future peptide sequencing. A peptide-based vaccine was used to treat another epidemic which had happened in West Africa (Ebola, 2013–2016). Ong et al. ([Deoras](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) used reverse vaccinology (RV) and ML to test various COVID-19 vaccines. RV was used to analyse bioinformatics pathogen genomes. Biological signals were forecasted using models called Vaxign and Vaxign-ML. Algorithms like vector support, RF, proximity neighbour, and overgrowth (XGB) strategies were used by the above models. Researchers from MIT (Rahman .2020) have developed an AI-based method that chooses peptides that give large vaccine numbers. The design software is named “OptiVax”. It designs new peptide vaccines by learning from the existing vaccines to improve composition. Immunoinformatic methods were used to beat the virus. Rahman et al. (Sarkar .2020), produced an antipeptid vaccine for COVID-19 which consisted of envelop, membrane and spike protein. A predictive system called Ellipro antibody epitope was used to predict B-specific epitopes in spike protein. To visualize and forecast a protein sequence, Ellipro used several models of ML. The epitope-based vaccine model of COVID-19 was also explored by Sarkar. (Prachar .2020). SVM technique was applied to predict the toxicity of choses epitopes. 174 epitopes of COVID-19 were forecasted by Prachar et al. (Xue .2018). IEDB (Immune Epitope Database) was used to collect data and PSSM (Position-Specific Scanning Matrix) and ANN was used for prediction.

**Drug development using ml for COVID-19**

AI systems can be used to battle infectious diseases like COVID-19, by easily classifying and predicting appropriate drugs. Drug development is a very costly, lengthy and dangerous phase. The success rate is only 2.01% and it takes a minimum of ten to 15 years to develop a completely new drug according to Eastern Research Group (ERG) (Mohanty .2020). Drug repurposing uses an experimental strategy to reuse pre-approved, discontinued and shelved medications. Normally, drug development involves 5 steps: Drug discovery and development, pre-clinical research, clinical research, FDA (Food and Drug Association) drug review and FDA drug aftermarket safety control. However, drug repurposing involves only 4 steps (Beck .2020): Compound identification, Drug clinical research and FDA drug after-market safety control and development. In one research, AI is used for drug discovery at the very molecular level. Beck et al. (Pahikkala .2015), recommended a DL-based drug-target interaction model, named Molecule Transformer Drug Target Interaction (MT-DTI). This model can be used to predict drugs for COVID-19. The MT-DTI model used amino-acid sequences and SMILES strings to classify target protein. The inhibitory potency of the chemical compound atazanavir (a drug to treat and prevent HIV) was the best according to the study. Remdesivir also produced positive results. Two similarity-based models KronRLS (He .[2017](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) and SimBoost (Öztürk .[2018](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) that use ML for drug discovery have been proposed. However, there are a couple of disadvantages. The representation of features is reduced drastically which can result in faulty products. Also, estimation of matrix of similarity limits the number of molecules. A DTI model which used DL, DeepDTA (K. Huang .2021), was introduced to improve the above models. It used a CNN based model that waives requirements related to feature engineering. From protein sequences and raw molecules, useful features are automatically detected by the model. Sim-Boost and KronRLS both used this model. Feature representation from SMILES strings and raw protein sequences were trained using a CNN model, and named it DeepDTA. Proteins and ligands pair-wise similarities were processed using Smith-Waterman and Pubchem similarity algorithms respectively. Three alternative combinations were trained to develop a more advanced DeepDTA model. These included training compound representations alone, training protein sequence representation alone and the third model used both protein and compound representation. The combined model was used in many COVID-19 drug repurposing researches (Anwar .2020; Pahikkala.[2015](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666); Zeng .2019). DeepDR (Anwar .2020), a network-based DL model was used for silico drug repurposing. DeepDR uses heterogenous networks to understand different high stage features using deep auto encoder. The AUC obtained by using the model was 90.08% surpassing all other ML and DL methods. DeepPurpose (Kojima .2020), kGCN (Ramsundar .2019), DeepChem (Y. Shi .2020), and D3Targets-2019-nCoV (Bung .2021) were the other models which used ML for drug repurposing. Bung et al. (Tang .2020), used deep learning technology to develop a modern chemical framework called SARS-CoV-2 3 CL pro. A RL based RRN method was established for protease inhibitor molecules. From CheMBL database, 2515 SMILES protease inhibitors molecules were trained. The SMILES sequence was used as location, symbol and time series. The 3CLpro structure was used to embed these molecules with a minimum force to select the participants in the development of anti-SARS-CoV-2 based on visualization. Tang et al. (Moskal .2020), pursued the development of anti-COVID-19 drugs using 3CLpro with a structure that was similar to SARS-CoV. Leading SARS-CoV-2 3CLpro compounds were developed using Q-learning network called ADQN-FBDD. 284 molecules of SARS-CoV-2 3CLpro inhibitors were collected and classified using an optimized BRICS technique in order to understand the SARS-CoV-2 3CLpro targeted library. The ADQN-FBDD model trained each target and classified the corresponding molecules. 47 compounds were identified, that could be used as anti-SARS-CoV2-drugs with the structure-based optimization policy. Table 8 shows ML and DL methods for drug development of COVID-19.

**Managing and controlling the pandemic**

Pandemic control deals with curbing the spread of this viral infection to the maximum extent. AI methods can be efficiently used in curbing the spread of the deadly virus. ML can be used to reduce the infection rate in many ways which are discussed below.

**Contact tracing**

Pandey .(Hashem .2020), developed an AI application called “Washkaro” that can track people who spread misinformation about COVID-19. The interactions of people are automatically recorded and all wrong information is monitored using a self-assessment tool. Prevention and detection of the spread of the pandemic were proposed in (Lee .2020). This machine learning model was used to forecast pandemics and epidemics which might occur further.

**2.3.10. Identification of COVID-19 cases**

Continuous inspection and monitoring are required to control the spread of this virus. In (Simsek & Kantarci, 2020), an agent which uses a call-based model is used to actively monitor and identify the people in South Korea and Japan. For optimal mobilization, mobile assessment agents were deployed in (Mashamba-Thompson & Crayton, 2020). In (Rao & Vazquez, 2020), a combination of blockchain and AI was used as a tracking system for COVID-19. A phone-based survey was used to identify cases in (Muthya .2020). Infections among several communities can be measured using particular indicators according to this paper. In (Pirouz .2020), symptoms like conjunctivitis, diarrhoea, nausea etc. were used to monitor different groups of the population.

**COVID-19 testing**

Testing is also very important in the battle against COVID-19. In (B. Wang .2020), ML was used to understand the correlation between the swab tests and the positive cases. The authors state that a correlation between swab tests and confirmed cases, placid cases, fatal cases and death rates definitely exists. ML along with graph theory is used to reduce the infection rate using an optimization problem called minimum weight vertex problem (Mowery .2020). In (Sangiorgio & Parisi, 2020), COVID-19 infection is tested using lateral flow assays (LFA) along with ML. The authors reported that the pairing of LFA’s obtained a better classification.

**Risk assesment**

In (Yan .2020), ANN-based algorithms were used for many urban districts to perform risk assessments. Community based risk assessment was proposed in (Uddin .2020). Based on the state, country and city, a specific risk was measured.

**Social distancing**

One of the most effective methods in curbing the virus is social distancing. A CNN based algorithm (Yang .2020) was proposed to monitor the people in public places. The algorithm checked whether the people had symptoms like cough and fever. In (Punn .2020), people who do not follow social distancing were caught using a machine vision algorithm. A surveillance video-monitored social distancing norm with the help of DL (Khandelwal .2020). In (Qin & Li, 2020), AI along with machine vision was used to detect social distancing violations among workers. Class projective geometry and deep learning techniques were combined. Wearing facemasks reduces the chances of getting infected by this virus. In (Sesagiri Raamkumar .2020), people were checked for face masks using image processing and machine learning. In (Johnstone, 2020), deep learning models were used to classify context on Facebook. The model was studied to understand the public perception towards face masks and social distancing.

**7.6. Helping various organisations**

In (Nawaz .2020), ML was used to help various charities to organize funds during the pandemic. Library services were optimized along with resource allocation (Bandyopadhyay & DUTTA, 2020). A neural network model is presented (T. Chen .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) to detect fraud transactions. The education system has been completely changed during this pandemic. New platforms have emerged to help students continue their education. Education satisfaction was modelled using ANN (Li .2020).

**Pyschological effect**

A lot of people have experienced unemployment, helplessness, hopelessness and social isolation during the pandemic. Many of them have had little or no psychological support. People’s mental health during COVID-19 is explained in (Hossain & McKyerand P. Ma, 2020). ML was used to manage the psychological effects in (Khattar .2020). In (Prakash Jha .2020), ML was used to understand the mental health of students of India. In (Ćosić .2020), ML was used to determine various factors which impact mental health of people during the lockdown period. Bayesian Network was used to understand the key factors. In (Cosic .2020), high distress among people was calculated which directly led to chronic mental health disorder. Mental health during pandemic was also studied in (Ehrlich & Ghimire, 2020).

**Other impacts**

The entire sporting industry has changed during the pandemic. Lack of fans, bio-bubbles, social distancing, etc. has impacted the performance of players. ML was used to study the same in (Rahman .2020). A clustering framework was used to study the economic impact during lockdown in (Ghamizi .2020). Economic crisis of various countries was studied using data driven models in (Tarrataca .2021). Epidemiological forecasting was also made with the aid of a deep learning method. A natural regression model was proposed (Allam & Jones, 2020) which predicted the effect of COVID-19 in Brazil. The idea of developing smart cities to tackle the effects of the pandemic have also been discussed. In (Minetto .2021), ML methods are used to understand various outbreaks and how smart cities can be built to increase data sharing during the pandemic. Human activities have changed drastically during COVID-19. To understand this change in pattern, a deep learning algorithm was used (D. Wang .2020). CNN is combined with strategic location sampling. A DNN was proposed in (Asheghi .[2020](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)) to understand various transportation trends during the pandemic.

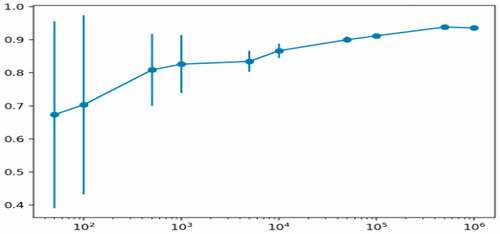
**2.4.5. Sensitivity analysis for developing ml models**

For predictive modelling, performance normally improves with the dataset quantity. This is dependent on the datasets and the model used, but it frequently means that using more data might result in greater performance and that discoveries made using smaller datasets to predict model performance often translates to using bigger datasets. The issue is that the relationship is unknown for every given dataset and model, and it might not exist for some datasets and models. Furthermore, if a co-relation exists, there may be a point or points of diminishing returns where adding additional data does not increase model performance or where datasets are too tiny to adequately capture a model’s potential at a wider scale. To address these difficulties, sensitivity analysis can be used to quantify the relationship between dataset size and model performance (May Tzuc .2019). We may then analyse the findings of the research and decide how much data is sufficient and how small a dataset should be in order to accurately estimate performance on larger datasets. One method to approach this challenge is to conduct a sensitivity analysis and see how the performance of your model on your dataset changes as you add or subtract data May Tzuc . 2019). This could entail comparing the same model with different dataset sizes to see if there is a link between dataset size and performance or if there is a point of diminishing returns. In most cases, especially for nonlinear models, there is a high correlation between training dataset size and model performance. As the dataset size grows, the model’s performance improves to a point and the predicted variance of the model decreases. This relationship between your model and dataset might be useful for a variety of reasons, including:

* Evaluation of models.
* Finding a more suitable model.
* Making the decision to collect more information.

We can quickly test a large number of model configurations on a smaller sample of the dataset and be confident that the results will generalize in a predictable manner to a larger training dataset. This may allow you to evaluate many more models and configurations than you would otherwise be able to given the time constraints, allowing you to find a model that performs better overall. We can also extrapolate and estimate the model’s expected performance to much bigger datasets, and determine whether it’s worth the effort or price to collect more training data. These results could be used to test additional model configurations and even different design types. The risk is that the different models will perform differently with more or less data, therefore it’s a good idea to run the sensitivity analysis again with a different model to make sure the association persists. Alternatively, you may run the study again using a variety of other model types. [Figure 5](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666" \l "f0005) compares the model performances with different dataset sizes.

Figure 5. Line plot with error bars of dataset size vs. model performance Brownlee, [n.d.](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666))

[](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)

When the probability estimates of a data point belonging to a class is critical, the model can be calibrated. Calibration is the comparing of a system’s actual output with its expected output Ranjan, [n.d.](https://www.tandfonline.com/doi/full/10.1080/23311916.2021.1958666)). In this process, we strive to enhance our model so that the probability predicted has a distribution and behavior that is similar to the probability distribution and behavior observed in training data. To get the best results out of AI, it is very important to perform sensitive analysis and calibration for the design models.

**2.4.6. Challenges and future directions**

This section elaborates the various challenges while using ML to fight COVID-19. Additionally, we give clear directions for future research.

**2.4.7. Challenges**

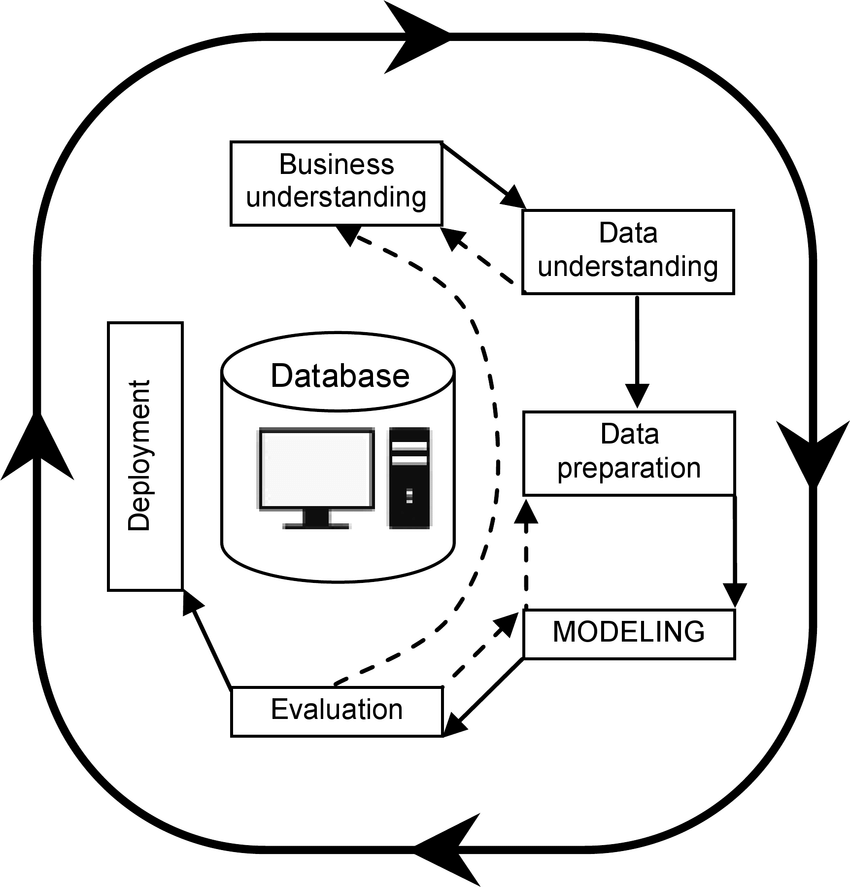
A lot of ML applications in COVID-19 research face many challenges, such as noisy data, scarcity and unavailability of large training data, rumors, limited awareness in interdisciplinary studies (computer science and medicine), security and data privacy issues and many more. The following have been discussed here.

* **Regulation**—As COVID-19 spreads and the number of confirmed cases grow, several measures, such as lockdown and social distancing, have been proposed to contain it. During a pandemic, authorities have a critical role in developing legislation and policies that can encourage citizens, researchers, scientists, business owners, medical centres, technology behemoths, and significant enterprises to participate in COVID-19 prevention.
* **Unavailability of large datasets**—Many AI-based DL approaches rely on large-scale training data, such as medical imaging and other applications. However, due to the rapid growth of COVID-19, there are insufficient datasets accessible for AI. In practice, evaluating training samples takes time and may necessitate the use of qualified medical specialists.
* **Noisy datasets and rumors**—The difficulties arise from relying on the internet with no significant adjustments. Thus, huge quantities of audio material and misleading reports regarding COVID-19 have been reported on various online sites. However, AI-based ML and DL algorithms appear to be slow when judging and filtering audio and erroneous data. Furthermore, when noisy data is used, the findings of AI-based ML and DL algorithms become skewed. This problem reduces the usefulness and performance of AI-based approaches, particularly in pandemic forecasting and spread assessment.
* **Lack of combined knowledge in medicine and computer fields**—Most AI-based ML and DL researchers have a computer science background, but substantial specialization in medical imaging, bioinformatics, virology and many other relevant domains are also necessary for the COVID-19 fight. To deal with COVID-19, it is therefore necessary to organize the participation of experts from diverse sectors and to include data from multiple investigations.
* **Data Privacy—**In the age of big data and AI, the cost of obtaining personal data is incredibly low. In response to public health challenges such as COVID-19, many governments intend to collect a variety of personal information, including ID numbers, contact numbers, personal data, and patient history. A concern that should be addressed is how to efficiently protect the privacy and human rights during AI-based research and analysis.
* **Unstructured Data (Numerical, text, image)—**It might be difficult to work with confusing and inaccurate information in text descriptions. A large amount of information from various sources may be false. Furthermore, excess data makes it impossible to extract meaningful information.

**Research Methodology**

The methodology is a formalized approach to the system development lifecycle (SDLC). It is also part of any analysis or research that is used to find out the type of data that is to be maintained, how it is found and how it is recorded. for this research, the cross-industry standard process for data mining (CRISP-DM) will be adopted.

CRISP-DM methodology is a standard cyclic framework used for the analysis of all data mining, predictive analysis, and data science projects (Pamanalytics, 2020). It helps one to plan, organize, and implement a data science project without stress. CRISP- MD methodology has six phases that naturally describe the data science life cycle as shown in figure 25 below. They have cycle iterations with the following steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.



**Figure 3.1: CRISP-DM Diagram of Data Science Life Cycle(Researchgate)**

The methodology flow begins with a proper understanding of the success criteria of the problem intending solving. The problem needs must be well-defined to know what the project can accomplish. The next step is to ascertain the type of data that is needed. Secondary data collection techniques to extract data from the official website of the Nigeria Center for Disease and Control. The site will be explored and searched for related data.

Data preparation refers to all the activities required to construct the final dataset out of the initial raw data. It involves the pre-processing of the dataset by cleaning the dataset and also checking for missing values and replacing them. About 85% of the success of a data science project is centered on the data preparation process. This phase takes the longest time because involves the following steps: selection of data, cleaning the data, constructing the data, integrating the data, and formatting the data. The next thing to do is feature extraction from the dataset which involves extracting useful information from the dataset. The dataset is classified into attributes and targets to suit the purpose of the research. The dimensionality reduction approach of feature extraction methods is studied in detail and computed in this proposed approach to make comparisons.

After data preparation comes modeling. A particular model is adopted that will fit into the context of the project. This phase is divided into four steps, namely:

1. Selection of modeling techniques.
2. Generate a test design based on the model technique by splitting the into training, test, and validation set.
3. Build the model(s) which can be executed by a few lines of code.
4. Iteratively assess the models until the best is selected.

During the evaluation, different classification algorithms will then be developed to train on the dataset and the one with the best accuracy will use to predict the system. The final phase is dependent on the nature of the requirements. It defines how stakeholders of the new system will access the result. The result can be as simple as generating a report or as complex as implementing a data mining process repeatedly in an organization.

**Benefits of the CRISP-DM methodolog**y

1. CRISP-DM is well-suited for machine learning projects
2. It has a uniform standard template that is made up of six phases thereby making it very simple.
3. A loose CRISP-DM is flexible as it adopts a principle that attracts the benefits of other methodologies like agile and waterfall.
4. This method is cost-effective as it integrates several processes.
5. CRISP-DM can be implemented in any data science project irrespective of its domain.
6. CRISP-DM is a cross-industry platform as such it is generally accepted.

**Data Gathering Techniques**

Based on the nature of this research, the method of data gathering technique that was adopted is secondary data. This involves the use of data that already exists from established sources. The research adopted secondary data because of the unavailability of covid-9 records which can be used for the development of the model. Therefore, online data was the source of our data. The data for the analysis was scrapped from the official website of the Nigeria Center for Disease and Control using the python beautiful soup module which is the only source to get a suitable dataset for the analysis. For this analysis, quantitative data was adopted because the dataset for the analysis comprises numerical details and information.

Analysis of the Existing System

The existing system of information gathering, tracking, and retrieving in Nigeria for the management of the COVID-19 outbreak is handled manually and unprofessionally. There was no platform for people to report suspected persons with COVID-19. That is to say, reporting of cases of COVID-19 is done mainly in isolation centers, and hence most times law enforcement agencies are not able to identify the person that entered the country either because of a bribe or no proper technology to do so. In transportation companies in Nigeria today, people still travel from state to state without proper medical checks for COVID-19, and even if the medical checks are done, fast reporting or call to the COVID-19 emergency response team are not fast contacted, therefore, the new system is designed to fill in this gap by providing a platform for private, public organizations, people to provide information on suspected persons with COVID-19 after medical examinations conducted. Furthermore, because the recording of infected or suspected persons of COVID-19 is done manually made accurate data storing, retrieving, tracking, and updating is difficult and time wastage sometimes leads to data loss. "There are no how restrictions to accurate data to researchers and technologist will not affect the rate of technological advancement or products development that could assist the medical personnel's and government fight the coronavirus". It was because of access to this information that made china design numerous technological devices that helped manage the disease in their country and other countries to determine if they are going for a total lockdown or partial lockdown based on the perdition of the system. Therefore, the researchers urge the Nigerian government to build a data bank or warehouse (if not in existence) and make it open for researchers and technologists to use in building software and hardware or techniques that could help in managing this deadly disease.

Furthermore, there are quite a several machine learning analyses that have been performed on covid-19 cases in Nigeria but all these models are based on mathematical modeling and also no predicting model was developed using the Nigeria covid-19 dataset from the official website. There is always a need for the development of a model capable of predicting the incidence of covid-19 based on the information gathered, this is what led to the development of the new system.

**Analysis of the new system**

The new system is an enhancement of the existing system. As such, all the features of the existing system were enhanced and improved upon by using a different and more robust algorithm in the analysing covid-19 dataset obtained from the official website of NCDC. The system was modeled to extract a dataset from the website and use the extracted dataset to build a model capable of analyzing the covid-19 cases in Nigeria. The process involves in developing the new system starts with the extraction of data from the site, till the final prediction was performed. These processes include data scrapping, data collection, Data exploration, model selection, model training, model evaluation, and finally model deployment.

**Data Scrapping and Data Collection**

Data collection is a crucial step in the machine learning process. Machine learning algorithms require large amounts of data to learn and make predictions. This data can come from a variety of sources, such as text files, databases, or web scraping.

The data for this analysis was automatically extracted from websites. It involves making HTTP requests to a website's server, downloading the HTML of the web page, and then parsing that HTML to extract the data you're interested in.

Python is a popular language for web scraping, with libraries like Beautiful Soup and Scrapy making it easy to extract information from HTML and XML documents. For this research, the extraction was done using a python library called Beautifulsoup.

**Data Exploration**

This stage involves analyzing and understanding a dataset before building a machine-learning model. It is an important step in the machine learning process as it helps to identify patterns, relationships, and trends in the data. This information can then be used to inform the model selection and feature engineering process. For this research, the following common techniques are used during the exploration stage using various libraries in python such as pandas, NumPy, and matplotlib. These libraries provide functions to perform data exploration and visualization.

1. Visualizing the data using plots and charts, such as histograms, scatter plots, and heatmaps, to gain insights into the distribution, relationship, and outliers of the features.
2. Summarizing the data by calculating statistics, such as mean, median, and standard deviation, to get a general understanding of the data.
3. Identifying correlations between features to understand how they relate to one another.
4. Checking for missing or duplicate values in the data, as these can impact the performance of the model.

**Model Selection and Training**

Model selection and training are two important steps in the machine-learning process.

Model selection refers to the process of choosing the appropriate algorithm or architecture for the problem at hand. This decision is based on the characteristics of the data and the problem to be solved. For this analysis, the polynomial regression model is considered for the problem because of the distributed nature of the dataset. Linear regression couldn't fit perfectly into the dataset because the parameter is not linearly proportional to each other. There are various libraries in Python used in achieving this such as sci-kit-learn. After the selection of the appropriate algorithm, the next stage is the training of the model. Training a machine learning model involves feeding it a large amount of data and adjusting the model's parameters so that it can make accurate predictions on new data. The process of training a model is also known as fitting the model. The first step of the training was splitting the data into testing and training data. The training data is used to optimize the model's parameters by minimizing a loss function, which measures the difference between the model's predictions and the true values of the output variable. This process is typically done using an optimization algorithm such as gradient descent.

3.6.4 Model Evaluation

Evaluating a polynomial regression model is similar to evaluating any other type of regression model. For this polynomial regression, we use visual evaluation methods. We plotted the predicted values against the actual values and observe how closely the predictions match the actual values. We can also plot the residuals, which are the differences between the predicted and actual values, and check if there is any pattern in the residuals, indicating a problem with the model.

Another way uses to evaluate a polynomial regression model is by comparing it with other models such as a linear regression model. This was done by comparing the performance of the models using the evaluation metrics mentioned above.

**Data Flow Diagram (DFD) of the New System**

The Data Flow Diagram of the new system is shown below

Feature extraction

Polynomial Regression

Result

ML Algorithms

Data classification

Machine learning model

dataset

Data pre-processing

Figure 3.2 Data flow diagram for the proposed system

### **Proposed System Design**

The design of this model was done using object–oriented modeling. The user actions, system actions, class diagrams, and sequence diagrams used to explain the flow of the algorithm design from data dataset preparation to the final model design are explained in this chapter.

#### **Interface Design**

The Graphical User Interface enables the attraction of users to the system. The part of the model, in which the user interacts, is the graphical user interface generated by the Jupiter Notebook which is viewed through the browser. Other visible entities in the model are the plots and graphs generated by matplotlib. The final deployed application is based on a web browser for viewing and input into the system.

**Sequence Diagram**

This is a UML tool that helps to know the flow of work to be done and how to accomplish it. The dataset is processed to remove inconsistencies in the data after which the data is trained using a machine learning algorithm on the train data and the performance of the algorithm is tested using the test data to generate the result of the prediction.

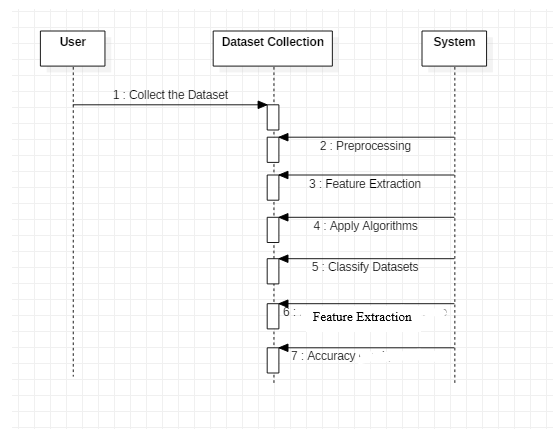


Figure 3.3: Sequence Diagram

### **Activity Diagram**

An activity diagram is a flow chart to represent the flow from one event to another. The event can be described as an operation of the system; therefore, the control flow is drawn from one operation to another. This flow can be parallel, sequential, branched, or concurrent. The flow of information for the model is described below

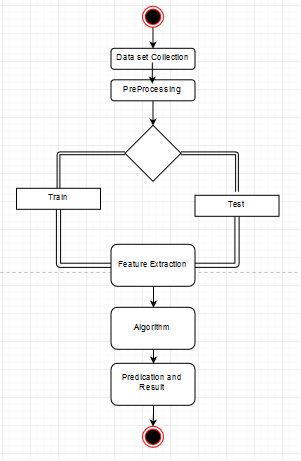


Figure 3.4: Activity Diagram

The activity diagram above illustrate the process involves in the development of the proposed system.

References

# References

1. Wang, D.; Hu, B.; Hu, C.; Zhu, F.; Liu, X.; Zhang, J.; Wang, B.; Xiang, H.; Cheng, Z.; Xiong, Y.; . Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA* **2020**, *323*, 1061–1069. [[CrossRef](http://dx.doi.org/10.1001/jama.2020.1585)] [[PubMed](http://www.ncbi.nlm.nih.gov/pubmed/32031570)]
2. Shen, K.; Yang, Y.; Wang, T.; Zhao, D.; Jiang, Y.; Jin, R.; Zheng, Y.; Xu, B.; Xie, Z.; Lin, L.; . Diagnosis, treatment, and prevention of 2019 novel coronavirus infection in children: Experts’ consensus statement. *World J. Pediatrics* **2020**, *16*, 223–231. [[CrossRef](http://dx.doi.org/10.1007/s12519-020-00343-7)] [[PubMed](http://www.ncbi.nlm.nih.gov/pubmed/32034659)]
3. Lauer, S.A.; Grantz, K.H.; Bi, Q.; Jones, F.K.; Zheng, Q.; Meredith, H.R.; Azman, A.S.; Reich, N.G.; Lessler, J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Intern. Med.* **2020**, *172*, 577–582. [[CrossRef](http://dx.doi.org/10.7326/M20-0504)] [[PubMed](http://www.ncbi.nlm.nih.gov/pubmed/32150748)]
4. Backer, J.A.; Klinkenberg, D.; Wallinga, J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance* **2020**, *25*, 2000062. [[CrossRef](http://dx.doi.org/10.2807/1560-7917.ES.2020.25.5.2000062)]
5. Remuzzi, A.; Remuzzi, G. COVID-19 and Italy: What next? *Lancet* **2020**, *395*, 1225–1228. [[CrossRef](http://dx.doi.org/10.1016/S0140-6736(20)30627-9)]
6. Tuite, A.R.; Ng, V.; Rees, E.; Fisman, D. Estimation of COVID-19 outbreak size in Italy. *Lancet Infect. Dis.*

**2020**, *20*, 537. [[CrossRef](http://dx.doi.org/10.1016/S1473-3099(20)30227-9)]

1. Tuite, A.R.; Bogoch, I.I.; Sherbo, R.; Watts, A.; Fisman, D.; Khan, K. Estimation of coronavirus disease 2019 (COVID-19) burden and potential for international dissemination of infection from Iran. *Ann. Intern. Med.* **2020**, *172*, 699–701. [[CrossRef](http://dx.doi.org/10.7326/M20-0696)]
2. Zhuang, Z.; Zhao, S.; Lin, Q.; Cao, P.; Lou, Y.; Yang, L.; He, D. Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: A modelling analysis based on overseas cases and air travel data. *Int. J.* *Infect. Dis.* **2020**, *94*, 29–31. [[CrossRef](http://dx.doi.org/10.1016/j.ijid.2020.03.019)]
3. Ceylan, Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ.* **2020**, *729*, 138817. [[CrossRef](http://dx.doi.org/10.1016/j.scitotenv.2020.138817)]
4. Batista, M. Estimation of the final size of the COVID-19 epidemic. *medRxiv* **2020**. [[CrossRef](http://dx.doi.org/10.1101/2020.02.16.20023606)]
5. Alboaneen, D.; Pranggono, B.; Alshammari, D.; Alqahtani, N.; Alyaffer, R. Predicting the Epidemiological Outbreak of the Coronavirus Disease 2019 (COVID-19) in Saudi Arabia. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4568. [[CrossRef](http://dx.doi.org/10.3390/ijerph17124568)] [[PubMed](http://www.ncbi.nlm.nih.gov/pubmed/32630363)]
6. Papastefanopoulos, V.; Linardatos, P.; Kotsiantis, S. COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population. *Appl. Sci.* **2020**, *10*, 3880. [[CrossRef](http://dx.doi.org/10.3390/app10113880)]
7. Hernandez-Matamoros, A.; Fujita, H.; Hayashi, T.; Perez-Meana, H. Forecasting of COVID19 per regions using ARIMA models and polynomial functions. *Appl. Soft Comput.* **2020**, *96*, 106610. [[CrossRef](http://dx.doi.org/10.1016/j.asoc.2020.106610)] [[PubMed](http://www.ncbi.nlm.nih.gov/pubmed/32834798)]
8. Qin, L.; Sun, Q.; Wang, Y.; Wu, K.F.; Chen, M.; Shia, B.C.; Wu, S.Y. Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2365. [[CrossRef](http://dx.doi.org/10.3390/ijerph17072365)] [[PubMed](http://www.ncbi.nlm.nih.gov/pubmed/32244425)]
9. Li, C.; Chen, L.J.; Chen, X.; Zhang, M.; Pang, C.P.; Chen, H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Eurosurveillance* **2020**, *25*, 2000199. [[CrossRef](http://dx.doi.org/10.2807/1560-7917.ES.2020.25.10.2000199)]
10. Li, J.; Xu, Q.; Cuomo, R.; Mackey, T.; Purushothaman, V. Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. *JMIR Public Health Surveill.* **2020**, *6*, 8700. [[CrossRef](http://dx.doi.org/10.2196/18700)]
11. Chen, F.-M.; Feng, M.-C.; Chen, T.-C.; Hsieh, M.-H.; Kuo, S.-H.; Chang, H.-L.; Yang, C.-J.; Chen, Y.-H. Big data integration and analytics to prevent a potential hospital outbreak of COVID-19 in Taiwan. *J. Microbiol.* *Immunol. Infect.* **2020**. [[CrossRef](http://dx.doi.org/10.1016/j.jmii.2020.04.010)]
12. Zhou, C.; Su, F.; Pei, T.; Zhang, A.; Du, Y.; Luo, B.; Cao, Z.; Wang, J.; Yuan, W.; Zhu, Y. COVID-19: Challenges to GIS with big data. *Geogr. Sustain.* **2020**, *1*, 77–87. [[CrossRef](http://dx.doi.org/10.1016/j.geosus.2020.03.005)]
13. Wieczorek, M.; Siłka, J.; Woz´niak, M. Neural network powered COVID-19 spread forecasting model.

*Chaos Solitons Fractals* **2020**, *140*, 110203. [[CrossRef](http://dx.doi.org/10.1016/j.chaos.2020.110203)]

1. Pinter, G.; Felde, I.; Mosavi, A.; Ghamisi, P.; Gloaguen, R. COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach. *Mathematics* **2020**, *8*, 890. [[CrossRef](http://dx.doi.org/10.3390/math8060890)]
2. Bragazzi, N.L.; Dai, H.; Damiani, G.; Behzadifar, M.; Martini, M.; Wu, J. How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3176. [[CrossRef](http://dx.doi.org/10.3390/ijerph17093176)]
3. Scopus. Analyze Search Results, TITLE-ABS-KEY (“COVID-19” or “CoViD-19” or “Hubei Pneumonia” or “Novel Corona Virus”). Available online: [https://08105qd69-1104-y-https-www-scopus-com.mplbci.ekb.](https://08105qd69-1104-y-https-www-scopus-com.mplbci.ekb.eg/term/analyzer.uri?sid=38aae2fabdccac31270f9b430acb9525&origin=resultslist&src=s&s=TITLE-ABS-KEY%28%22COVID-19%22%2BOR%2B%22CoViD-19%22%2BOR%2B%22Hubei%2BPneumonia%22%2BOR%2B%22Novel%2Bcorona%2Bvirus%22%29&sort=plf-f&sdt=b&sot=b&sl=84&count=8069&analyzeResults=Analyze%2Bresults&txGid=fc5eb31e2f64a965b06d73db82b71d2e) [eg/term/analyzer.uri?sid=38aae2fabdccac31270f9b430acb9525&origin=resultslist&src=s&s=TITLE-ABS-](https://08105qd69-1104-y-https-www-scopus-com.mplbci.ekb.eg/term/analyzer.uri?sid=38aae2fabdccac31270f9b430acb9525&origin=resultslist&src=s&s=TITLE-ABS-KEY%28%22COVID-19%22%2BOR%2B%22CoViD-19%22%2BOR%2B%22Hubei%2BPneumonia%22%2BOR%2B%22Novel%2Bcorona%2Bvirus%22%29&sort=plf-f&sdt=b&sot=b&sl=84&count=8069&analyzeResults=Analyze%2Bresults&txGid=fc5eb31e2f64a965b06d73db82b71d2e) [KEY%28%22COVID-19%22+OR+%22CoViD-19%22+OR+%22Hubei+Pneumonia%22+OR+%22Novel+](https://08105qd69-1104-y-https-www-scopus-com.mplbci.ekb.eg/term/analyzer.uri?sid=38aae2fabdccac31270f9b430acb9525&origin=resultslist&src=s&s=TITLE-ABS-KEY%28%22COVID-19%22%2BOR%2B%22CoViD-19%22%2BOR%2B%22Hubei%2BPneumonia%22%2BOR%2B%22Novel%2Bcorona%2Bvirus%22%29&sort=plf-f&sdt=b&sot=b&sl=84&count=8069&analyzeResults=Analyze%2Bresults&txGid=fc5eb31e2f64a965b06d73db82b71d2e) [corona+virus%22%29&sort=plf-f&sdt=b&sot=b&sl=84&count=8069&analyzeResults=Analyze+results&](https://08105qd69-1104-y-https-www-scopus-com.mplbci.ekb.eg/term/analyzer.uri?sid=38aae2fabdccac31270f9b430acb9525&origin=resultslist&src=s&s=TITLE-ABS-KEY%28%22COVID-19%22%2BOR%2B%22CoViD-19%22%2BOR%2B%22Hubei%2BPneumonia%22%2BOR%2B%22Novel%2Bcorona%2Bvirus%22%29&sort=plf-f&sdt=b&sot=b&sl=84&count=8069&analyzeResults=Analyze%2Bresults&txGid=fc5eb31e2f64a965b06d73db82b71d2e) [txGid=fc5eb31e2f64a965b06d73db82b71d2e](https://08105qd69-1104-y-https-www-scopus-com.mplbci.ekb.eg/term/analyzer.uri?sid=38aae2fabdccac31270f9b430acb9525&origin=resultslist&src=s&s=TITLE-ABS-KEY%28%22COVID-19%22%2BOR%2B%22CoViD-19%22%2BOR%2B%22Hubei%2BPneumonia%22%2BOR%2B%22Novel%2Bcorona%2Bvirus%22%29&sort=plf-f&sdt=b&sot=b&sl=84&count=8069&analyzeResults=Analyze%2Bresults&txGid=fc5eb31e2f64a965b06d73db82b71d2e) (accessed on 23 May 2020).
4. Wang, W.; Tang, J.; Wei, F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *J. Med Virol.* **2020**, *92*, 441–447. [[CrossRef](http://dx.doi.org/10.1002/jmv.25689)]
5. Garnier-Crussard, A.; Forestier, E.; Gilbert, T.; Krolak-Salmon, P. Novel Coronavirus (COVID-19) Epidemic: What Are the Risks for Older Patients? *J. Am. Geriatr. Soc.* **2020**, *68*, 939–940. [[CrossRef](http://dx.doi.org/10.1111/jgs.16407)]
6. Wei, M.; Yuan, J.; Liu, Y.; Fu, T.; Yu, X.; Zhang, Z.-J. Novel Coronavirus Infection in Hospitalized Infants Under 1 Year of Age in China. *JAMA* **2020**, *323*, 1313–1314. [[CrossRef](http://dx.doi.org/10.1001/jama.2020.2131)]
7. Ivorra, B.; Ferrández, M.R.; Vela-Pérez, M.; Ramos, A.M. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Commun. Nonlinear. Sci. Numer. Simul.* **2020**, *88*, 105303. [[CrossRef](http://dx.doi.org/10.1016/j.cnsns.2020.105303)]
8. Shen, C.; Chen, A.; Luo, C.; Zhang, J.; Feng, B.; Liao, W. Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study. *J. Med.* *Internet Res.* **2020**, *22*, 9421. [[CrossRef](http://dx.doi.org/10.2196/19421)]
9. Ayyoubzadeh, S.M.; Ayyoubzadeh, S.M.; Zahedi, H.; Ahmadi, M.; Niakan Kalhori, R.S. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. *JMIR Public Health Surveill.* **2020**, *6*, 8828. [[CrossRef](http://dx.doi.org/10.2196/18828)]
10. Hashim, H.A.; El-Ferik, S.; Lewis, F.L. Neuro-adaptive cooperative tracking control with prescribed performance of unknown higher-order nonlinear multi-agent systems. *Int. J. Control* **2019**, *92*, 445–460. [[CrossRef](http://dx.doi.org/10.1080/00207179.2017.1359422)]
11. Hashim, H.A.; El-Ferik, S.; Lewis, F.L. Adaptive synchronisation of unknown nonlinear networked systems with prescribed performance. *Int. J. Syst. Sci.* **2017**, *48*, 885–898. [[CrossRef](http://dx.doi.org/10.1080/00207721.2016.1226984)]
12. Araújo, R.d.A.; Nedjah, N.; Oliveira, A.L.I.; Meira, S.R.d.L. A deep increasing–decreasing-linear neural

network for financial time series prediction. *Neurocomputing* **2019**, *347*, 59–81. [[CrossRef](http://dx.doi.org/10.1016/j.neucom.2019.03.017)]

1. Eltoukhy, A.E.E.; Wang, Z.X.; Chan, F.T.S.; Fu, X. Data analytics in managing aircraft routing and maintenance staffing with price competition by a Stackelberg-Nash game model. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *122*, 143–168. [[CrossRef](http://dx.doi.org/10.1016/j.tre.2018.12.002)]
2. Eltoukhy, A.E.E.; Wang, Z.X.; Chan, F.T.S.; Chung, S.H.; Ma, H.; Wang, X.P. Robust Aircraft Maintenance Routing Problem Using a Turn-Around Time Reduction Approach. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, 1–14. [[CrossRef](http://dx.doi.org/10.1109/TSMC.2019.2937648)]
3. Chung, S.H.; Ma, H.L.; Chan, H.K. Cascading Delay Risk of Airline Workforce Deployments with Crew Pairing and Schedule Optimization. *Risk Anal.* **2017**, *37*, 1443–1458. [[CrossRef](http://dx.doi.org/10.1111/risa.12746)]
4. Misiunas, N.; Oztekin, A.; Chen, Y.; Chandra, K. DEANN: A healthcare analytic methodology of data envelopment analysis and artificial neural networks for the prediction of organ recipient functional status. *Omega* **2016**, *58*, 46–54. [[CrossRef](http://dx.doi.org/10.1016/j.omega.2015.03.010)]
5. Tsai, T.-H.; Lee, C.-K.; Wei, C.-H. Neural network based temporal feature models for short-term railway passenger demand forecasting. *Expert Syst. Appl.* **2009**, *36*, 3728–3736. [[CrossRef](http://dx.doi.org/10.1016/j.eswa.2008.02.071)]
6. Magesh, S.; Niveditha, V.R.; Rajakumar, P.S.; RamMohan, S.R.; Natrayan, L. Pervasive computing in the context of COVID-19 prediction with AI-based algorithms. *Int. J. Pervasive Comput. Commun.* **2020**, in press.
7. Tuli, S.; Tuli, S.; Tuli, R.; Gill, S.S. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet Things* **2020**, *11*, 100222. [[CrossRef](http://dx.doi.org/10.1016/j.iot.2020.100222)]
8. WorldOmeter. COVID-19 CORONAVIRUS PANDEMIC. Available online: [https://www.worldometers.info/](https://www.worldometers.info/coronavirus/) [coronavirus/](https://www.worldometers.info/coronavirus/) (accessed on 6 May 2020).
9. Our World in Data. The COVID-19 Pandemic Slide Deck. Available online: [https://ourworldindata.org/the-](https://ourworldindata.org/the-covid-19-pandemic-slide-deck) [covid-19-pandemic-slide-deck](https://ourworldindata.org/the-covid-19-pandemic-slide-deck) (accessed on 30 April 2020).
10. World Bank. World Bank Open Data. Available online: <https://data.worldbank.org/>(accessed on 7 May 2020).
11. United Nations Development Programme. Human Development Reports. Available online: [http://hdr.undp.](http://hdr.undp.org/en/data) [org/en/data](http://hdr.undp.org/en/data) (accessed on 8 May 2020).
12. Agency, T.N. Country of Four Seasons: Iran A World inside A Country. Available online: https://www. tasnimnews.com/en/news/2016/10/03/1202507/country-of-four-seasons-iran-a-world-inside-a-country (accessed on 20 August 2020).
13. Menezes, J.M.P.; Barreto, G.A. Long-term time series prediction with the NARX network: An empirical evaluation. *Neurocomputing* **2008**, *71*, 3335–3343. [[CrossRef](http://dx.doi.org/10.1016/j.neucom.2008.01.030)]
14. Xie, H.; Tang, H.; Liao, Y.-H. Time series prediction based on NARX neural networks: An advanced approach. In Proceedings of the 2009 International Conference on Machine Learning and Cybernetics, Baoding, China, 12–15 July 2009; pp. 1275–1279.
15. Boussaada, Z.; Curea, O.; Remaci, A.; Camblong, H.; Mrabet Bellaaj, N. A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation. *Energies* **2018**, *11*, 620. [[CrossRef](http://dx.doi.org/10.3390/en11030620)]
16. Marcjasz, G.; Uniejewski, B.; Weron, R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. *Int. J. Forecast.* **2019**, *35*, 1520–1532. [[CrossRef](http://dx.doi.org/10.1016/j.ijforecast.2017.11.009)]
17. Rai, A.; Upadhyay, S. The use of MD-CUMSUM and NARX neural network for anticipating the remaining useful life of bearings. *Measurement* **2017**, *111*, 397–410. [[CrossRef](http://dx.doi.org/10.1016/j.measurement.2017.07.030)]
18. Eltoukhy, A.E.E.; Wang, Z.X.; Chan, F.T.S.; Chung, S.H. Joint optimization using a leader–follower Stackelberg game for coordinated configuration of stochastic operational aircraft maintenance routing and maintenance staffing. *Comput. Ind. Eng.* **2018**, *125*, 46–68. [[CrossRef](http://dx.doi.org/10.1016/j.cie.2018.08.012)]
19. Plaehn, D.; Horne, J. A regression-based approach for testing significance of “just-about-right” variable penalties. *Food Qual. Prefer.* **2008**, *19*, 21–32. [[CrossRef](http://dx.doi.org/10.1016/j.foodqual.2007.06.003)]
20. Kim, Y.-S.; Yum, B.-J. Robust design of multilayer feedforward neural networks: An experimental approach.

*Eng. Appl. Artif. Intell.* **2004**, *17*, 249–263. [[CrossRef](http://dx.doi.org/10.1016/j.engappai.2003.12.005)]

1. Packianather, M.S.; Drake, P.R.; Rowlands, H. Optimizing the parameters of multilayered feedforward neural networks through Taguchi design of experiments. *Qual. Reliab. Eng. Int.* **2000**, *16*, 461–473. [CrossRef]
2. Tabrizi, S.; Ghodsypour, S.H.; Ahmadi, A. Modelling three-echelon warm-water fish supply chain: A bi-level optimization approach under Nash–Cournot equilibrium. *Appl. Soft Comput.* **2018**, *71*, 1035–1053. [[CrossRef](http://dx.doi.org/10.1016/j.asoc.2017.10.009)]
3. Menebo, M.M. Temperature and precipitation associate with Covid-19 new daily cases: A correlation study between weather and Covid-19 pandemic in Oslo, Norway. *Sci. Total Environ.* **2020**, *737*, 139659. [[CrossRef](http://dx.doi.org/10.1016/j.scitotenv.2020.139659)]
4. S¸ahin, M. Impact of weather on COVID-19 pandemic in Turkey. *Sci. Total Environ.* **2020**, *728*, 138810.

[[CrossRef](http://dx.doi.org/10.1016/j.scitotenv.2020.138810)]

1. Tosepu, R.; Gunawan, J.; Effendy, D.S.; Ahmad, L.O.A.I.; Lestari, H.; Bahar, H.; Asfian, P. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Sci. Total Environ.* **2020**, *725*, 138436. [[CrossRef](http://dx.doi.org/10.1016/j.scitotenv.2020.138436)]
2. Eltoukhy, A.E.E.; Chan, F.T.S.; Chung, S.H.; Niu, B. A model with a solution algorithm for the operational aircraft maintenance routing problem. *Comput. Ind. Eng.* **2018**, *120*, 346–359. [[CrossRef](http://dx.doi.org/10.1016/j.cie.2018.05.002)]
3. Eltoukhy, A.E.E.; Chan, F.T.S.; Chung, S.H.; Niu, B.; Wang, X.P. Heuristic approaches for operational aircraft maintenance routing problem with maximum flying hours and man-power availability considerations. *Ind. Manag. Data Syst.* **2017**, *117*, 2142–2170. [[CrossRef](http://dx.doi.org/10.1108/IMDS-11-2016-0475)]
4. Milman, O. Coronavirus in America: Why the US has struggled to tackle a growing crisis. *The Guardian*.

24 March 2020. Available online: https://www.theguardian.com/us-news/2020/mar/24/coronavirus-america-

[donald-trump-crisis](https://www.theguardian.com/us-news/2020/mar/24/coronavirus-america-donald-trump-crisis) (accessed on 7 July 2020).

1. BBC NEWS. Coronavirus: Brazil Records third-highest Covid-19 Infection Level. Available online: [https:](https://www.bbc.com/news/world-latin-america-52719391)

[//www.bbc.com/news/world-latin-america-52719391#](https://www.bbc.com/news/world-latin-america-52719391) (accessed on 7 July 2020).

1. Collyns, D. Peru’s coronavirus response was ‘right on time’—So why isn’t it working? *The Guardian*. 20 May 2020. Available online: https://www.theguardian.com/global-development/2020/may/20/peru-[coronavirus-lockdown-new-cases](https://www.theguardian.com/global-development/2020/may/20/peru-coronavirus-lockdown-new-cases) (accessed on 7 July 2020).
2. Npr News. COVID-19 Numbers Are Bad in Ecuador. The President Says The Real Story Is Even Worse. Available online: https://www.npr.org/sections/goatsandsoda/2020/04/20/838746457/covid-19-numbers-are- [bad-in-ecuador-the-president-says-the-real-story-is-even-wo](https://www.npr.org/sections/goatsandsoda/2020/04/20/838746457/covid-19-numbers-are-bad-in-ecuador-the-president-says-the-real-story-is-even-wo) (accessed on 8 June 2020).