



# Epidemiological challenges in pandemic coronavirus disease (COVID-19): Role of artificial intelligence

Abhijit Dasgupta<sup>1</sup> | Abhisek Bakshi<sup>2</sup> | Srijani Mukherjee<sup>1</sup> | Kuntal Das<sup>1</sup> |  
Soumyajeet Talukdar<sup>1</sup> | Pratyayee Chatterjee<sup>1</sup> | Sagnik Mondal<sup>1</sup> |  
Puspita Das<sup>1</sup> | Subhrojit Ghosh<sup>1</sup> | Archisman Som<sup>1</sup> | Pritha Roy<sup>1</sup> |  
Rima Kundu<sup>1</sup> | Akash Sarkar<sup>1</sup> | Arnab Biswas<sup>1</sup> | Karnelia Paul<sup>3</sup> |  
Sujit Basak<sup>4</sup>  | Krishnendu Manna<sup>5</sup> | Chinmay Saha<sup>6</sup> |  
Satinath Mukhopadhyay<sup>7</sup> | Nitai P. Bhattacharyya<sup>7</sup> | Rajat K. De<sup>8</sup> 

<sup>1</sup>Department of Data Science, School of Interdisciplinary Studies, University of Kalyani, Kalyani, Nadia, West Bengal, India

<sup>2</sup>Department of Information Technology, Bengal Institute of Technology, Kolkata, West Bengal, India

<sup>3</sup>Department of Biotechnology, University of Calcutta, Kolkata, West Bengal, India

<sup>4</sup>Department of Physiology and Biophysics, Stony Brook University, Stony Brook, New York, USA

<sup>5</sup>Department of Food and Nutrition, University of Kalyani, Kalyani, Nadia, West Bengal, India

<sup>6</sup>Department of Genome Science, School of Interdisciplinary Studies, University of Kalyani, Kalyani, Nadia, West Bengal, India

<sup>7</sup>Department of Endocrinology and Metabolism, Institute of Post Graduate Medical Education and Research and Seth Sukhlal Karnani Memorial Hospital, Kolkata, West Bengal, India

<sup>8</sup>Machine Intelligence Unit, Indian Statistical Institute, Kolkata, West Bengal, India

## Correspondence

Rajat K. De, Machine Intelligence Unit,  
Indian Statistical Institute, 203 B.T. Road,  
Kolkata 700108, West Bengal, India.  
Email: [rajat@isical.ac.in](mailto:rajat@isical.ac.in)

**Edited by:** Mehmed Kantardzic,  
Associate Editor and Witold Pedrycz,  
Editor-in-Chief

## Abstract

World is now experiencing a major health calamity due to the coronavirus disease (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus clade 2. The foremost challenge facing the scientific community is to explore the growth and transmission capability of the virus. Use of artificial intelligence (AI), such as deep learning, in (i) rapid disease detection from x-ray or computed tomography (CT) or high-resolution CT (HRCT) images, (ii) accurate prediction of the epidemic patterns and their saturation throughout the globe, (iii) forecasting the disease and psychological impact on the population from social networking data, and (iv) prediction of drug-protein interactions for repurposing the drugs, has attracted much attention. In the present study, we describe the role of various AI-based technologies for rapid and efficient detection from CT images complementing quantitative real-time polymerase chain reaction and immunodiagnostic assays. AI-based technologies to anticipate the current pandemic pattern, prevent the spread of disease, and face mask detection are also discussed. We inspect how the virus transmits depending on different factors. We investigate the deep learning technique to assess the affinity of the most probable drugs to treat COVID-19.

Abhijit Dasgupta and Abhisek Bakshi contributed equally to this study.

This article is categorized under:

Application Areas > Health Care

Algorithmic Development > Biological Data Mining

Technologies > Machine Learning

#### KEYWORDS

deep learning, drug affinity, EHR, social media, x-ray/CT/HRCT

## 1 | INTRODUCTION

Infection with severe acute respiratory syndrome coronavirus clade 2 (SARS-CoV-2) causes devastating pneumonia COVID-19. Symptoms of the disease are similar to SARS-CoV and Middle East respiratory syndrome coronavirus infections. The worldwide public health is badly affected by the recent pneumonia outbreak caused by SARS-CoV-2. The common symptoms of coronavirus disease (COVID-19) are fever and dry cough at the onset of illness, but the most distinctive symptom of the patient is suffering from respiratory problems. The health of many patients in the intensive care has been worsened in a short phase of time, leading to death because of respiratory failure (Y.-C. Li, Bai, et al., 2020). Thus, speedy and accurate detection of SARS-CoV-2 is essential for appropriate management and preventing the spread. Presently, there is no specific treatment for the disease. Thus, it is very necessary to establish a quick standard diagnostic test for the detection of COVID-19 to avert consequent secondary spread among individuals.

Reverse transcriptase mediated quantitative real-time polymerase chain reaction (qRT-PCR) is a gold standard test for the diagnosis of COVID-19 infection (Udugama et al., 2020). RT-PCR is expensive and cannot be used for diagnosis of large number of people as screening cum detection test. A number of external factors can also affect qRT-PCR test results. These include sampling operations (sample collection and handling), sample stability (decrease in viral counts due to sample handling time), sampling time (different period of the disease development), availability, and performance of detection kits. Thus, qRT-PCR technique is difficult to use at population level and is unable to control epidemic COVID-19 at very initial phase.

Preliminary screening can be done through the presence of six symptoms (fever, cough, respiratory distress, loss of taste, loss of smell, and conjunctivitis) associated with COVID-19, followed by normal oxygen saturation level measurement using simple pulse oximeter. Normal oxygen saturation for most individuals is 94%–100%, whereas COVID-19 pneumonia patients may have oxygen saturation as low as 50%. Direct confirmatory qRT-PCR test should be performed for symptomatic cases. It will help in differentiating between COVID-19 and other flu cases. Remaining individuals in the community, that is, presymptomatic and asymptomatic, may further have to undergo a rapid antibody testing. Those who are positive for rapid antibody test, confirmatory qRT-PCR test should be performed for them. Negative cases should be repeated once again for rapid antibody test after 7 days and so. Thus, to detect high-risk COVID-19 pneumonia patients, artificial intelligence (AI)-based image processing technology on chest x-ray images can take immense role in identifying those patients in early treatment with oxygen. In addition, improved AI-based image processing of chest computed tomography (CT)/high-resolution CT (HRCT) images can take immense role in identifying COVID-19 patients for early quarantine/isolation/treatment leading to prevention of transmission in near future.

AI can play an important role in predicting the pattern of worldwide outbreak of COVID-19 and its saturation time (Maier & Brockmann, 2020; Roosa et al., 2020). It has been observed in various databases about COVID-19<sup>1,2,3</sup> that different countries have inconsistent infection and death rate patterns. Thus, it immediately arises some striking questions:

1. What are the variation of important factors (yet to be explored) responsible for various outbreak patterns in different countries?
2. How are viruses transmitted so fast?
3. What are the possible technologies or measures or principles to prevent the spread?
4. Is there any role of social media data in forecasting the pattern of disease spread?
5. How can COVID-19 patients be detected rapidly to take quick quarantine/isolation/treatment step managing huge number of affected people worldwide?

6. What are the immediate most probable drugs available in market to treat COVID-19 before the invention of the most appropriate ones?

We are going to find, in this article, suitable answers to the aforementioned questions. We will see how AI can help experimental biologist, doctors, and pharmacologists for addressing the above questions.

In this context, we have manually curated the existing literature to assist the scientists in determining the track of investigations in the future. Since the pandemic has started, scientists have investigated numerous approaches with the help of various AI techniques such as machine learning, deep learning, natural language processing, cognitive computing, and many more to arrest the growth of the infection. We have scrutinized various investigations that have employed several AI techniques to anticipate the status of the pandemic all over the world.

Researchers have employed several supervised machine learning frameworks (Borghi et al., 2021; Car et al., 2020; V. K. Gupta et al., 2021) to anticipate the trajectory of the pandemic in terms of the number of deaths, infection cases, and many more. On the other side, several unlabeled data set has influenced the researchers to incorporate unsupervised machine learning techniques. For instance, recently, researchers have proposed a clustering algorithm (Zarikas et al., 2020) based on time-series data associated with COVID-19 infection. Here, the authors have uncovered the similarity among countries concerning active cases and active cases per population. More recently, Vahabi et al. (2021) have scrutinized the mortality to incidence ratio (MIR) to develop a model-based clustering algorithm. Here, the researchers have incorporated the longitudinal generalized estimating equations (GEE) model (Diggle et al., 2002) to determine the risk factor. Afterward, they have developed a latent growth mixture model (LGMM) to identify the risk factors of several clusters separately. Such identification may help the policymakers to determine the allocation of various medical resources and other emergencies. However, extracting useful features from high-dimensional data sets often motivates the researchers to employ the autoencoder-based deep learning model in the framework. Dastider et al. (2021) have employed an autoencoder along with a convolutional neural network (CNN)-based DenseNet-201 model to anticipate the severity of COVID-19.

More recently, a similar type of investigation has been presented by Khozeimeh et al. (2021). Here, the authors have predicted the chance of survival of COVID-19 patients based on clinical information. Recent observations on COVID-19 diagnosis have revealed a significant role of various medical imaging methods such as x-ray and CT scans (G. Li, Fan, et al., 2020). In recent years, classification tasks (Zhou, 2020) using CNNs based on deep learning techniques have demonstrated promising results. For example, a recent investigation (Xu et al., 2020) has employed multiple CNN models to assess the severity of COVID-19 infection from CT image samples. Covent (L. Li, Qin, et al. 2020) has reported a similar type of investigation based on ResNet architecture. Scientists have attempted to predict the nature of the pandemic from several aspects throughout the globe. Respective authorities in all the countries have updated numerous disease-related information such as the number of infected cases, number of hospital admissions, and much more regularly. All such information can generate an extensive time-series data set.

To exploit the data set for several predictive tasks, recurrent neural network (RNN)-based framework can be the most suitable choice. For example, a group of scientists has incorporated RNN-based framework (Lee et al., 2021) on the electronic health record (EHR) data set to anticipate the severity of the disease concerning ventilation, tracheostomy, or death of a COVID-19-infected patient. The foremost benefit of the technique is that a medical practitioner can predict the future hazard without any immediate pathological parameters. In this context, a recent investigation (Arun Kumar et al., 2021) has reported 60 days ahead forecast of the pandemic in terms of confirmed cases, recovered cases, and fatalities in country wise using gated recurrent units (GRUs) and long short-term memory (LSTM) cells along with RNN. Arora et al. (2020) have predicted a similar type of task with the help of a modified version of RNN, termed deep LSTM, convolutional LSTM, and bi-directional LSTM.

In addition, we have reported numerous previous observations on the correlation of various searching keywords with the pandemic in Section 4.1.2. For example, AI-based Natural Language Processing techniques (Ayyoubzadeh et al., 2020; Qin et al., 2020; Singh et al., 2018) could highly be beneficial for early precautions to arrest the pandemic. The unstoppable nature of the disease has introduced panic, fear, and anxiety among the population. Scientists have perceived that several social policies such as isolation, social distancing, work from home, and many more may inhibit the devastating nature of the disease (Javed et al., 2020). A recent report (C. Wang, Pan, et al., 2020) has estimated that more than 50% of the Chinese population have experienced several psychiatric stresses such as anxiety and depression. Moreover, Flesia et al. (2020) have investigated an ML model to identify the people of Italy having a greater risk of developing mental imbalance. Here, the researchers have utilized socio-demographic circumstances along with a set of questionnaires indicating various psychological hallmarks. Similarly, Roma et al. (2020) have attempted to reveal the

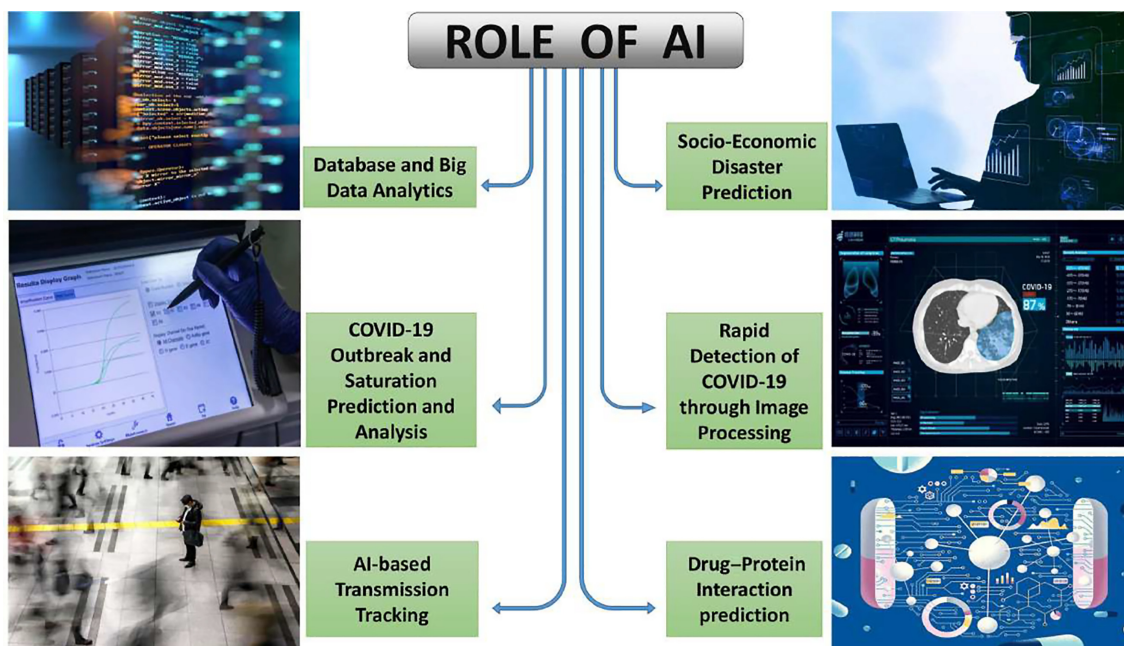


FIGURE 1 The role of artificial intelligence (AI) in addressing epidemiological challenges in pandemic COVID-19

behavioral change among the population with the help of a set of questionnaires. Now, to achieve an efficient antiviral treatment, AI-based predictive model may help in exploring possible effective drugs for COVID-19 using available drug databases.<sup>4</sup> We are going to illustrate more on this topic in the upcoming sections.

Before addressing the aforementioned questions related to current epidemic challenges, we are going to discuss different databases (Section 2) which are essential these challenges in various perspectives. In the following section (Section 3), AI-based image processing techniques to detect COVID-19 preliminary before confirmatory test will be explained elaborately. Next section (Section 4) will deal with different AI-based models predicting the outbreak pattern, transmission, and saturation of the epidemic worldwide. Subsequently, role of social media in understanding the pandemic pattern (Section 4.2) will be elaborated. This section will also include different types of transmission with some influencing factors (Section 4.3). The succeeding section (Section 5) will present AI-based models for future drug discovery and vaccination. Recently, the policymakers of various countries have perceived the importance of face masks to arrest the pandemic. Now, reducing the transmission rate necessitates strict monitoring of wearing a face mask. We will discuss this topic in Section 6. We have identified several major challenges by manually curating existing literature. In this regard, we will discuss probable solutions to the circumstances in Section 7. Finally, the article will conclude with a discussion and elaboration of future direction (Section 8). In this regard, Figure 1 depicts the role of AI in response to epidemiological challenges in pandemic COVID-19.

## 2 | DATABASES FOR EPIDEMIOLOGICAL STUDIES

Recently, Governments of several countries have decided to take immediate action to obtain a solution against such growing pandemic caused by the coronavirus. They have started involving the data science community and shared information related to the disease worldwide. On the other hand, World Health Organization (WHO) and other well-known data science communities too have shared various data sets related to COVID-19. Thus, data science community should concentrate on how to explore these publicly available databases to predict some important features of the disease. We have summarized some key characteristics of the publicly available data sets in Supporting Information to restrict the size of the article. Here, we are going to discuss the kinds of predictive analysis possible using these databases.

All such data sets have motivated us to visualize the pattern of the disease with respect to various types of parameters worldwide. Here, we can apply the machine learning technique to contemplate how the rate of infection is related to global health security index (GHSI) score which determines how different nations have the risk of pandemic/epidemic attack. In addition, the score tells us the appropriate precautions before the pandemic activities become out of control in a large geographical region. Consequently, we can make cluster of various countries according to their GHSI score as high vulnerability, low vulnerability, and medium vulnerability zones. Such clustering may be highly significant for the government to reserve sufficient medical equipment to fight against the novel coronavirus.

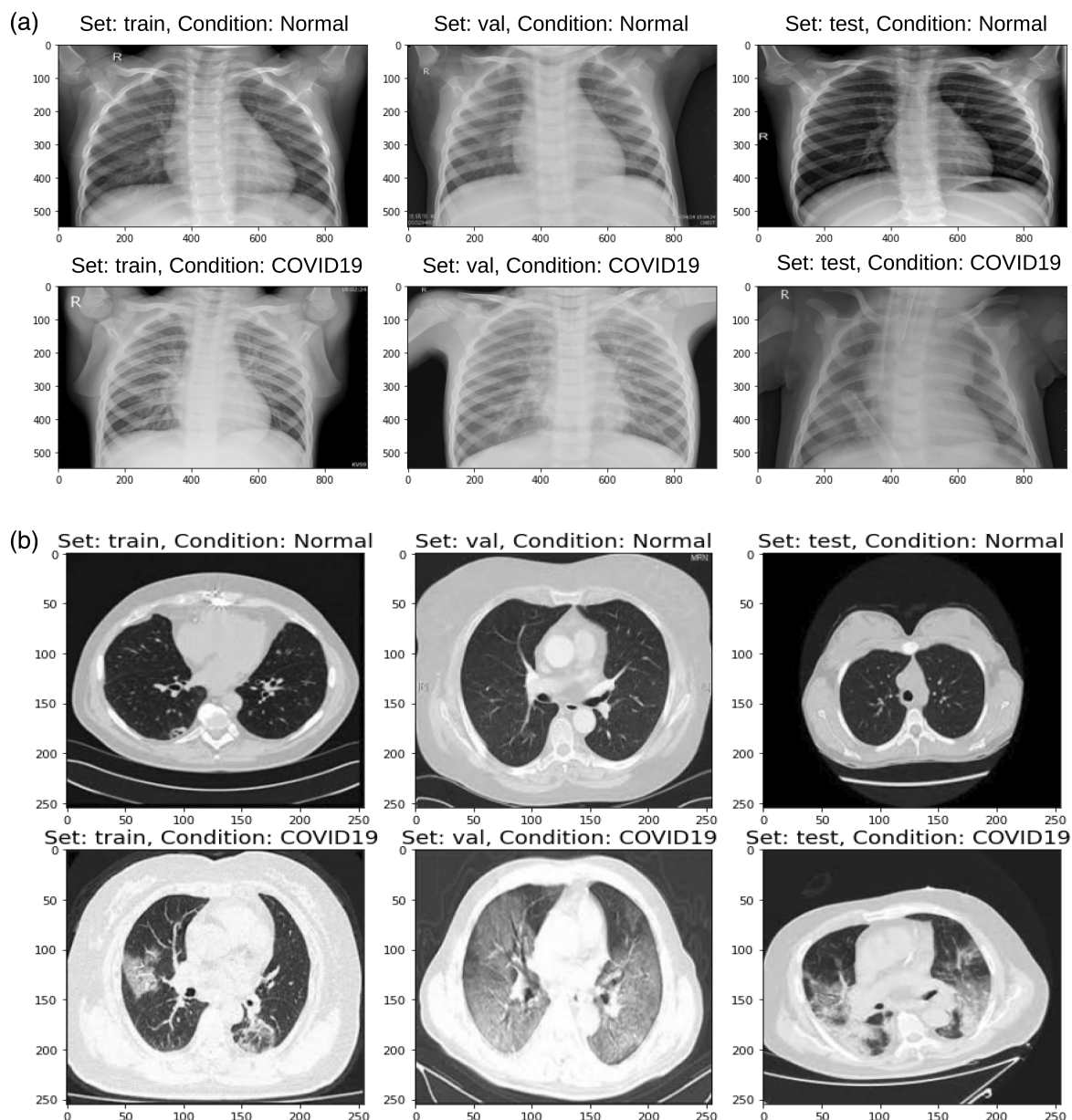
### 3 | AI-BASED DETECTION

Rapid diagnosis is a main part of health care, because it provides details about the health problem of a patient and informs about successive health care decisions. The diagnostic process is a mutual and complex activity that involves clinical analysis and gives information to determine health problem of a patient. In recent decades, modern methods have been developed for the rapid diagnostics of different infections. The main advantages of rapid detection techniques are the possibility of earlier intervention and faster focused action to potential problems, but also improved throughput of analysis. Here, we are going to discuss how AI-based image processing techniques using chest x-ray/CT/HRCT images can help in cost-effective preliminary screening before confirmatory detection by qRT-PCR.

Chest x-ray/CT/HRCT imaging technique can be a useful option to detect suspected COVID-19 pneumonia patients with high risk at early stage. In this regard, efficient image processing techniques based on the concept of machine learning may accurately detect COVID-19 pneumonia confirmed cases. A recent study (D. Wang, Hu, et al., 2020) has uncovered the shape of COVID-19 as bilateral distribution of patchy shadows and ground glass opacity. In addition, to extract the significant features from chest CT scan images, deep learning approach (Gómez et al., 2019), such as CNN, may be employed. CNN has already shown improved performance to recognize the nature of pulmonary nodules (Choe et al., 2019). Subsequently, it has also accurately uncovered the position of polyps from colonoscopies and many more (Kermany et al., 2018; Negassi et al., 2020; P. Wang, Xiao, et al., 2018). In this pandemic, AI-based image processing techniques (Fang et al., 2020; Xie et al., 2020) have already shown good efficiency to detect COVID-19 using chest CT scan of affected individuals at very early phase of the disease. In comparison with qRT-PCR, it has been found (Fang et al., 2020) that approximately 98% of COVID-19 cases have initially been detected accurately by noncontrast chest CT scan, while qRT-PCR has become successful in detecting 71% of the such initial cases. In addition, it has been shown that 3% of 167 patients had negative qRT-PCR data for the virus despite chest CT findings of viral pneumonia (Xie et al., 2020). It shows that the use of chest CT can be significant to decrease false-negative data (Xie et al., 2020) during the early phase of COVID-19.

To explore the efficacy of deep learning-based image processing techniques in detecting COVID-19 at early stage, we have implemented a CNN-based existing method.<sup>5</sup> Here, mostly chest x-ray images (with a very few chest CT/HRCT images) of COVID-19-affected patients have been used from publicly available databases.<sup>6,7,8</sup> A total of 5412 chest images (mostly x-ray with very few CT/HRCT) for training (75 of normal individuals and 5337 of COVID-19-affected patients) and 42 such images for test (21 normal images and 21 COVID-19-positive images) have been considered. Figure 2a depicts a few samples of training and test chest x-ray images. This method shows test accuracy of 78.57% with precision of 75%, sensitivity of 85.71%, specificity 71.43%, and F1-score of 80. In addition, Figure 2b depicts a few successful classifications of chest CT/HRCT images of normal individuals and COVID-19 patients based on the aforementioned methodology. However, test accuracy is not significant due to lack of chest CT/HRCT images of COVID-19 patients for training the model. Future improvement in algorithm with more number of chests CT/HRCT images for training leads to more rapid and accurate detection of COVID-19 pneumonia patients with high risk at very early stage.

Similar to the aforesaid method, a number of recent investigations (X. Chen et al., 2020; Islam et al., 2020; Khan et al., 2020; Ozturk et al., 2020) have used CNN to identify COVID-19-positive patients from the CT scan/x-ray images. Islam et al. (2020) has employed a hybrid architecture consisting of a series of CNN layers followed by LSTM and fully connected feed forward neural network. Such architecture has improved test accuracy to become more than 90%. Thus, it is clear that on availability of more training samples of CT/HRCT images of all possible respiratory syndromes along with more efficient and accurate deep learning-based algorithms can rapidly detect not only high-risk COVID-19 pneumonia patients but also other respiratory diseases in upcoming days. In addition, presymptomatic CT/HRCT/x-ray



**FIGURE 2** (a) Samples of x-ray images of both normal individuals and COVID-19 patients, used for training, validation and test. A deep learning-based image processing method<sup>5</sup> shows test accuracy of 78.57% with precision of 75%, sensitivity of 85.71%, specificity 71.43%, and *F1*-score of 80 during rapid detection of COVID-19 pneumonia patients with high risk at very early phase. (b) Here, the prediction based on CT/HRCT images is depicted. Here, in the manuscript (Section 3), we have discussed more improved methods in this regard. CT, computed tomography; HRCT, high-resolution computed tomography

images can be used for early detection of COVID-19 patients based on deep learning-based techniques leading us a step ahead to fight against the current as well as upcoming pandemic. Here, extracting useful features from the CT/HRCT images is an important task to accurately detect the disease or normal ones.

#### 4 | OUTBREAK, SATURATION, TRANSMISSION, AND PREVENTION

Past few decades have experienced a lot of technological innovation in various scientific fields, particularly in medical science. The application of natural language processing (Afshar et al., 2019; Funk et al., 2020; S. Wang, Ren, et al., 2018), medical image processing (Lambin et al., 2017; Parmar et al., 2018), and deep learning techniques

(Thomas et al., 2019; Wei et al., 2020) are some examples of such advancement. The influence of AI has added a lot of potential in healthcare sector to anticipate the diagnosis in a faster way. Thus, it must have a potential role to fight against the current global health calamity due to pandemic COVID-19 (Huang et al., 2020).

Previous investigations (Ksiazek et al., 2003; Peiris et al., 2003) have claimed a close association of novel coronavirus with SARS (Poutanen et al., 2003). However, a recent study (Ting et al., 2020) has argued that the ultimate spread of COVID-19 may become more influencing over SARS in 2003. Thus, to prevent the outbreak of such a disease, it requires a set of appropriate guidelines. However, strictly imposing such guidelines on the common people throughout a large geographical area is a very hard task to accomplish. Here, AI can guide us by predicting different aspects of this pandemic based on noise free publicly available databases (as discussed earlier) about COVID-19. Even social networking data have played a crucial role in forecasting a number of important facts in health care sectors (Andersen et al., 2012; Moorhead et al., 2013; Rozenblum & Bates, 2013). In the following subsections, we are going to discuss how AI-based prediction on COVID-19 outbreak, saturation, transmission, and prevention can help us to take necessary steps against the epidemic.

## 4.1 | Outbreak and saturation

According to the WHO, COVID-19 has become pandemic from the second week of March, 2020. It is clear from worldwide research and reports of WHO that the exponential spread of this disease may continue. In this context, statistics and machine learning-based efficient models may be effective to estimate the growth of future affected people as well as its saturation time.

### 4.1.1 | Previous models predicting outbreak pattern of Ebola virus diseases

Here, we can look back some previously developed models, namely, a transmission model (Kraemer et al., 2019) based on human movement data to predict the spatial outbreak of Ebola virus diseases (EVD) in West Africa during 2014–2016. This model has used weekly probable and confirmed case data obtained from WHO. In addition, it has predicted human movements based on three definite mathematical formulations such as gravity model, radiation model, and adjacency network considering each district as a vertex. Based on the Poisson regression, the first two models have tried to predict the total number of individuals moving from one district to another depending on some parameters, such as, population size at origin and destination, and distance between two districts among others. The predicted human movement data have helped in estimating spatial spread of EVD with a certain speed as well as its saturation period. Similar kind of studies (D. N. Fisman et al., 2013; D. Fisman et al., 2014; Kelly et al., 2019; Roosa et al., 2020) have been found on predicting spread and saturation of EVD and SARS in the past. All these studies have estimated the best-fit model solution to the reported data using nonlinear least squares fitting. However, it has been claimed that the uncertainty of the models can be reduced with the availability of more data. In short, we can summarize that all such aforesaid models can be highly beneficial for designing a real-time decision-making system.

### 4.1.2 | AI-based model predicting the outbreak pattern of current pandemic

Based on the previous study (Anderson, 1991), there are various nonlinear ordinary differential equation-based models, such as susceptible-infected-removed (SIR) model (Di Giambardino & Iacoviello, 2017), susceptible-infected-recovered-deceased (SIRD) model (Matadi, 2014) and susceptible-exposed-infectious-removed model (Chowell, 2017), to analyze the spread of infection in a large population. These models mostly depict the movement of the population across various states, such as susceptible, infected, recovered, and deceased over time. Recently, Maier and Brockmann (2020) have proposed a more realistic model, entitled as SIR-X model, considering “containment” and “quarantine” states of symptomatic population. Such aforementioned models have considered several parameters remaining constant over time. However, during a pandemic situation such parameter values may vary because of complex behavioral changes among population. In addition, the data about infected population may highly deviate from the actual number due to inefficient documentation process. To overcome such challenges, recently, Calafiore et al. (2020) have introduced a parameter-varying modification of the SIRD model. However, identifying such parameters is both nontrivial and non-convex problems due to the fast growing data of infected, recovered, death cases day by day throughout the world.

To tackle the aforementioned problem, various deep learning techniques may be useful to anticipate the outbreak pattern of recent COVID-19 pandemic. In this context, we have found that an investigation (Poirier et al., 2020) has employed least absolute shrinkage and selection operator regression model for 2-days-ahead real-time prediction of COVID-19 spread in 32 Chinese provinces during February 3, 2020 to February 21, 2020. Similar investigation (Tiwari et al., 2020) can be found to predict the number of confirmed, recovered, and death cases 22-days-ahead in India. Here, the authors have utilized Kaggle database<sup>9</sup> provided by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. However, these models have not been verified to check their ability to determine the epidemic peaks in corresponding areas.

During current pandemic situation, prediction of the mortality rate may help the medical practitioners to decide the treatment strategy in early stage for an infected person. A recent investigation (Nemati et al., 2020) has depicted a comparative study of different statistical and machine learning techniques for survival analysis of 1182 patients in China. Here, gradient boosting survival model has shown highest accuracy among the others. Thus, the machine learning (ML)-based model can be used as the remote application for real-time prediction. However, worldwide laboratory findings and characteristics of infected cases are frequently changing day by day based on various time-dependent parameters, such as mobility of a person, previous health-related issues and many more. Such parameters can be taken care with the help of RNN-based deep learning techniques for more accurate prediction of the current pandemic outbreak pattern.

Considering the aforesaid time-dependent parameters, a recent investigation (Alakus & Turkoglu, 2020) has evaluated the performance of six ML-based models, including CNN, RNN, LSTM, two hybrid models CNN-LSTM, and CNN-RNN, to predict the COVID-19 outbreak. Among them, CNN-LSTM has been proved as the best one. Likewise, another investigation (Zeroual et al., 2020) has compared different ML-based models, including bidirectional LSTM (BiLSTM), GRUs, and variational autoencoder (VAE), among others, in terms of the performance measures (e.g., root mean square error [RMSE], mean absolute error [MAE], mean absolute percentage error, and root mean-squared logarithmic error) to forecast new COVID-19 cases in various countries. Here, VAE model has shown promising accuracy over other models despite lack of large training data set. Similar kind of work (P. Wang, Zheng, et al., 2020) has proposed a LSTM-based model using an update mechanism based on real-time data of COVID-19 cases as well as predicted confirmed cases for the next day.

In this context, it should be mentioned that asymptomatic patients (having no such observable symptoms) play a vital role in COVID-19 outbreak because such individuals are susceptible to spread the virus through close contacts. In this situation, “herd immunity” can be considered as the indirect protection from COVID-19. It can be achieved when the population become resistant against the virus either through vaccination or immunity developed due to previous infection. Thus, the researchers have become curious about the time and cost to achieve the herd immunity. Recently, an investigation (Kaushal et al., 2020) has presented a model, entitled as susceptible-asymptomatic-infected-removed (SAIR) model, to uncover the hidden asymptomatic population. According to this study, the herd immunity may be achieved after 10%–25% of the population getting infected. Another similar type of study (Hoque et al. 2020) has employed SIRD model to reveal the dynamics of the disease outbreak in Bangladesh. In addition, the authors have presented a clustering method based on age-group to predict the population having high probability to be infected. They have further claimed that the herd immunity threshold may be reduced to 31% considering some assumptions on the population having low to mediocre income.

## 4.2 | Prediction through social media

In the course of pandemic activities during COVID-19, the early prediction of growth of the disease is the major task. A group of scientists have employed a statistical approach (Dong et al., 2019) based on general global optimization algorithm (Cheng et al., 2011) to estimate the dynamics. In addition, it has been observed that internet-based query through social networking websites has played a crucial role in prediction of various prevalent incidents throughout the world.

### 4.2.1 | Natural language processing

Recently, C. Li, Chen, et al. (2020) have observed that the most searched words in various social networks and search engines, such as Google Trends, Baidu Index, and many more, are “coronavirus” and “pneumonia.” Such a high



correlation between the searched words and future circumstances has been noticed almost 2 weeks earlier than the newly confirmed cases as well as newly suspected cases in China have been reported. Analyzing the previous investigations, we can realize that the future effect of COVID-19 can be predicted with the help of an intelligent social networking data mining tool with respect to some specific keywords. It can anticipate the dynamics of the epidemic in the early stage helping the Government in preparing the medical facilities to resist such pandemic activity. Researchers (Qin et al., 2020) have collected various social media keywords, such as chest distress, coronavirus, and pneumonia among others, and feed through a number of machine learning models to predict new suspected/confirmed COVID-19 cases 10-days-ahead. Similar kind of study (Ayyoubzadeh et al., 2020) has employed linear regression and LSTM-based model to estimate the number of confirmed cases in Iran using Google trends website. Besides, a study (Arpaci et al., 2020) has analyzed a huge number of tweets collected over a week to elucidate the expression pattern of common people due to such pandemic situation. Here, AI-based data summarization technique using K-means clustering has indicated that the “Coronavirus” term has been dominating over other unigram terms in Twitter during the span of current pandemic. However, the term “COVID-19” has become more important in the later stage followed by other bigram keywords, such as “stay home” and “Coronavirus pandemic.”

As COVID-19 has become a highly dominating topic in social media, researchers (Singh et al., 2018) have shown that social media can play a crucial role in controlling the emergency situation as well as in sentiment analysis of people in response to this pandemic. In this context, an investigation (Chakraborty et al., 2020) has employed a fuzzy logic-based deep learning classifier to classify such sentiments as positive, negative, and neutral scores. However, the study does not attempt to analyze multilingual tweets. Consequently, the psychological effect of such pandemic situation can further be analyzed by emotional intelligence that uncovers any unwanted psychological state found particularly in senior citizens.

#### 4.2.2 | Psychological impact analysis

The rapid number of exponential global COVID-19 cases has initiated panic, fear, and anxiety among people. Such mental and physical health of the global population has been found to be directly proportional to current pandemic crisis. In this context, a recent investigation (Flesia et al., 2020) has revealed that 53.8% of Chinese population have experienced acute negative psychological effects, including anxiety, depression, and stress (W. Cao et al., 2020; Huremović, 2019; C. Wang, Pan, et al., 2020, among others) due to COVID-19 outbreak. In addition, the fear of self-isolation as well as social distancing has a significant impact on mental health in terms of frustration (Brooks et al., 2020; Fontanesi et al., 2020; Mazza et al., 2020). Thus, identification of people with greater risk of having mental imbalance is one of the foremost challenges.

Nowadays, researchers have investigated various machine learning techniques that can identify such group of people suffering from mental trauma during current pandemic situation. Recently, a ML-based model (Flesia et al., 2020) has become able to identify the people experiencing a significant risk of developing various psychological complications. Here, 1628 participants from Italy have been categorized as 2 types of individuals, suffering from high and low psychological stress, respectively, based on a set of questionnaires related to COVID-19 threats, psychological traits, socio-demographic factors, such as gender, age, education, and monthly household income among others. This model has revealed that female individuals are more likely to be associated with higher psychological stress. Besides, it has shown a significant correlation between psychological stress and family income. However, since the data have been collected only from Italy, the result should not be considered as general tendency of common people throughout the world. A similar type of ML-based investigation (Roma et al., 2020) has been carried out to predict the behavioral changes of population due to COVID-19 outbreak with the help of a number of psychological and psycho-social variables presented in the online questionnaires.

### 4.3 | Transmission

One of the major reasons behind such pandemic by the novel coronavirus is that it can spread in various ways. In addition, different factors can influence its transmission. One of the most important ways is respiratory droplets. Here, angiotensin-converting enzyme 2 and SARS-CoV receptor have been found in lungs and gastrointestinal tract of human body. Therefore, it spreads through respiratory droplets over a relatively close distance. The statistical analysis (Geller et al., 2012) on the experimental data shows that SARS-CoV remains infectious in suspension as well as in the temperature below 60°C.

The stability of the virus in various surfaces is also a great enhancer for sustainability of the disease in a large geographical region. Although airborne transmission of the virus has not been proven till date, but in accordance with a report of WHO, the transmission may occur by aerosol-generating medical procedure (Tran et al., 2012). A recent evidence (van Doremalen et al., 2020) has shown that SARS-CoV-2 remains infectious in aerosols (<5  $\mu\text{m}$ ), and its intensity reduces with respect to time in exponential manner as in the case of SARS-CoV. To estimate the decay rate, they have employed a Bayesian regression model elucidating that SARS-CoV-2 is more stable on plastic and stainless steel surfaces than on copper and card-board surfaces. Moreover, it exists up to 4 h on the copper surface, whereas in the case of plastic and stainless steel surfaces, it survives more than 72 h. Thus, with the capability to resist different environmental factors, the virus becomes more threatening in terms of transferring between contaminated hosts via the hands and surfaces.

A previous study (McMichael et al., 2008) has uncovered a significant correlation between environmental temperature and mortality rate of human population. In the same way, virus may need a specific temperature range (particularly, during winter) to survive. A recent investigation (Prata et al., 2020) has supported the claim. Prata et al. (2020) have employed a generalized additive model (GAM) for (sub)tropical regions, such as Brazil, to depict the inversely proportional relation between the number of confirmed cases and the temperature range from 16.8 to 27.4°C. However, this study does not prove that the confirmed cases will decrease if environmental temperature enhances greater than 25.8°C. Another investigation (Qi et al., 2020) has also utilized GAM for understanding the similar relationship between viral transmission and average temperature as well as relative humidity in 30 Chinese provinces. Thus, researchers have expected that transmission capability may be reduced during summer (Tobías & Molina, 2020). Scientists (A. Gupta et al., 2020) have also explored the correlation of COVID-19 transmission with long term climate, social, and meteorological factors in India. A positive association between death rate due to COVID-19 and daily temperature in Wuhan, China has been reported in a study (Y. Ma et al., 2020). Even transmission throughout the world has been found higher in particular areas with significantly low temperature of subtropical countries (Poole, 2020). Similarly, a study (Shi et al., 2020) has depicted a negative correlation between daily temperature in China and COVID-19 transmission.

Researchers have observed that infants, elderly people, and immunosuppressed patients are under high risk of getting affected by COVID-19 (Geller et al., 2012). In addition, the study has revealed that COVID-19 can promote many other diseases such as digestive dysfunctions, necrotizing enterocolitis in newborns, diarrhea, and other gastrointestinal symptoms associated with heart problems among others (Geller et al., 2012). According to an investigation (Y. Wu et al., 2020), people with blood group of “A” are more prone to COVID-19, whereas persons with blood group of “O” are less susceptible than the others. However, this finding is still to be experimentally verified.

According to a recent study (Ferretti et al., 2020), transmission can broadly be classified in four categories, such as, “symptomatic” (a direct transference from symptomatically infected individuals to a noninfected one), “presymptomatic” (direct transmission from an infected individual before he/she experiences noticeable symptoms), “asymptomatic” (a direct transmission from an infected individual who has never experienced any symptoms) and “environmental” transmission (transference through contamination without a typically attributable path or source). This study shows that presymptomatic contact has 45% probability to spread COVID-19, which differs significantly from the outbreak studies for previous SARS-CoV.

To arrest the pandemic of a viral disease, an in-depth investigation of every individual in a particular region is highly needed. Unfortunately, due to some unwanted behavior of a set people, the chance of blocking community transmission of the disease may become under threat. In most of the cases, it has been revealed that hiding actual information about the health status of the potential patients may be the foremost cause behind that. In this situation, a set of tactful questionnaires may help any health care professionals to identify the potential patients who may be at a risk of getting infected by the viral disease, like COVID-19. Here, AI can help in preventing the transmission using online surveys and contact tracing mobile app.

#### 4.3.1 | AI-based prevention: Online surveys

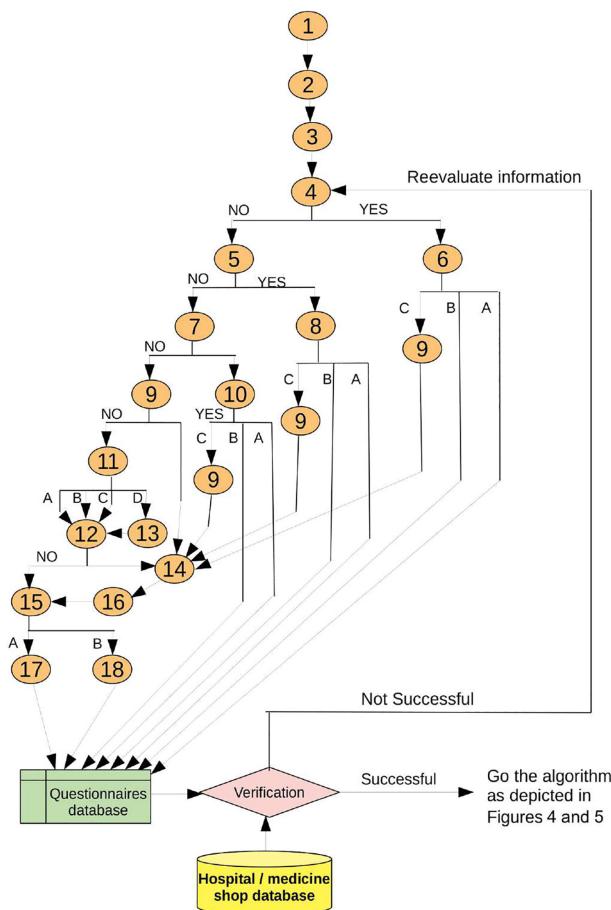
To get right information about the health condition of an individual, a hypothesis to identify such patients can be developed. Here, we may require a set of questionnaires that does not comprise any question related to specific disease symptom directly. In spite of that, the answers of the indirect questions signify most important information about the health status of the individuals. On the other hand, it can be assumed that the EHR of various patients is maintained at corresponding local hospitals. In addition, the availability of medicine purchase history of those patients from corresponding local medicine shop as well as from different online delivery companies is taken into account.

**TABLE 1** Set of sample questionnaires for maintaining transparency and availability of right information related to health of an individual

Serial number	Question	Question type	Option
1	What is your name?	Short answer	Not applicable
2	What is your contact number?	Short answer	Not applicable
3	What is your age?	Short answer	Not applicable
4	Do you have fever more than 100°?	Multiple choice	Yes, no
5	If you have fever more than 100° then how many days you are facing such problem?	Multiple choice	<1 week, >1 week and <2 week, >2 week
6	Do you have dry cough?	Multiple choice	Yes, no
7	Do you have body pain?	Multiple choice	Yes, no
8	If you have dry cough then how many days you are facing such problem?	Multiple choice	< 1 week. > 1 week and <2 week. > 2 week.
9	Have you taken any medicine in last 2 weeks?	Multiple choice	Yes, no
10	If you have body pain then how many days you are facing such problem?	Multiple choice	<1 week, >1 week and <2 week, > 2 week
11	How many days ago you have visited the hospital?	Multiple choice	>1 year, > 1 month, < 1 month, <2 week
12	Please tell me the list of medicines, doctor advised.	Descriptive	Not applicable
13	If you have visited the hospital what 2 weeks then what was the symptom?	Multiple choice	Cough, body pain, fever, all of the above
14	Please tell me the list of medicines, you have consumed.	Descriptive	Not applicable
15	From where you have bought your medicine?	Multiple choice	Yes, no
16	Did your doctor advised to take vitamin tablet?	Multiple choice	Yes, no
17	If you select local shop, please provide the details of the shop.	Descriptive	Not applicable
18	If you select online delivery, please provide the details of the company.	Descriptive	Not applicable

Collecting all such information from various sources, we can verify the answers of the questionnaires with the databases to ensure the correctness of the information. If the correctness percentage does not cross the threshold, we can reevaluate the data by asking similar questionnaires again. In Table 1, we have presented a set of sample questionnaires related to a set of symptoms of the disease caused by coronavirus. Subsequently, Figure 3 depicts the order of the questions to be asked to an individual. If an individual is found suspected to be COVID-19 patient based on his/her satisfactory correct health information collected by the proposed approach, the person may undergo with another survey to detect his/her severity in two ways:

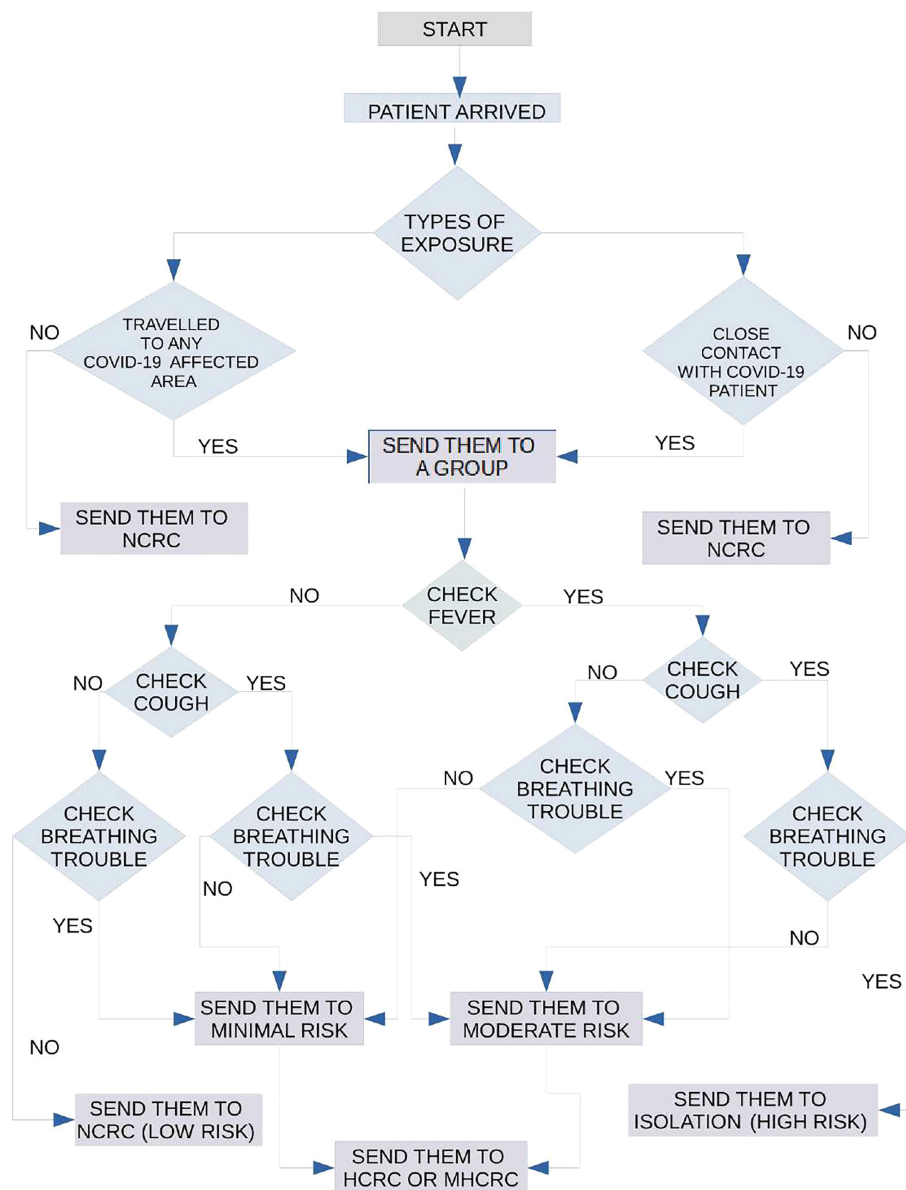
- *Online survey before treatment:* Online survey through particular websites, email, and mobile apps may be used to predict possible COVID-19-affected populations more quickly with the help of machine learning techniques. Recently, a group of scientists (Rao & Vazquez, 2020) has already done a mobile phone-based survey to collect address, age, gender, traveling information for the last 14 days, particulars about direct close contact with COVID-19-affected people or not and health condition among other details. Here, Figure 4 depicts flowchart of their algorithm to identify possible affected people using the aforementioned survey data. This algorithm may help us to identify actual patients suffering from COVID-19 and provide them early treatment/isolation/quarantine preferred to other suspected ones.
- *EHR-based survey:* It may be better to design a survey based on EHR using machine learning techniques to detect the severity of patients suffering from COVID-19 as well as risk of being affected by the viral disease. Here, EHR helps in finding any kind of association among patients and their disease(s), and treatment trajectory(s). Here, we try to give an idea of an EHR-based survey, which can be designed as follows.



**FIGURE 3** The figure illustrates the order of the questions to be asked. The answers should be verified with databases of local hospitals/ medical shops/online delivery companies to track false health information of an individual in a particular area

Let us assume that some regular medicines enhance the immunity of human to fight against COVID-19. They have taken the medicine just before the outbreak or take regularly. We also consider a positive correlation of getting affected by COVID-19 with some specific disease in their medical history. To illustrate the scenario, we have considered six categories of sample medicines, such as M1, M2, ..., M6. Subsequently, 10 various types of diseases, represented by D1, D2, ..., D10 are taken into consideration. Thereafter, we have presented a flow chart (Figure 5) that examines previous medicines consumed as well as diseases in medical history. In accordance with several combinations of these two considerations, we have partitioned the population into 14 disjoint groups (represented by  $G_i$ , where  $i = 1, 2, \dots, 14$ ). Finally, we have assessed the risk of getting affected by the coronavirus into five different statuses as depicted in Table 2: (i) “very high risk,” (ii) “high risk,” (iii) “moderate risk,” (iv) “low risk,” and (v) “no risk.” However, the data present in the EHR should be error free to obtain a satisfactory predictive result using the proposed AI-based algorithm.

Such huge information in EHR has influenced the researchers to develop AI-based predictive model in big data paradigm. Previous studies (Ben-Assuli et al., 2012; Paley et al., 2011; Silow-Carroll et al., 2011) have revealed that medical decision using EHR can be a useful option for cost-effective health care in terms of cost of patient care, unnecessary admission, faulty diagnosis, and many more. Moreover, EHR can play as a common platform for the communication among various clinicians in case of diagnosis of a patient having multiple chronic conditions. Such system motivated the scientists to design a shared decision-making model by which patients and clinicians can interact regarding medical decision. Prediction of future health status and treatment trajectories from EHR necessitate some intelligent decision-making tool. However, a major difficulty for building such predictive model based on EHR is temporal irregularity. The models based on Markovian assumption are not capable to handle such irregularity. In such cases, machine learning approaches can show improved performance for disease prediction and prognosis. For example, RNN-based



**FIGURE 4** The flowchart shows how we can track mobile health check recommendation for COVID-19 (MHCRC/HCRC) and nonidentified respondents (NCRC) people based various if-else condition (received answers from mobile based survey). Thus, using this artificial intelligence-based approach, we can provide immediate isolation to high-risk patients suffering from COVID-19 and prevent the spread of the disease

models, such as LSTM and GRU among others, are significantly applicable to capture temporal relationship between diagnosis trajectories and recurrent observation of a patient. During COVID-19, the threat toward health care management has influenced the scientific communities to employ AI tool to capture the temporal dynamics of the pandemic.

For example, Wagner et al. (2020) have presented a deep learning-based data mining tool to anticipate various clinical notes including disease and its symptoms. Here, the researchers have employed the Mayo Clinic EHR of 77,167 patients with positive/negative COVID-19 detected by qRT-PCR. Besides, for the SARS-CoV2-positive patients, they have identified a total of 26 symptom categories along with 145 synonyms or synonymous phrases. Recently, another investigation (Osborne et al., 2020) has utilized the Care Assessment Need (CAN) with 1-year mortality score for bivariate analysis and chi-square test. Here, the CAN score is an existing risk estimation mechanism within the Veterans Health Administration (VA) that generates a score from 0 to 99. Such a score is automatically quantified from the national EHR database of the VA. Moreover, the score positively correlates the risk of the disease. In addition, the

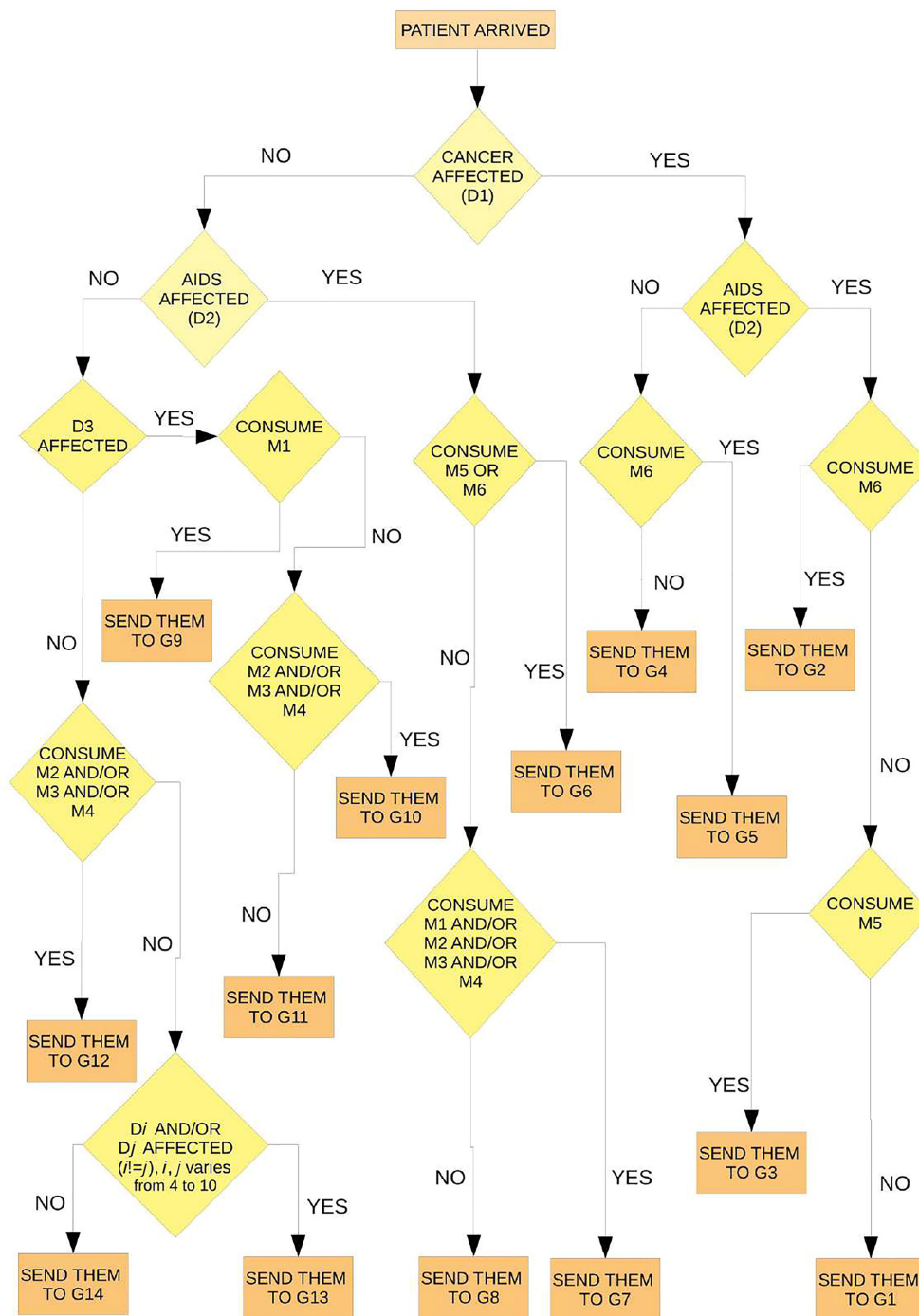


FIGURE 5 Flowchart of proposed EHR-based survey using artificial intelligence technique to detect severity of COVID-19-affected patients as well as risk of being affected by the viral disease. EHR, electronic health record

researchers have employed a logistic regression model to anticipate the risk of disease in terms of high or low using the CAN score. The result has uncovered a significant association of CAN score (>50) with several outcomes after COVID-19-positive cases, such as hospital admission, prolonged hospital stay, ICU admission, and many more.

### 4.3.2 | AI-based prevention: Contact tracing mobile app

Contact tracing and case isolation are the two main keys to contain this epidemic. The contact tracing has the ability to identify suspected individuals within the population quickly to arrest the outbreak. The technique has already shown

**TABLE 2** Set of decisions as per the flowchart of Figure 5

Serial number	Status	Groups
1	Very high risk	G-1, G-8
2	High risk	G-2, G-3, G-4, G-7
3	Moderate risk	G-5, G-6, G-11, G-13
4	Low risk	G-9, G-10, G-12
5	No risk	G-14

its efficiency to detect outbreak pattern of previous pandemic cases, such as tuberculosis (Ha et al., 2013), Ebola outbreak in 2014 (Danquah et al., 2019), and many more. A manual contact tracing is a huge job. However, it can be made easier, faster, and more efficient using a contact tracing mobile application. The application can collect various non-personal data continuously to track the location of the person along with a list of social contacts in real-time basis. For example, South Korea has started tracking phone location through GPS. In addition, it has collected credit card history, surveillance video recording, and many more for contact tracing of COVID-19 infection. Generally, such huge data are highly heterogeneous in nature like big data. Thus, contact tracing technique can be considered as in big data paradigm. Big data emphasize a situation where the size of the data surpasses the traditional computational power in terms of complexity, storage, and processor speed among others (Lefebvre-Naré et al., 2017). In other words, such huge data cannot be handled by conventional methods. Here, big data framework offers the advanced technologies for storing, distributing, and managing the information (Senthilkumar et al., 2018). In addition, it ensures analyzing information for predicting the disease transmission rate in large scale.

In this context, MapReduce (Dean & Ghemawat, 2008) is a novel big data framework offered by Google. It is a highly scalable programming model for processing big data of terrabyte volume. In this model, users can specify a map<sup>10</sup> and a reduce<sup>11</sup> function. The framework deals with partitioning of the input data, scheduling of code execution across a set of machines, handling failures, and managing the required overall inter-machine communication (Dean & Ghemawat, 2008). Consequently, Hadoop (White, 2012) is another open-source MapReduce package in distributed framework released by Apache. The framework offers a software component known as MapReduce along with storage of data in distributed fashion with the help of Hadoop Distributed File System (Borthakur, 2008).

The technique essentially involves contact identification, contact listing, and contact follow-up (Kricka et al., 2020). Researchers have investigated various digital technologies (Kretzschmar et al., 2020; Y. J. Park et al., 2020; Salathé et al., 2020) for surveillance of the physical distancing measures to efficiently control the pandemic. Several countries have supported the development of some contact tracing applications, such as “Close contact detector”<sup>12</sup> by China, TraceTogether<sup>13</sup> by Singapore, eRouška (eFacemask)<sup>14</sup> in the Czech Republic, GH Covid-19 Tracker App<sup>15</sup> of Ghana, and Aarogya Setu<sup>16</sup> by India, among others. Thus, we may conclude that contact tracing using mobile apps may work as an escape from strict lockdown measures. Keeping the idea of contact tracing, recently, Apple and Google have decided to work together on the development of such AI-based mobile app.<sup>17</sup> However, the previous investigations have not utilized any AI-based tools yet that we can include in the present report.

## 5 | AI-BASED PREDICTION ON POSSIBLE EFFECTIVE DRUGS IN MARKET AND VACCINATION

Since the outbreak of COVID-19 has been declared as “Global Pandemic,” the critical strategies to manage COVID-19 outbreak are divided into multiple parameters of early identification, early diagnosis, and rapid isolation along with the early protection (Sanders et al., 2020). Unfortunately, there is no targeted therapy exists for SARS-CoV-2 management. Supportive medicines have been prescribed with diffident outcomes.

It is clear from previous subsection that exploring appropriate drug in market for COVID-19 treatment has become inevitable need of mankind nowadays. In this regard, deep neural networks have shown satisfactory level of accuracy for prediction of various bioactivities as well as molecular properties for drug design.

Years ago, a group of scientists have applied Bayesian machine learning models to uncover the active compounds against Ebola virus (Ekins et al., 2015). The in silico technique is able to detect three in vitro active molecules such as

tilorone, pyronaridine, and quinacrine. Besides, Salama et al. (2016) have developed a novel machine learning technique to predict the possible point mutations that appear on alignments of primary RNA sequence structure. In addition, they are successful in anticipating the genotype of each nucleotide in the RNA sequence. They have employed an algorithm based on rough set to extract the point mutation patterns with satisfactory level of accuracy. Subsequently, Ekins and Perlstein (2018) have studied the usefulness of machine learning techniques to realize the treatments for rare diseases. The authors illustrated the methodology with three major iterative steps: (i) associating the targets with rare diseases, (ii) designing a model for such targets related to those diseases, and (iii) applying a machine learning technique that may discover other molecules for future test and validation.

The genome sequence analysis of SARS-CoV-2 has revealed a significant identity homology with  $\beta$ -coronavirus subtype, which is implicated as the causative factor of severe respiratory distress syndrome (Lu et al., 2020). The  $\beta$  subtype is originated from bat, which has been the leading player behind the global outbreak of SARS in 2003 (Lu et al., 2020). The invention of new antiviral agents against SARS-CoV has been started during the time of the epidemic. Unfortunately, no such antiviral formulation has successfully been reported due to the complex disease pathophysiology and host–pathogen interaction. However, information obtained from the broad area of research findings and the successive endeavors might be useful to identify the potent therapeutics.

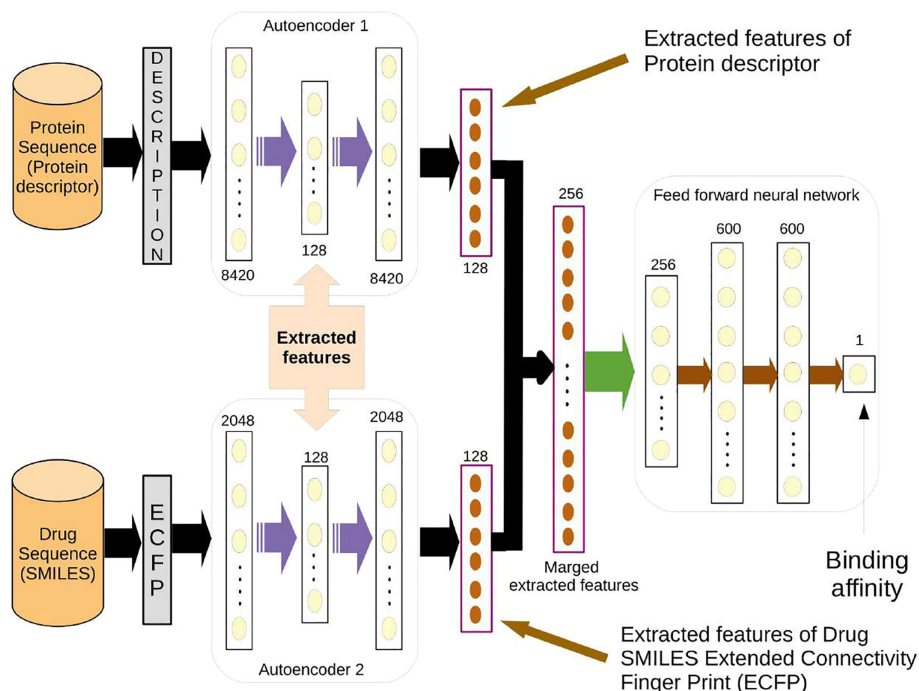
Researchers and data analysts everywhere throughout the globe are perpetually working to explore the infection modalities alongside conceivable treatment regimens by finding dynamic remedial agents. Severe complexity in the drug design and development process, as well as the elongated regulatory protocol of clinical trial, sometimes decelerates the new drug discovery to stop the pandemic. Drug repurposing can be beneficial to combat the outbreak when no such exclusively standardized therapeutics available in the market (Pawar, 2020). The objective of repurposing drugs is to find ways to employ the existing approved drug molecules to establish an innovative path in epidemic warfare.

Nowadays, a remarkable advancement in the computational power coupled with artificially intelligent methodology has been adopted to figure out the putative drug molecules (Vamathevan et al., 2019) as mentioned in some of the previous investigations. Based on the machine learning-based prediction, a known antiviral, broad-spectrum antiparasitic, antibiotic, conventional protease inhibitor can be identified (Zoffmann et al., 2019), which may be useful for the successful therapy of COVID-19. The predictive model of AI and its structured algorithm can accurately envisage the drug–target interactions underlying the complex information regarding the chemical bond formation.

Here, following the previous investigations (L. Wang, You, et al., 2018; M. Wen et al., 2017), we demonstrate how to implement a deep learning-based AI model to anticipate the affinity of COVID-19 protein sequences as the target of drugs (can be extracted from Drugbank<sup>18</sup>). Such a model usually requires comprehensive learning of various drug–target interactions. Thus, we need protein–drug combinations (can be obtained from the Drug Bank<sup>18</sup> and Pubchem database<sup>19</sup>) to train a feed-forward network. Here, the protein sequence composition descriptor (D.-S. Cao et al., 2013) can provide the structural and physicochemical features of proteins. In addition, an extended connectivity fingerprint (Rogers & Hahn, 2010) based on the Morgan algorithm (Morgan, 1965) can characterize the chemical structure of the drugs. The two autoencoders extract the high-level features of proteins and drugs. Such features become the inputs of the feed-forward network. As a result, the model might efficiently predict the known drug–protein interactions with high accuracy. Thus, we could successfully assess the probability of the proteins associated with COVID-19 as the target of unknown candidate drugs. Figure 6 depicts the outline of the deep learning-based model for predicting probable drugs to treat COVID-19. Such a predictive model might reveal the possible inhibitory roles of several drugs in spike protein–cell fusion machinery and speed up the clinical trial.

Thus, different AI-based models have demonstrated their efficiency to anticipate drug design in terms of structure of the molecules (H. Chen et al., 2018). Specifically, AI-based models have offered a number of predictive tasks, such as compound property prediction (J. Ma et al., 2015), activity prediction (Zhavoronkov et al., 2019), and reaction prediction (Fooshee et al., 2018). For example, Pollastri et al. (2002) have predicted the secondary structure of a protein with the help of RNN and LSTM networks. On the other hand, design of the vaccine has been benefited by the AI-based models. For example, Doytchinova and Flower (2007) have employed “Reverse Vaccinology” (Rappuoli, 2000) approaches for antigen prediction using machine learning. Recently, a deep learning-based investigation (J. Wu et al., 2019) has shown promising result to anticipate the binding affinity between neoantigens and their human leukocyte antigen (HLA). Here, autoencoder can play a crucial role to uncover the features of HLA-A as shown by an investigation (Miyake et al., 2018). Recently, researchers (Keshavarzi Arshadi et al., 2020) have discussed about the different steps of AI-based drug discovery and vaccine development for COVID-19.





**FIGURE 6** The figure illustrates the conceptual framework of the deep learning technique to predict the binding affinity of various readily available drugs with proteins, associated with coronavirus. The framework employs two autoencoders to extract the features from the drugs and proteins. Thereafter, the extracted features have been concatenated and fed through an artificial neural network to assess the binding affinity

## 6 | AI TECHNIQUES FOR FACE MASK DETECTION IN THE CONTEXT OF COVID-19

The infectious behavior of the coronavirus has imposed several social guidelines in daily life all over the world. Several policymakers have realized that wearing a mask in any public places should be the foremost task to arrest the spread of the disease (Feng et al., 2020). The goal of such a decision is to provide initial protection to every individual in any mass gathering. Unfortunately, a large population is not yet cautious about following such guideline. In this situation, real-time monitoring of the population may be the solution. However, manual tracking of the implementation of such policy is highly complicated in a large geographical region. Thus, the policymakers necessitate an AI-based automated face mask detection technique to control the large population in various countries across the world.

We can examine the face mask detection technique as an object recognition technique within a large image. Several researchers have involved AI-based methods for the task. For example, Ejaz et al. (2019) have investigated a principal component analysis (PCA)-based machine learning technique to differentiate masked and unmasked face. Here, the researchers have first attempted to detect the face within an image with the help of the Viola–Jones algorithm (Vikram & Padmavathi, 2017). Afterward, they have employed PCA to identify the inherent facial component. In the end, an unmasked face is predicted based on the nearest-neighbor classifier distance between the target image and the PCA feature space of the training image. As a result, the researchers have observed that the model can recognize an unmasked face more accurately over a masked face. Now, previous studies (Khojasteh et al., 2019; L. Wen et al., 2019) have observed that ResNet-50 has performed a significantly good result in extracting features. Thus, a group of scientists have proposed a transfer learning-based masked face detection classifier (Loey et al., 2021).

Such model can be integrated with a surveillance camera to identify people without wearing a mask in an online fashion. Here, the researchers have incorporated three traditional machine learning classifiers, such as support vector machine (SVM), decision tree, and ensemble learning. They have further evaluated the performance of the model with three data sets with various training and testing strategies. The investigation has revealed that SVM has outperformed other classifiers for accurate detection of unmasked faces. However, researchers have observed that many people wear a mask in the wrongly way such that it cannot cover the required portion within a face. In that situation, such methods may fail. Hussain et al. (2021) have proposed a technique to overcome such constraint. Here, a group of scientists have

presented an Internet of Things and deep transfer learning-based framework that can discriminate a face with a mask, a face without a mask, and a face with a mask in the wrong position on the face in a real-time manner. The authors have incorporated several pretrained deep learning models such as VGG-16, MobileNetV2, Inception V3, ResNet-50, and CNN to detect face mask wear in three classes. Furthermore, the investigation has successfully classified the mask into two categories as N-95 and a surgical mask. The VGG1-based transfer learning model has shown significant accuracy for both classification tasks.

Now, predicting the identity of a person with a mask sometimes become difficult. Such a limitation may provide an advantage to a person to perform an unauthenticated job. In this regard, several studies have attempted to predict the missing region after the removal of an unwanted object from an image. Hays and Efros (2007) have assessed the missing pixel of an object based on the pixel information of the candidate image that is most similar to the input sample. Subsequently, other studies (Criminisi et al., 2004; J. Wang et al., 2014) have anticipated the missing region based on the surrounding information of the image. On the other side, J.-S. Park et al. (2005) have introduced a modified PCA-based reconstruction technique to uncover a face without eyeglasses. Previous investigations (Iizuka et al., 2017; Y. Li et al., 2017) have engaged a generative adversarial network-based machine learning model to reconstruct the damaged part of an object. However, the method may fail in the case of a high-resolution image. Subsequently, the technique may suffer during the removal of large objects in facial images. Recently, several scientists have investigated a similar technique as Iizuka et al. (2017) to anticipate the unmasked version of the face based on the representation of a masked face. To retain the structural and appearance consistency of the predicted output, the researchers have incorporated a two-step process. In the first step, they have extracted the global structure of the masked face. Afterward, they have identified the missing region from the masked face in the next step. For the training process, the authors have employed a synthetic data set created from the CelebA data set. As a result, the model has significantly outperformed the previous models (Iizuka et al., 2017; Y. Li et al., 2017).

## 7 | CHALLENGES AND PROBABLE SOLUTIONS

Designing an AI-based automated system usually requires the data in a correct format. Moreover, sometimes hospital administrations are unable to maintain the patients' data with a proper time stamp. Such chaos has created several challenges for the researchers to apply AI algorithm in various predictive tasks. In this section, we have discussed probable solution to the present challenges that may motivate the scientific community to investigate a new era of research.

Based on several latest investigations, it is clear that computer scientists have confronted a major challenge in designing highly accurate AI models because of the absence of sufficiently large publicly available data sets. To tackle the situation, several organizations and web-based repositories have collected and systematically stored the COVID-19-related data. CSSE at Johns Hopkins University,<sup>20</sup> Kaggle repository "Chest X-Ray Images (Pneumonia)",<sup>21</sup> Italian COVID-19 Lung Ultrasound Database,<sup>22</sup> Novel Corona Virus 2019 data set (2020)<sup>23</sup> are well known among them. In this context, Roberts et al. (2021) have investigated the efficacy of several AI models concerning their clinical relevance. According to the researchers, most of such models are highly biased because of misguided control of the public data sets. Here, an appropriate feature engineering on the data sets can make the models more rigorous to reveal the pattern of the infection. Besides, designing an efficient AI model requires a reliable validation data set. External data sets (data sets collected from a different source) can be a good choice to consider for such validation. Furthermore, the rapid collection of data may create an imbalanced data set that may cause a severe problem for decision-making (Rahman & Davis, 2013a). We can address the challenge by adjusting the number of samples of various classes using several oversampling (Chawla et al., 2002) and under sampling techniques.

On the other side, the recent advancement of medical science strongly demands an intelligent clinical decision support system with the help of EHRs. Experts have observed that patients having several comorbidities have faced critical health issues due to the COVID-19 pandemic. Moreover, the probability of getting affected by post-COVID complications such as lung disease, heart disease, frailty, and mental health disorders has also increased (Jiang & McCoy, 2020). Avoiding such complexities may require precise knowledge about previous health issues in the form of an EHR. Previous investigations have employed EHR in several applications, viz., data-driven prediction of drug effects and interactions (Tatonetti et al., 2012), Type 2 diabetes subgroup identification (L. Li et al., 2015), classifying patients for clinical trials (Miotto & Weng, 2015), and many more. In these circumstances, personalized medicine (PM) can provide customized patient treatment based on the predicted response or disease risk. A personalized model can capture specific characteristics of the patients with the help of data from similar patient cohorts. Efficient foresight of post-COVID risks may

require capturing the transient relation among various clinical observations. Several deep learning-based frameworks have proved their efficacy in this context. For example, recently, researchers have proposed a novel time-fusion CNN-based framework (Suo et al., 2017) that can preserve the temporal relationship across adjacent visits to improve the accuracy of disease prediction. Although implementing a PM framework has several challenges, it can be beneficial in the clinical research domain for COVID-19.

In contrast to EHR, several biobanks such as genetic banks, tissue banks, and others have facilitated the researcher to uncover the intricate association of COVID-19 with different factors like environments (Hall et al., 2014), lifestyle choice (Rutten-Jacobs et al., 2018; Said et al., 2018), and genetics. The foremost advantage of biobanking is that researchers can utilize the data from a unified data repository having standardized data collection protocols. Now, accurate diagnosis of the patients essentially requires the collaboration of clinicians with AI experts (Abdulkareem & Petersen, 2021).

Now, understanding the disease trajectory often necessitate recording clinical observation promptly. Unfortunately, several hospitals usually capture clinical inspection on irregular timing basis in ICU. Moreover, the frequency of the measurements as well as the observation parameters may vary across patients. Such inconsistent study may result in several missing values in the EHR (Beaulieu-Jones & Moore, 2017). Several causes of missing data in the data set may influence the prediction of the survival of a patient. Here, an AI-based framework can avoid the limitation by imputing missing values in respective places. For example, Jerez et al. (2010) have investigated the efficacy of six methods such as mean, hot-deck, MI, MLP, SOM, and KNN to impute missing values in “El Álamo-I” (Martín et al., 2004; Ruiz et al., 2005) breast cancer data set. Afterward, they have employed the ANN model to anticipate breast cancer recurrence. The researchers have further observed that predicted the missing values using all the methods other than hot-deck increased the accuracy of the predictive model. More recently, Rahman and Davis (2013b) have studied a similar type of task.

Clinical studies have observed that a significant number of patients have experienced several prolonged symptoms following recovery from COVID-19. Such cases are often termed long COVID (Sykes et al., 2021). In general, long COVID symptoms include extreme tiredness (fatigue), shortness of breath, difficulty sleeping (insomnia), and many more.<sup>24</sup> Therefore, predicting the probability of having long COVID symptoms based on medical history is a big challenge for the researchers. Here, we can utilize the EHR and various OMICS information to design an AI model for such a task.

During the COVID-19 pandemic, we have observed that the virus has mutated very rapidly. At the same time, scientists have suggested that vaccine can be the best solution to arrest the emergency across the world. Unfortunately, we have observed several cases where the patient got affected even after taking the vaccine. Different strains of the mutated virus are responsible for the epidemics in the various demographic regions. Naturally, such a rapid mutation may affect the process of designing proper vaccines. In this situation, early knowledge about the mutated strains may help scientists to develop more effective vaccines. Scientists can employ a genetic algorithm to uncover a set of new strains of the virus having more deadly characteristics in terms of infectious nature. In addition, genome editing tool (Cyranoski, 2016) can be a considerable opportunity to fight against the disease. CRISPR/Cas has influenced the scientific community for disease diagnosis and treatment. The reason is that the technique is cost-effective, quicker as well as more accurate (Chekani-Azar et al., 2020). Scientist can exploit the technology for drug development against novel coronavirus disease. Besides, the researchers can concentrate on pharmacogenomics (Relling & Evans, 2015) that may allow the development of effective drugs for several health disorders and pandemics like COVID-19.

## 8 | DISCUSSION AND FUTURE DIRECTION

In the present study, we have tried to find some fruitful ways to contain current pandemic from many aspects with the help of AI. First of all, we have discussed elaborately about different databases, helping us in analyzing current scenario in various perspectives. To obtain an efficient AI model, developing a proper data set is crucial. Improper sampling may create a model highly biased toward a particular or a set of features. For example, collecting the data from a particular demographic region may create an AI model based on a particular cohort. Therefore, an efficient sampling of features, as well as instances, is vital.

In demand of rapid testing, CNN-based chest image (mostly, x-ray with very few CT/HRCT) processing techniques have shown promising result to detect COVID-19 pneumonia in the early stage of disease. In addition, improved AI-based image processing techniques for chest CT/HRCT images can be developed in near future to detect high risky

COVID-19 pneumonia more accurately and rapidly. Thus, it may be a good support of qRT-PCR and antigen detection method. However, asymptomatic patients may not undergo such primary detection techniques. Such ignorance can trigger a negative impact on the motivation of the outbreak control. In this situation, different digital tools, such as contact tracing mobile app, can be figured out to stop the transmission of COVID-19 without vigorous lockdown and mass quarantine. In addition, genomic profiling of the traced person along with their EHR can play a crucial role. As EHR contain historical time-series data about previous disease and medicine consumed, RNN-based deep learning technique can be applied to predict any unwanted future circumstances. Accordingly hospitals can take appropriate decision in the early stage of the asymptotically infected cases. Such planning may not only decrease the number of death cases but also control the exponential growth of the pandemic situation.

On the other hand, our primary concern has dealt with how to address different patterns of confirmed infections and deaths in different countries. Because the promising control on the disease outbreak requires accurate prediction of such pattern. We have discussed a number of AI-based models in this regard. However, more robust and accurate mathematical models are needed to capture the variation in infection and death throughout the world. In addition, efficient machine learning algorithms for sentiment analysis based on social networking data have been discussed. However, in social media fake news is well known since decades. As the predictive result depends on hoax information related to COVID-19, it may instigate a number of wrong decisions. Thus, before applying AI technique on social media data, false data should be removed. Presently, a popular way of spreading fake news is in the form of clickbaits (Y. Chen et al., 2015). In this context, a recent study (Aldwairi & Alwahedi, 2018) has investigated a number of machine learning classifiers, such as Bayes Net, Logistic, Random Tree, and Naive Bayes, to recognize and eliminate fake social media news as well as fake sites. Here, logistic classifier has shown promising accuracy. The fake news can broadly be categorized in three ways, such as the text of an article, the number of user response, and the source users promoting it. A study (Ruchansky et al., 2017) has proposed a RNN-based deep learning technique to detect various types of fake information. In short, an AI-based predictive model to anticipate disease transmission pattern or “hotspot” requires a two-layer pipeline. In the first layer, the model should identify and eliminate all the fake information. Thereafter, the clean data can be fed into the next AI layer to get the desired output.

Moreover, we have explored different factors and medium of the infection transmission. Recent investigations have observed a significant association of environment temperature and humidity with COVID-19-infected cases. Nevertheless, we have not found proper validation of a previous finding claiming that people with blood group “A” is more prone to be affected by COVID-19. Such observation may help the medical practitioners to identify potential-infected cases before any adverse situation arises.

However, during treatment, medical practitioners have found common symptoms in COVID-19-affected patients including fever and nonproductive cough. Due to unavailability of any deterministic way to arrest the disease, the doctors depend on a set of supportive medicines for treatment. Although some random clinical trials have been performed to uncover the effect of various drugs, its results have not ensured a set of particular drugs against the virus. Moreover, a considerable amount of time needs to be invested for each trial. In this consequence, AI may help to anticipate the effect of various drugs in a faster way. A deep learning-based framework can assess the binding affinity between drugs and the proteins associated with coronavirus. Here, the autoencoders extract the highly significant features from the molecular fingerprint of the drugs and proteins. Besides, the artificial neural network predicts the binding affinity between readily available drugs and proteins. Thus, more efficient deep learning-based techniques need to be developed to speed up the drug design process.

The governments of all the countries have already convinced that face mask is an effective practice to decrease the exponential growth of the disease. Although all the nations have imposed a guideline on wearing a face mask, strict supervising is required to maintain the rule. We have discussed several latest AI-based investigations in this context. In addition, we have encapsulated various automated tasks related to COVID-19 based on the class of AI techniques. We have further recognized few challenges during the predictive goal using present methods. We have realized that EHR can be an effective tool to identify patients' cohorts under significant threat during the pandemic. Previous investigations have influenced us to employ EHR coupled with OMICS information to find a PM solution. The world has already experienced several mutated strains of the virus. Some of them have more deadly characteristics than their ancestor strains. Various mutated strains may be the foremost obstacle to designing an effective vaccine. Thus, early knowledge of the fatal features of mutated strains may help the scientific community to develop effective drugs. In this context, we may utilize genetic algorithm-based techniques for such prediction.

Interestingly, if approximately 50%–70% of the population become immune against a disease epidemic, the epidemic itself ceases. For example, a COVID-19 patient can infect approximately four people, but if the latter becomes immune

for the disease then the former will be incapable of infecting them. Thus, the virus remains in the body of the COVID-19 patient and dies maximum within 4 weeks. In this way, an epidemic can be controlled. The concept of vaccine is to increase the herd immunity (Fine et al., 2011). Various epidemics, such as, polio and measles, have been eradicated in this way. It would be interesting to see whether AI can help in predicting the people with low clinical risk and also immune individuals, leading to building up rapid and high herd immunity in near future.

Any infection may lead to formation of two types of antibody in the body, that is, IgM and IgG (Duan et al., 2020). Presence of IgM in the body of the host means that it still contains acute infection. Presence of both IgM and IgG in the body means that the host has started acquiring immunity against the disease but the infection is still persisting as IgM is still present. Absence of IgM and presence of only IgG means that the host has probably become immune to the disease and therefore further spread of the disease is not possible. As it takes a minimum of 1 week for the antibodies to be produced, rapid antibody testing is preferably done in the 7th day, then 14th day, and so on, until IgM becomes negative and IgG becomes positive. Theoretically, an immune person can neither get infected and nor can spread the disease. However, as COVID-19 is a new disease and apparently many individuals are getting infected second time, it cannot be surely declared that one is completely immune after being infected for the first time. Thus, it would be very useful if AI can help in detecting the individuals who are completely immune to COVID-19.

During testing, once qRT-PCR test proves a person COVID-19 negative, we should wait for a month to watch whether that person is getting reinfected or not. If not then it is expected that the plasma of this person contains lots of IgG. This plasma can be collected and injected in critically ill patients to see whether condition of the patient is improving or not. A few case reports say that such condition is improved on applying the above method (Z. Li, Yi, et al., 2020), whereas many cases show that there is no significant improvement. Therefore, it would be helpful if AI can facilitate identifying the individuals who are immune to COVID-19. It will help in deciding whether their plasma can be collected for the therapy for saving coronavirus patients. In this regard, social networking as well as EHR data related to COVID-19 cases may help in developing such AI-based model in upcoming days.

## AUTHOR CONTRIBUTIONS

**Abhisek Bakshi:** Data curation (equal); methodology (equal); resources (equal); software (equal); visualization (equal); writing – review and editing (supporting). **Srijani Mukherjee:** Conceptualization (supporting); formal analysis (supporting); methodology (supporting); resources (supporting); validation (supporting); writing – review and editing (supporting). **Kuntal Das:** Data curation (supporting); methodology (supporting); resources (supporting); software (supporting); validation (supporting); visualization (supporting); writing – original draft (supporting). **Soumyajeet Talukdar:** Data curation (supporting); formal analysis (supporting); methodology (supporting); resources (supporting); validation (supporting); visualization (supporting); writing – original draft (supporting). **Pratyayee Chatterjee:** Data curation (supporting); formal analysis (supporting); methodology (supporting); resources (supporting); software (supporting); validation (supporting); visualization (supporting); writing – original draft (supporting). **Sagnik Mondal:** Data curation (supporting); formal analysis (supporting); resources (supporting); validation (supporting); writing – original draft (supporting). **Puspita Das:** Data curation (supporting); resources (supporting); software (supporting); writing – original draft (supporting). **Subhrojit Ghosh:** Data curation (supporting); formal analysis (supporting); resources (supporting); software (supporting); writing – original draft (supporting). **Archisman Som:** Resources (supporting); writing – original draft (equal). **Pritha Roy:** Visualization (equal); writing – original draft (equal). **Rima Kundu:** Data curation (supporting); writing – original draft (supporting). **Akash Sarkar:** Validation (supporting); writing – original draft (supporting). **Arnab Biswas:** Writing – original draft (supporting). **Karnelia Paul:** Validation (equal); writing – original draft (equal); writing – review and editing (supporting). **Sujit Basak:** Investigation (equal); validation (equal); writing – original draft (equal); writing – review and editing (equal). **Krishnendu Manna:** Formal analysis (equal); investigation (equal); validation (equal); writing – original draft (equal); writing – review and editing (equal). **Chinmay Saha:** Conceptualization (equal); formal analysis (equal); investigation (equal); project administration (equal); supervision (equal); validation (equal); writing – original draft (equal); writing – review and editing (equal). **Satinath Mukhopadhyay:** Writing – review and editing (equal). **Nitai P. Bhattacharyya:** Conceptualization (equal); supervision (equal); writing – review and editing (equal). **Rajat K. De:** Conceptualization (equal); writing – review and editing (lead). **Abhijit Dasgupta:** Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); project administration (lead); resources (equal); software (lead); supervision (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal).

## ACKNOWLEDGMENT

Rajat K. De acknowledges the project sanctioned under MATRICS Special call under COVID-19 (File Number: MSC/2020/000350) in the Indian Statistical Institute by the Science and Engineering Research Board (SERB) under the Department of Science and Technology (DST), Government of India.

## CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

Sujit Basak  <https://orcid.org/0000-0002-9845-6326>

Rajat K. De  <https://orcid.org/0000-0001-6080-1131>

## RELATED WIRES ARTICLE

[Sentiment analysis for mining texts and social networks data: Methods and tools](#)

## ENDNOTES

- <sup>1</sup> <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- <sup>2</sup> <https://www.worldometers.info/coronavirus/>
- <sup>3</sup> <https://coronavirus.jhu.edu/map.html>
- <sup>4</sup> <https://www.drugbank.ca/>
- <sup>5</sup> <https://github.com/JordanMicahBennett/>
- <sup>6</sup> <https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset>
- <sup>7</sup> <https://www.kaggle.com/parthachakraborty/pneumonia-chest-x-ray>
- <sup>8</sup> <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
- <sup>9</sup> <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>
- <sup>10</sup> Processes a key/value pair to generate a set of intermediate key/value pairs.
- <sup>11</sup> Integrates all intermediate values associated with the same intermediate key.
- <sup>12</sup> <https://thenextweb.com/china/2020/03/03/chinas-covid-19-app-reportedly-color-codes-people-and-shares-data-with-cops/>
- <sup>13</sup> <https://en.m.wikipedia.org/wiki/TraceTogether>
- <sup>14</sup> <https://www.praguemorning.cz/erouska-app-can-trace-your-contacts-without-sacrificing-your-privacy/>
- <sup>15</sup> <https://ghcovid19.com/>
- <sup>16</sup> <https://www.mygov.in/aarogya-setu-app/>
- <sup>17</sup> <https://www.economist.com/science-and-technology/2020/04/16/app-based-contact-tracing-may-help-countries-get-out-of-lockdown>
- <sup>18</sup> <https://www.drugbank.ca/>
- <sup>19</sup> <https://pubchem.ncbi.nlm.nih.gov/>
- <sup>20</sup> <https://github.com/CSSEGISandData/COVID-19>
- <sup>21</sup> <http://tuberculosis.by>
- <sup>22</sup> <https://iclus-web.bluetensor.ai/>
- <sup>23</sup> <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset/>
- <sup>24</sup> <https://www.nhs.uk/conditions/coronavirus-covid-19/long-term-effects-of-coronavirus-long-covid/>

## Further Reading

Ahmed, F., Ahmed, N., Pissarides, C., & Stiglitz, J. (2020). Why inequality could spread COVID-19. *The Lancet Public Health*, 5, e240.

Buheji, M., da Costa Cunha, K., Beka, G., Mavric, B., de Souza, Y., da Costa Silva, S. S., Hanafi, M., & Yein, T. C. (2020). The extent of COVID-19 pandemic socio-economic impact on global poverty. A global integrative multidisciplinary review. *American Journal of Economics*, 10, 213–224.

- Caly, L., Druce, J. D., Catton, M. G., Jans, D. A., & Wagstaff, K. M. (2020). The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Research*, *178*, 104787.
- Chan, K. C. C., Hu, P.-W., & You, Z.-H. (2016). Large-scale prediction of drug-target interactions from deep representations. In *2016 International Joint Conference on Neural Networks (IJCNN)* (pp. 1236–1243). IEEE.
- Chen, X.-Y., & Xu, Z. (2016). Artemisinin and plant secondary metabolism. *Science Bulletin*, *61*, 1–2.
- Kissler, S., Tedijanto, C., Goldstein, E., Grad, Y., & Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the post-pandemic period. *Science*, *368*, 860–888.
- Lv, Z., Chu, Y., & Wang, Y. (2015). HIV protease inhibitors: A review of molecular selectivity and toxicity. *HIV/AIDS (Auckland, NZ)*, *7*, 95.
- Magar, R., Yadav, P., & Farimani, A. B. (2021). Potential neutralizing antibodies discovered for novel corona virus using machine learning. *Scientific Reports*, *11*, 1–11.
- McClellan, K., & Perry, C. M. (2001). Oseltamivir. *Drugs*, *61*, 263–283.
- Roy, S., He, R., Kapoor, A., Forman, M., Mazzone, J. R., Posner, G. H., & Arav-Boger, R. (2015). Inhibition of human cytomegalovirus replication by artemisinins: Effects mediated through cell cycle modulation. *Antimicrobial Agents and Chemotherapy*, *59*, 3870–3879.
- Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: Current knowledge. *Virology Journal*, *16*, 69.
- Sen Gupta, P. S., Biswal, S., Panda, S. K., Ray, A. K., & Rana, M. K. (2020). Binding mechanism and structural insights into the identified protein target of COVID-19 and importin- $\alpha$  with in-vitro effective drug ivermectin. *Journal of Biomolecular Structure and Dynamics*, *40*, 1–10.
- Sumner, A., Hoy, C., & Ortiz-Juarez, E. (2020). *Estimates of the impact of COVID-19 on global poverty* (pp. 800–809). UNU-WIDER.

## REFERENCES

- Abdulkareem, M., & Petersen, S. E. (2021). The promise of AI in detection, diagnosis, and epidemiology for combating COVID-19: Beyond the hype. *Frontiers in Artificial Intelligence*, *4*, 53.
- Afshar, M., Dligach, D., Sharma, B., Cai, X., Boyda, J., Birch, S., Valdez, D., Zelisko, S., Joyce, C., Modave, F., & Price, R. (2019). Development and application of a high throughput natural language processing architecture to convert all clinical documents in a clinical data warehouse into standardized medical vocabularies. *Journal of the American Medical Informatics Association*, *26*, 1364–1369.
- Alakus, T. B., & Turkoglu, I. (2020). Comparison of deep learning approaches to predict covid-19 infection. *Chaos, Solitons & Fractals*, *140*, 110120.
- Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, *141*, 215–222.
- Andersen, K. N., Medaglia, R., & Henriksen, H. Z. (2012). Social media in public health care: Impact domain propositions. *Government Information Quarterly*, *29*, 462–469.
- Anderson, R. M. (1991). Discussion: The Kermack–Mckendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, *53*, 1.
- Arora, P., Kumar, H., & Panigrahi, B. K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons & Fractals*, *139*, 110017.
- Arpaci, I., Alshehaby, S., Al-Emran, M., Khasawneh, M., Mahariq, I., Abdeljawad, T., & Hassanien, A. E. (2020). Analysis of twitter data using evolutionary clustering during the COVID-19 pandemic. *Computers, Materials & Continua*, *65*, 193–204.
- Arun Kumar, K., Kalaga, D. V., Kumar, C. M. S., Kawaji, M., & Brenza, T. M. (2021). Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUS) and long short-term memory (LSTM) cells. *Chaos, Solitons & Fractals*, *146*, 110861.
- Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Kalhori, S. R. N. (2020). Predicting COVID-19 incidence through analysis of Google trends data in Iran: Data mining and deep learning pilot study. *Journal of Medical Internet Research Public Health and Surveillance*, *6*, e18828.
- Beaulieu-Jones, B. K., & Moore, J. H. (2017). Missing data imputation in the electronic health record using deeply learned autoencoders. *Pacific Symposium on Biocomputing*, *22*, 207–218.
- Ben-Assuli, O., Leshno, M., & Shabtai, I. (2012). Using electronic medical record systems for admission decisions in emergency departments: Examining the crowdedness effect. *Journal of Medical Systems*, *36*, 3795–3803.
- Borghi, P. H., Zakordonets, O., & Teixeira, J. P. (2021). A COVID-19 time series forecasting model based on MLP ANN. *Procedia Computer Science*, *181*, 940–947.
- Borthakur, D. (2008) *HDFS architecture guide*. Hadoop Apache Project, 53.
- Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *The Lancet*, *395*, 912–920.
- Calafiore, G. C., Novara, C., & Possieri, C. (2020). A time-varying SIRD model for the COVID-19 contagion in Italy. *Annual Reviews in Control*, *50*, 361–372.
- Cao, D.-S., Xu, Q.-S., & Liang, Y.-Z. (2013). propy: A tool to generate various modes of Chou's PseAAC. *Bioinformatics*, *29*, 960–962.
- Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., & Zheng, J. (2020). The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Research*, *287*, 112934.
- Car, Z., Baressi Šegota, S., Anelić, N., Lorencin, I., & Mrzljak, V. (2020). Modeling the spread of COVID-19 infection using a multilayer perceptron. *Computational and Mathematical Methods in Medicine*, *2020*, 1–10.
- Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*, *97*, 106754.

- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Chekani-Azar, S., Gharib Mombeni, E., Birhan, M., & Yousefi, M. (2020). CRISPR/Cas9 gene editing technology and its application to the coronavirus disease (COVID-19), a review. *Journal of World's Poultry Research*, 10, 1–9.
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today*, 23, 1241–1250.
- Chen, X., Tang, Y., Mo, Y., Li, S., Lin, D., Yang, Z., Yang, Z., Sun, H., Qiu, J., Liao, Y., Xiao, J., Chen, X., Wu, X., Wu, R., & Dai, Z. (2020). A diagnostic model for coronavirus disease 2019 (COVID-19) based on radiological semantic and clinical features: A multi-center study. *European Radiology*, 30, 4893–4902.
- Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). Misleading online content: Recognizing clickbait as “false news”. In *Proceedings of the 2015 ACM on workshop on multimodal deception detection* (pp. 15–19), Association for Computing Machinery.
- Cheng, X., Chai, F., Gao, J., & Zhang, K. (2011). 1st opt and global optimization platform-comparison and case study. *Proceedings of the 4th IEEE International Conference on Computer Science and Information Technology*, Chengdu, China.
- Choe, J., Lee, S. M., Do, K.-H., Lee, G., Lee, J.-G., Lee, S. M., & Seo, J. B. (2019). Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology*, 292, 365–373.
- Chowell, G. (2017). Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2, 379–398.
- Criminisi, A., Pérez, P., & Toyama, K. (2004). Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13, 1200–1212.
- Cyranoski, D. (2016). CRISPR gene-editing tested in a person for the first time. *Nature*, 539, 479.
- Danquah, L. O., Hasham, N., MacFarlane, M., Conteh, F. E., Momoh, F., Tedesco, A. A., Jambai, A., Ross, D. A., & Weiss, H. A. (2019). Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: A proof-of-concept study. *BMC Infectious Diseases*, 19, 810.
- Dastider, A. G., Sadik, F., & Fattah, S. A. (2021). An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound. *Computers in Biology and Medicine*, 132, 104296.
- Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51, 107–113.
- Di Giambardino, P., & Iacoviello, D. (2017). Optimal control of SIR epidemic model with state dependent switching cost index. *Biomedical Signal Processing and Control*, 31, 377–380.
- Diggle, P., Liang, K.-Y., & Zeger, L. S. (2002). *Analysis of longitudinal data*. Oxford University Press.
- Dong, L., Zhang, Y., Tao, Q., Deng, S., & Li, N. (2019). Statistical estimate of epidemic trend, suggestions and lessons for public safety from the 2019 novel coronavirus (COVID-19). *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.353966>
- Doytchinova, I. A., & Flower, D. R. (2007). Vaxijen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BioMed Central Bioinformatics*, 8, 4.
- Duan, K., Liu, B., Li, C., Zhang, H., Yu, T., Qu, J., Zhou, M., Chen, L., Meng, S., Hu, Y., Peng, C., Yuan, M., Huang, J., Wang, Z., Yu, J., Gao, X., Wang, D., Yu, X., Li, L., ... Yang, X. (2020). Effectiveness of convalescent plasma therapy in severe COVID-19 patients. *Proceedings of the National Academy of Sciences*, 117, 9490–9496.
- Ejaz, M. S., Islam, M. R., Sifatullah, M., & Sarker, A. (2019). Implementation of principal component analysis on masked and non-masked face recognition. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (pp. 1–5). IEEE.
- Ekins, S., Freundlich, J. S., Clark, A. M., Anantpadma, M., Davey, R. A., & Madrid, P. (2015). Machine learning models identify molecules active against the Ebola virus in vitro. *F1000Research*, 4, 1091.
- Ekins, S., & Perlstein, E. O. (2018). Doing it all-how families are reshaping rare disease research. *Pharmaceutical Research*, 35, 192.
- Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., & Ji, W. (2020). Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *Radiology*, 296, E115–E117.
- Feng, S., Shen, C., Xia, N., Song, W., Fan, M., & Cowling, B. J. (2020). Rational use of face masks in the COVID-19 pandemic. *The Lancet Respiratory Medicine*, 8, 434–436.
- Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., & Fraser, C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368, eabb6936.
- Fine, P., Eames, K., & Heymann, D. L. (2011). “Herd immunity”: A rough guide. *Clinical Infectious Diseases*, 52, 911–916.
- Fisman, D., Khoo, E., & Tuite, A. (2014). Early epidemic dynamics of the west African 2014 Ebola outbreak: Estimates derived with a simple two-parameter model. *PLoS Currents*, 6. <https://doi.org/10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571>
- Fisman, D. N., Hauck, T. S., Tuite, A. R., & Greer, A. L. (2013). An idea for short term outbreak projection: Nearcasting using the basic reproduction number. *PLoS One*, 8, e83622.
- Flesia, L., Monaro, M., Mazza, C., Fietta, V., Colicino, E., Segatto, B., & Roma, P. (2020). Predicting perceived stress related to the COVID-19 outbreak through stable psychological traits and machine learning models. *Journal of Clinical Medicine*, 9, 3350.
- Fontanesi, L., Marchetti, D., Mazza, C., Di Giandomenico, S., Roma, P., & Verrocchio, M. C. (2020). The effect of the COVID-19 lockdown on parents: A call to adopt urgent measures. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12, S79–S81.
- Fooshee, D., Mood, A., Gutman, E., Tavakoli, M., Urban, G., Liu, F., Huynh, N., Van Vranken, D., & Baldi, P. (2018). Deep learning for chemical reaction prediction. *Molecular Systems Design & Engineering*, 3, 442–452.



- Funk, B., Sadeh-Sharvit, S., Fitzsimmons-Craft, E. E., Trockel, M. T., Monterubio, G. E., Goel, N. J., Balantekin, K. N., Eichen, D. M., Flatt, R. E., Firebaugh, M.-L., Jacobi, C., Graham, A. K., Hoogendoorn, M., Wilfley, D. E., & Taylor, C. B. (2020). A framework for applying natural language processing in digital health interventions. *Journal of Medical Internet Research*, *22*, e13855.
- Geller, C., Varbanov, M., & Duval, R. E. (2012). Human coronaviruses: Insights into environmental resistance and its influence on the development of new antiseptic strategies. *Viruses*, *4*, 3044–3068.
- Gómez, P., Semmler, M., Schützenberger, A., Bohr, C., & Döllinger, M. (2019). Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. *Medical & Biological Engineering & Computing*, *57*, 1451–1463.
- Gupta, A., Banerjee, S., & Das, S. (2020). Significance of geographical factors to the COVID-19 outbreak in India. *Modeling Earth Systems and Environment*, *6*, 2645–2653.
- Gupta, V. K., Gupta, A., Kumar, D., & Sardana, A. (2021). Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model. *Big Data Mining and Analytics*, *4*, 116–123.
- Ha, Y. P., Littman-Quinn, R., Antwi, C., Seropola, G., Green, R. S., Tesfalul, M. A., Ho-Foster, A., Luberti, A. A., Holmes, J. H., Steenhoff, A. P., & Kovarik, C. L. (2013). A mobile health approach to tuberculosis contact tracing in resource-limited settings. *Studies in Health Technology and Informatics*, *192*, 1188.
- Hall, M. A., Dudek, S. M., Goodloe, R., Crawford, D. C., Pendergrass, S. A., Peissig, P., Brilliant, M., McCarty, C. A., & Ritchie, M. D. (2014). Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield personalized medicine research project biobank. In *Biocomputing 2014* (pp. 200–211). World Scientific.
- Hays, J., & Efros, A. A. (2007). Scene completion using millions of photographs. *Association for Computing Machinery Transactions on Graphics (TOG)*, *26*, 4–es.
- Hoque, E., Islam, M. S., Das, S. K., Mitra, D. K., & Amin, M. R. (2021). Adjusted dynamics of COVID-19 pandemic due to herd immunity in Bangladesh. *International Journal of Modern Physics C*, *32*(10), 2150140.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., ... Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*, 497–506.
- Huremović, D. (2019). *Psychiatry of pandemics: A mental health response to infection outbreak*. Springer.
- Hussain, S., Yu, Y., Ayoub, M., Khan, A., Rehman, R., Wahid, J. A., & Hou, W. (2021). IoT and deep learning based approach for rapid screening and face mask detection for infection spread control of COVID-19. *Applied Sciences*, *11*, 3495.
- Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2017). Globally and locally consistent image completion. *Association for Computing Machinery Transactions on Graphics (TOG)*, *36*, 1–14.
- Islam, M. Z., Islam, M. M., & Asraf, A. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using x-ray images. *Informatics in Medicine Unlocked*, *20*, 100412.
- Javed, B., Sarwer, A., Soto, E., & Mashwani, Z. (2020). Is Pakistan's response adequate enough to stem a coronavirus (SARS-CoV-2) outbreak. *Frontiers in Medicine*, *7*. <https://doi.org/10.3389/fmed.2020.00158>
- Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, *50*, 105–115.
- Jiang, D. H., & McCoy, R. G. (2020). Planning for the post-COVID syndrome: How payers can mitigate long-term complications of the pandemic. *Journal of General Internal Medicine*, *35*, 3036–3039.
- Kaushal, S., Rajput, A. S., Bhattacharya, S., Vidyasagar, M., Kumar, A., Prakash, M. K., & Ansumali, S. (2020). Estimating the herd immunity threshold by accounting for the hidden asymptomatics using a COVID-19 specific model. *PLoS One*, *15*, e0242132.
- Kelly, J. D., Park, J., Harrigan, R. J., Hoff, N. A., Lee, S. D., Wannier, R., Selo, B., Mossoko, M., Njoloko, B., Okitolonda-Wemakoy, E., Mbala-Kingebeni, P., Rutherford, G. W., Smith, T. B., Ahuka-Mundeké, S., Muyembe-Tamfum, J. J., Rimoin, A. W., & Schoenberg, F. P. (2019). Real-time predictions of the 2018–2019 Ebola virus disease outbreak in The Democratic Republic of the Congo using Hawkes point process models. *Epidemics*, *28*, 100354.
- Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., Pei, J., Ting, M. Y. L., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., ..., Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*, 1122–1131.
- Keshavarzi Arshadi, A., Webb, J., Salem, M., Cruz, E., Calad-Thomson, S., Ghadirian, N., Collins, J., Diez-Cecilia, E., Kelly, B., Goodarzi, H., & Shiun Yuan, J. (2020). Artificial intelligence for COVID-19 drug discovery and vaccine development. *Frontiers in Artificial Intelligence*, *3*, 65.
- Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine*, *196*, 105581.
- Khojasteh, P., Júnior, L. A. P., Carvalho, T., Rezende, E., Aliahmad, B., Papa, J. P., & Kumar, D. K. (2019). Exudate detection in fundus images using deeply-learnable features. *Computers in Biology and Medicine*, *104*, 62–69.
- Khozeimeh, F., Sharifrazi, D., Izadi, N. H., Joloudari, J. H., Shoeibi, A., Alizadehsani, R., Gorriz, J. M., Hussain, S., Sani, Z. A., Moosaei, H., Khosravi, A., Nahavandi, S., & Islam, S. M. S. (2021). Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients. *Scientific Reports*, *11*, 1–18.
- Kraemer, M., Golding, N., Bisanzio, D., Bhatt, S., Pigott, D., Ray, S., Brady, O., Brownstein, J., Faria, N., Cummings, D., Pybus, O., Smith, D., Tatem, A., Hay, S., & Reiner Jr, R. (2019). Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Scientific Reports*, *9*, 1–11.

- Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C., van Boven, M., van de Wijger, J. H., & Bonten, M. J. (2020). Impact of delays on effectiveness of contact tracing strategies for COVID-19: A modelling study. *The Lancet Public Health*, 5, e452–e459.
- Kricka, L. J., Polevikov, S., Park, J. Y., Fortina, P., Bernardini, S., Satchkov, D., Kolesov, V., & Grishkov, M. (2020). Artificial intelligence-powered search tools and resources in the fight against COVID-19. *Electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine*, 31, 106–116.
- Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., ... the SARS Working Group. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *New England Journal of Medicine*, 348, 1953–1966.
- Lambin, P., Leijenaar, R. T., Deist, T. M., Peerlings, J., de Jong, E. E. C., van Timmeren, J., Sanduleanu, S., Larue, R. T., Even, A. J., Jochems, A., van Wijk, Y., Woodruff, H., van Soest, J., Lustberg, T., Roelofs, E., van Elmp, W., Dekker, A., Mottaghy, F. M., Wildberger, J. E., & Walsh, S. (2017). Radiomics: The bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, 14, 749–762.
- Lee, J., Ta, C., Kim, J. H., Liu, C., & Weng, C. (2021). Severity prediction for COVID-19 patients via recurrent neural networks. *American Medical Informatics Association Annual Symposium Proceedings*, 2021, 374.
- Lefebvre-Naré, F., Lemire, S., & Petersson, G. J. (2017). What is big data? In *Cyber society, big data and evaluation*. Transaction Publishers.
- Li, C., Chen, L. J., Chen, X., Zhang, M., Pang, C. P., & Chen, H. (2020). Retrospective analysis of the possibility of predicting the COVID-19 outbreak from internet searches and social media data, China, 2020. *Eurosurveillance*, 25, 2000199.
- Li, G., Fan, Y., Lai, Y., Han, T., Li, Z., Zhou, P., Pan, P., Wang, W., Hu, D., Liu, X., Zhang, Q., & Wu, J. (2020). Coronavirus infections and immune responses. *Journal of Medical Virology*, 92, 424–432.
- Li, L., Cheng, W.-Y., Glicksberg, B. S., Gottesman, O., Tamler, R., Chen, R., Bottinger, E. P., & Dudley, J. T. (2015). Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Science Translational Medicine*, 7, 311ra174.
- Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., Cao, K., Liu, D., Wang, G., Xu, Q., Fang, X., Zhang, S., Xia, J., & Xia, J. (2020). Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*, 296, E65–E71.
- Li, Y., Liu, S., Yang, J., & Yang, M.-H. (2017). Generative face completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3911–3919).
- Li, Y.-C., Bai, W.-Z., & Hashikawa, T. (2020). The neuroinvasive potential of SARS-CoV2 may be at least partially responsible for the respiratory failure of COVID-19 patients. *Journal of Medical Virology*, 92, 552–555.
- Li, Z., Yi, Y., Luo, X., Xiong, N., Liu, Y., Li, S., Sun, R., Wang, Y., Hu, B., Chen, W., Zhang, Y., Wang, J., Huang, B., Lin, Y., Yang, J., Cai, W., Wang, X., Cheng, J., Chen, Z., ... Ye, F. (2020). Development and clinical application of a rapid IGM-IGG combined antibody test for SARS-CoV-2 infection diagnosis. *Journal of Medical Virology*, 92, 1518–1524.
- Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, 167, 108288.
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., ... Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet*, 395, 565–574.
- Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., & Svetnik, V. (2015). Deep neural nets as a method for quantitative structure–activity relationships. *Journal of Chemical Information and Modeling*, 55, 263–274.
- Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., & Luo, B. (2020). Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of the Total Environment*, 724, 138226.
- Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, 368, 742–746.
- Martín, M., Llombart-Cussac, A., Lluch, A., Alba, E., Munárriz, B., Tusquets, I., Barnadas, A., Balil, A., Dorta, J., Picó, C., & Grupo GEICAM. (2004). Epidemiological study of the GEICAM group about breast cancer in Spain (1990–1993): El Alamo project. *Medicina Clinica*, 122, 12–17.
- Matadi, M. B. (2014). The SISRD epidemic model. *Far East Journal of Applied Mathematics*, 1, 1–14.
- Mazza, C., Ricci, E., Marchetti, D., Fontanesi, L., Di Giandomenico, S., Verrocchio, M. C., & Roma, P. (2020). How personality relates to distress in parents during the COVID-19 lockdown: The mediating role of child's emotional and behavioral difficulties and the moderating effect of living with other people. *International Journal of Environmental Research and Public Health*, 17, 6236.
- McMichael, A. J., Wilkinson, P., Kovats, R. S., Pattenden, S., Hajat, S., Armstrong, B., Vajanapoom, N., Niciu, E. M., Mahomed, H., Kingkeow, C., Kosnik, M., O'Neill, M. S., Romieu, I., Ramirez-Aguilar, M., Barreto, M. L., Gouveia, N., & Nikiforov, B. (2008). International study of temperature, heat and urban mortality: The “ISOTHERM” project. *International Journal of Epidemiology*, 37, 1121–1131.
- Miotto, R., & Weng, C. (2015). Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. *Journal of the American Medical Informatics Association*, 22, e141–e150.
- Miyake, J., Kaneshita, Y., Asatani, S., Tagawa, S., Niioka, H., & Hirano, T. (2018). Graphical classification of DNA sequences of HLA alleles by deep learning. *Human Cell*, 31, 102–105.
- Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*, 15, e85.

- Morgan, H. L. (1965). The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5, 107–113.
- Negassi, M., Suarez-Ibarrola, R., Hein, S., Miernik, A., & Reiterer, A. (2020). Application of artificial neural networks for automated analysis of cystoscopic images: A review of the current status and future prospects. *World Journal of Urology*, 38, 1–10.
- Nemati, M., Ansary, J., & Nemati, N. (2020). Machine-learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data. *Patterns*, 1, 100074.
- Osborne, T. F., Veigulis, Z. P., Arreola, D. M., Rössli, E., & Curtin, C. M. (2020). Automated EHR score to predict COVID-19 outcomes at US Department of Veterans Affairs. *PLoS One*, 15, e0236554.
- Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of COVID-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine*, 121, 103792.
- Paley, L., Zornitzki, T., Cohen, J., Friedman, J., Kozak, N., & Schattner, A. (2011). Utility of clinical examination in the diagnosis of emergency department patients admitted to the department of medicine of an academic hospital. *Archives of Internal Medicine*, 171, 1393–1400.
- Park, J.-S., Oh, Y. H., Ahn, S. C., & Lee, S.-W. (2005). Glasses removal from facial image using recursive error compensation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27, 805–811.
- Park, Y. J., Choe, Y. J., Park, O., Park, S. Y., Kim, Y.-M., Kim, J., Kweon, S., Woo, Y., Gwack, J., Kim, S. S., Lee, J., Hyun, J., Ryu, B., Jang, Y. S., Kim, H., Shin, S. H., Yi, S., Lee, S., Kim, H. K., ... COVID-19 National Emergency Response Center, Epidemiology and Case Management Team. (2020). Contact tracing during coronavirus disease outbreak, South Korea, 2020. *Emerging Infectious Diseases*, 26, 2465–2468.
- Parmar, C., Barry, J. D., Hosny, A., Quackenbush, J., & Aerts, H. J. (2018). Data analysis strategies in medical imaging. *Clinical Cancer Research*, 24, 3492–3499.
- Pawar, A. Y. (2020). Combating devastating COVID-19 by drug repurposing. *International Journal of Antimicrobial Agents*, 56, 105984.
- Peiris, J., Lai, S., Poon, L., Guan, Y., Yam, L., Lim, W., Nicholls, J., Yee, W., Yan, W., Cheung, M., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K., Yuen, K. Y., & SARS study group. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *The Lancet*, 361, 1319–1325.
- Poirier, C., Liu, D., Clemente, L., Ding, X., Chinazzi, M., Davis, J., Vespignani, A., & Santillana, M. (2020). Real-time forecasting of the COVID-19 outbreak in Chinese provinces: Machine learning approach using novel digital data and estimates from mechanistic models. *Journal of Medical Internet Research*, 22, e20285.
- Pollastri, G., Przybylski, D., Rost, B., & Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins: Structure, Function, and Bioinformatics*, 47, 228–235.
- Poole, L. (2020) Seasonal influences on the spread of SARS-COV-2 (COVID-19), causality, and forecastability (3-15-2020). *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.3554746>
- Poutanen, S. M., Low, D. E., Henry, B., Finkelstein, S., Rose, D., Green, K., Tellier, R., Draker, R., Adachi, D., Ayers, M., Chan, A. K., Skowronski, D. M., Salit, I., Simor, A. E., Slutsky, A. S., Doyle, P. W., Krajden, M., Petric, M., Brunham, R. C., & McGeer, A. J. (2003). Identification of severe acute respiratory syndrome in Canada. *New England Journal of Medicine*, 348, 1995–2005.
- Prata, D. N., Rodrigues, W., & Bermejo, P. H. (2020). Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil. *Science of the Total Environment*, 729, 138862.
- Qi, H., Xiao, S., Shi, R., Ward, M. P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X., & Zhang, Z. (2020). COVID-19 transmission in mainland China is associated with temperature and humidity: A time-series analysis. *Science of the Total Environment*, 728, 138778.
- Qin, L., Sun, Q., Wang, Y., Wu, K.-F., Chen, M., Shia, B.-C., & Wu, S.-Y. (2020). Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. *International Journal of Environmental Research and Public Health*, 17, 2365.
- Rahman, M. M., & Davis, D. N. (2013a). Addressing the class imbalance problem in medical datasets. *International Journal of Machine Learning and Computing*, 3, 224–228.
- Rahman, M. M., & Davis, D. N. (2013b). Machine learning-based missing value imputation method for clinical datasets. In *IAENG transactions on engineering technologies* (pp. 245–257). Springer.
- Rao, A. S. S., & Vazquez, J. A. (2020). Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine. *Infection Control & Hospital Epidemiology*, 41, 1–18.
- Rappuoli, R. (2000). Reverse vaccinology. *Current Opinion in Microbiology*, 3, 445–450.
- Relling, M. V., & Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature*, 526, 343–350.
- Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A. I., Etmann, C., McCague, C., Beer, L., Weir-McCall, J., Teng, Z., Gkrania-Klotsas, E., AIX-COVNET, Rudd, J., Sala, E., & Schonlieb, C. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*, 3, 199–217.
- Rogers, D., & Hahn, M. (2010). Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50, 742–754.
- Roma, P., Monaro, M., Muzi, L., Colasanti, M., Ricci, E., Biondi, S., Napoli, C., Ferracuti, S., & Mazza, C. (2020). How to improve compliance with protective health measures during the COVID-19 outbreak: Testing a moderated mediation model and machine learning algorithms. *International Journal of Environmental Research and Public Health*, 17, 7252.
- Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J., Yan, P., & Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, 5, 256–263.

- Rozenblum, R., & Bates, D. W. (2013). Patient-centred healthcare, social media and the internet: The perfect storm? *British Medical Journal Quality and Safety*, 22, 183–186.
- Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797–806). ACM.
- Ruiz, A., Lluch, A., Martín, M., Munárriz, B., Antón, A., Alba, E., Martínez Del Prado, P., Plazaola, A., Fernández-Aramburo, A., & Mahillo, E. (2005). Spanish breast cancer research group (geicam) population-based study on breast cancer outcomes: El álamo project (1990–1997). *Journal of Clinical Oncology*, 23, 585.
- Rutten-Jacobs, L. C., Larsson, S. C., Malik, R., Rannikmäe, K., Sudlow, C. L., Dichgans, M., Markus, H. S., & Traylor, M. (2018). Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: Cohort study of 306 473 UK biobank participants. *British Medical Journal*, 363. <https://doi.org/10.1136/bmj.k4168>
- Said, M. A., Verweij, N., & van der Harst, P. (2018). Associations of combined genetic and lifestyle risks with incident cardiovascular disease and diabetes in the UK biobank study. *Journal of the American Medical Association Cardiology*, 3, 693–702.
- Salama, M. A., Hassaniien, A. E., & Mostafa, A. (2016). The prediction of virus mutation using neural networks and rough set techniques. *European Association for Speech Signal and Image Processing Journal on Bioinformatics and Systems Biology*, 2016, 10.
- Salathé, M., Althaus, C. L., Neher, R., Stringhini, S., Hodcroft, E., Fellay, J., Zwahlen, M., Senti, G., Battegay, M., Wilder-Smith, A., Eckerle, I., Egger, M., & Low, N. (2020). COVID-19 epidemic in Switzerland: On the importance of testing, contact tracing and isolation. *Swiss Medical Weekly*, 150. <https://doi.org/10.4414/smw.2020.20225>
- Sanders, J. M., Monogue, M. L., Jodlowski, T. Z., & Cutrell, J. B. (2020). Pharmacologic treatments for coronavirus disease 2019 (COVID-19): A review. *Journal of the American Medical Association*, 18, 1824–1836.
- Senthilkumar, S., Rai, B. K., Meshram, A. A., Gunasekaran, A., & Chandrakumarmangalam, S. (2018). Big data in healthcare management: A review of literature. *American Journal of Theoretical and Applied Business*, 4, 57–69.
- Shi, P., Dong, Y., Yan, H., Zhao, C., Li, X., Liu, W., He, M., Tang, S., & Xi, S. (2020). Impact of temperature on the dynamics of the COVID-19 outbreak in China. *Science of the Total Environment*, 728, 138890.
- Silow-Carroll, S., Edwards, J. N., & Lashbrook, A. (2011). Reducing hospital readmissions: Lessons from top-performing hospitals. *Care Management*, 17, 14.
- Singh, R., Singh, R., & Bhatia, A. (2018). Sentiment analysis using machine learning technique to predict outbreaks and epidemics. *International Journal of Advanced Science and Research*, 3, 19–24.
- Suo, Q., Ma, F., Yuan, Y., Huai, M., Zhong, W., Zhang, A., & Gao, J. (2017). Personalized disease prediction using a cnn-based similarity learning method. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 811–816). IEEE.
- Sykes, D. L., Holdsworth, L., Jawad, N., Gunasekera, P., Morice, A. H., & Crooks, M. G. (2021). Post-COVID-19 symptom burden: What is long-COVID and how should we manage it? *Lung*, 199, 113–119.
- Tatonetti, N. P., Patrick, P. Y., Daneshjou, R., & Altman, R. B. (2012). Data-driven prediction of drug effects and interactions. *Science Translational Medicine*, 4, 125ra31–125ra31.
- Thomas, A. W., Müller, K.-R., & Samek, W. (2019). Deep transfer learning for whole-brain fMRI analyses. In *OR 2.0 context-aware operating theaters and machine learning in clinical neuroimaging* (pp. 59–67). Springer.
- Ting, D. S. W., Carin, L., Dzau, V., & Wong, T. Y. (2020). Digital technology and COVID-19. *Nature Medicine*, 26, 1–3.
- Tiwari, S., Kumar, S., & Guleria, K. (2020). Outbreak trends of coronavirus disease–2019 in India: A prediction. *Disaster Medicine and Public Health Preparedness*, 14, 1–6.
- Tobías, A., & Molina, T. (2020). Is temperature reducing the transmission of COVID-19? *Environmental Research*, 186, 109553.
- Tran, K., Cimon, K., Severn, M., Pessoa-Silva, C. L., & Conly, J. (2012). Aerosol generating procedures and risk of transmission of acute respiratory infections to healthcare workers: A systematic review. *PLoS One*, 7, e35797.
- Udugama, B., Kadhiresan, P., Kozłowski, H. N., Malekjahani, A., Osborne, M., Li, V. Y., Chen, H., Mubareka, S., Gubbay, J. and Chan, W. C. (2020) American Chemical Society Nano, Diagnosing COVID-19: The disease and tools for detection.
- Vahabi, N., Salehi, M., Duarte, J. D., Mollalo, A., & Michailidis, G. (2021). County-level longitudinal clustering of COVID-19 mortality to incidence ratio in the United States. *Scientific Reports*, 11, 1–22.
- Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18, 463–477.
- van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., Tamin, A., Harcourt, J. L., Thornburg, N. J., Gerber, S. I., Lloyd-Smith, J. O., Wit, E., & Munster, V. J. (2020). Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England Journal of Medicine*, 382, 1564–1567.
- Vikram, K., & Padmavathi, S. (2017). Facial parts detection using viola jones algorithm. In *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1–4). IEEE.
- Wagner, T., Shweta, F., Murugadoss, K., Awasthi, S., Venkatakrishnan, A., Bade, S., Puranik, A., Kang, M., Pickering, B. W., O'Horo, J. C., Bauer, P. R., Razonable, R. R., Vergidis, P., Temesgen, Z., Rizza, S., Mahmood, M., Wilson, W. R., Challener, D., Anand, P., ... Soundararajan, V. (2020). Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis. *eLife*, 9, e58227.
- Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *International Journal of Environmental Research and Public Health*, 17, 1729.

- Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *Journal of the American Medical Association*, 323, 1061–1069.
- Wang, J., Lu, K., Pan, D., He, N., & Bao, B.-K. (2014). Robust object removal with an exemplar-based image inpainting approach. *Neurocomputing*, 123, 150–155.
- Wang, L., You, Z.-H., Chen, X., Xia, S.-X., Liu, F., Yan, X., Zhou, Y., & Song, K.-J. (2018). A computational-based method for predicting drug–target interactions by using stacked autoencoder deep neural network. *Journal of Computational Biology*, 25, 361–373.
- Wang, P., Xiao, X., Glissen Brown, J. R., Berzin, T. M., Tu, M., Xiong, F., Hu, X., Liu, P., Song, Y., Zhang, D., Yang, X., Li, L., He, J., Yi, X., Liu, J., & Liu, X. (2018). Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nature Biomedical Engineering*, 2, 741–748.
- Wang, P., Zheng, X., Ai, G., Liu, D., & Zhu, B. (2020). Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. *Chaos, Solitons & Fractals*, 140, 110214.
- Wang, S., Ren, F., & Lu, H. (2018). A review of the application of natural language processing in clinical medicine. In *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 2725–2730). IEEE.
- Wei, Q., Ji, Z., Li, Z., du, J., Wang, J., Xu, J., Xiang, Y., Tiryaki, F., Wu, S., Zhang, Y., Tao, C., & Xu, H. (2020). A study of deep learning approaches for medication and adverse drug event extraction from clinical text. *Journal of the American Medical Informatics Association*, 27, 13–21.
- Wen, L., Li, X., & Gao, L. (2019). A transfer convolutional neural network for fault diagnosis based on resnet-50. *Neural Computing and Applications*, 32, 1–14.
- Wen, M., Zhang, Z., Niu, S., Sha, H., Yang, R., Yun, Y., & Lu, H. (2017). Deep-learning-based drug–target interaction prediction. *Journal of Proteome Research*, 16, 1401–1409.
- White, T. (2012). *Hadoop: The definitive guide*. O'Reilly Media, Inc.
- Wu, J., Wang, W., Zhang, J., Zhou, B., Zhao, W., Su, Z., Gu, X., Wu, J., Zhou, Z., & Chen, S. (2019). DeepHLA: A deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity. *Frontiers in Immunology*, 10, 2559.
- Wu, Y., Feng, Z., Li, P., & Yu, Q. (2020). Relationship between ABO blood group distribution and clinical characteristics in patients with COVID-19. *Clinica Chimica Acta*, 509, 220–223.
- Xie, X., Zhong, Z., Zhao, W., Zheng, C., Wang, F., & Liu, J. (2020). Chest CT for typical 2019-nCoV pneumonia: Relationship to negative RT-PCR testing. *Radiology*, 296, E41–E45. <https://doi.org/10.1148/radiol.202000343>
- Xu, X., Jiang, X., Ma, C., du, P., Li, X., Lv, S., Yu, L., Ni, Q., Chen, Y., Su, J., Lang, G., Li, Y., Zhao, H., Liu, J., Xu, K., Ruan, L., Sheng, J., Qiu, Y., Wu, W., ... Li, L. (2020). A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering*, 6, 1122–1129.
- Zarikas, V., Pouloupoulos, S. G., Gareiou, Z., & Zervas, E. (2020). Clustering analysis of countries using the COVID-19 cases dataset. *Data in Brief*, 31, 105787.
- Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-series data: A comparative study. *Chaos, Solitons & Fractals*, 140, 110211.
- Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R. R., Zhebrak, A., Minaeva, L. I., Zagribelnyy, B. A., Lee, L. H., Soll, R., Madge, D., ... Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37, 1038–1040.
- Zhou, D.-X. (2020). Theory of deep convolutional neural networks: Downsampling. *Neural Networks*, 124, 319–327.
- Zoffmann, S., Vercruyse, M., Benmansour, F., Maunz, A., Wolf, L., Marti, R. B., Heckel, T., Ding, H., Truong, H. H., Prummer, M., Schmucki, R., Mason, C. S., Bradley, K., Jacob, A. I., Lerner, C., Rosario, A. A., Burcin, M., Amrein, K. E., & Prunotto, M. (2019). Machine learning-powered antibiotics phenotypic drug discovery. *Scientific Reports*, 9, 1–14.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Dasgupta, A., Bakshi, A., Mukherjee, S., Das, K., Talukdar, S., Chatterjee, P., Mondal, S., Das, P., Ghosh, S., Som, A., Roy, P., Kundu, R., Sarkar, A., Biswas, A., Paul, K., Basak, S., Manna, K., Saha, C., Mukhopadhyay, S., ... De, R. K. (2022). Epidemiological challenges in pandemic coronavirus disease (COVID-19): Role of artificial intelligence. *WIREs Data Mining and Knowledge Discovery*, 12(4), e1462. <https://doi.org/10.1002/widm.1462>