

## **Modern Statistical Methodologies in the COVID–19 Pandemic: A Review from a Statistical Perspective**

**Das, A. and Muralidharan, K.**

*Department of Statistics, Faculty of Science,  
The Maharaja Sayajirao University of Baroda, Vadodara*

Received: 28 May 2021; Revised: 14 June 2021; Accepted: 16 June 2021

---

### **Abstract**

Ever since the world has been introduced to the novel corona virus, we have been privy to multiple statistical, epidemiological, and deep learning models to understand the biological prevalence of the disease. Neural Networks having played a pivotal part in this sphere, aimed at varied objectives. The most common among them have been (1) forecasting of transmission dynamics based on daily data, (2) classification of infected and non-infected patients based on various scans and images and (3) to analyze survival data to determine incubation period / time to death / duration of hospitalization, or to analyze competing risks, among others. This paper aims to showcase some of those Neural Networks that have been crucial in either handling or monitoring or in some cases, providing useful insight into the pandemic.

*Key words:* Corona virus; Pandemic; Artificial Neural network; Artificial Intelligence; Machine Learning; Deep Learning; Classification; Regression modeling.

---

### **1. Introduction**

The advent of the novel corona virus, or technically SARS-CoV-2, or as it is more popularly known, COVID – 19, in Wuhan, Hubei province, People’s Republic of China, as reported by the Wuhan Municipal Health Commission, on December 31, 2019, is a global phenomenon today, with more than 154 million infected, 90.3 million recovered and 3.22 million deaths to its name, as per the World Health Organization, as of May 4, 2021. During this time, a lot of time, effort, and money has been allotted by organizations and governments worldwide, to mitigate numerous micro-tragedies occurring in spurts across the globe.

A lot of research facilities around the world have invested their resources in developing epidemiological, statistical, and mathematical models to handle crises both at the micro as well as at the macro level, or to monitor transmission dynamics of the virus, or to gain some insight regarding how the virus might spread, the virus' mutation dynamics, patient's survivability in case of mild-to-moderate-to-severe lung damage, optimization of existing tests based on patient vitals, among others. An effort has also been made to monitor and subsequently, minimize the spread of false information in this time of crisis. The second wave of the disease has further boosted the necessity of model-based study to support the transactional issues of the pandemic worldwide.

The neural network model, especially the deep neural network model, since its inception, or more appropriately, since its computational barriers have subsided, has been in

the limelight for its prediction accuracy and applicability across varied data types. In practice, a deep neural network is any neural network having a large number of layers of neurons. Neural networks are a set of algorithms, modeled loosely after the human brain, that is designed to recognize patterns. They interpret sensory data through various kinds of machine perception, labeling, or clustering of raw inputs. The patterns they recognize are numerical, contained in vectors, into which all real-world data, be it images, sound, text, or time series, must be translated. This motivated us to make a review of neural network-based models available so far in statistical literature in understanding the behavior pattern of the COVID-19 pandemic and its mitigation potential.

## 2. Literature Review

The review is attempted in two statistical perspectives here. In the immediate section, we explore the modern-day techniques involving neural network-related studies and examine their potential in describing the virus patterns. We also extract those studies where the use of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) techniques are effectively used for modeling the pandemic. Section 2.2 will explore the statistical studies which are attempted using probability models and inferences therein.

It may be noted that most of the papers on models addressing transmission dynamics of Covid-19 are associated the reproduction number,  $R$ , which is a key epidemiological parameter that quantifies the average number of new infections caused by a single infected individual. When a pathogen emerges in an entirely susceptible population, this parameter is referred to as the basic reproduction number,  $R_0$ .

Utmost care is taken to establish the hierarchy of the published articles.

### 2.1. Neural network models in the pandemic

The paper by Chatterjee *et al.* (2020) addresses the potential of data science to assess the risk factors correlated with COVID-19, after analyzing existing data sets available in “ourworldindata.org” and newly simulated datasets, following the analysis of different univariate Long Short-Term Memory(LSTM) models for forecasting new cases and resulting deaths. They discussed the findings related to the statistical analysis on simulated datasets. In a related paper, Pirouz *et al.* (2020) aim to study the possible correlation between the numbers of swab tests and the trend of confirmed cases of infection, while paying particular attention to the sickness level. The statistical analysis of this article provided the basis for an AI study by Artificial Neural Network (ANN). Also, they use multivariate linear regression approach to validate their findings.

Shams *et al.* (2020) have investigated the importance of using Generative Adversarial Networks (GAN) as a preprocessing stage to applied DNN (DotNetNuke) for image data augmentation. They presented a case study of using GAN networks for limited COVID-19 X-Ray Chest images. The results indicated that the proposed system based on using GAN-DNN is powerful with a minimum loss function for detecting COVID-19 X-Ray Chest images. Stochastic gradient descent and Improved Adam optimizers are used during the training process of the COVID-19 X-Ray images, and the evaluation results depend on loss function are determined to ensure the reliability of the proposed GAN architecture.

In the study by Yasar and Ceylan (2020), which aims at early diagnosis of Covid-19 disease using X-ray images, the deep-learning approach, a state-of-the-art artificial intelligence method, was used, and automatic classification of images was performed using Convolutional Neural Networks(CNNs). In the first training-test data set there were 230 X-ray images, of which 150 were Covid-19 and 80 were non-Covid-19, while in the second training-test data set there were 476 X-ray images, of which 150 were Covid-19 and 326 were non-Covid-19. Thus, classification results have been provided for two data sets, containing predominantly Covid-19 images and predominantly non-Covid-19 images. A23-layer CNN architecture and a 54-layer CNN architecture were developed, of which results were obtained in the form of chest X-ray images directly in the training-test procedures and the sub-band images obtained by applying dual-tree complex wavelet transform (DT-CWT) to the above-mentioned images. The same experiments were repeated using the image obtained by applying the local binary pattern to the chest X-ray images. Within the scope of the study, four new result generation pipeline algorithms having been put forward additionally, it was ensured that the experimental results were combined and the success of the study was improved. The training sessions were carried out using a23-fold cross-validation method for both datasets.

The aim of the study by Apostolopoulos & Mpesiana (2020) was to evaluate the performance of state-of-the-art CNN architectures proposed for the classification of medical images. A dataset of X-ray images from patients with common bacterial pneumonia confirmed Covid-19 disease, and normal incidents were utilized for the automatic detection of Covid-19. Following two datasets have been utilized in this experiment:

- A collection of 1427 X-ray images including 224 images with confirmed Covid-19 disease, 700 images with confirmed common bacterial pneumonia, and 504 images of normal conditions.
- A dataset including 224 images with confirmed Covid-19 disease, 714 images with confirmed bacterial and viral pneumonia, and 504 images of normal conditions.

The data was collected from the available X-ray images on public medical repositories. The results suggest that Transfer Learning with X-ray imaging may extract significant biomarkers related to the Covid-19 disease, while the best accuracy, sensitivity, and specificity obtained are 96.78%, 98.66%, and 96.46% respectively.

In another related study by Ozturk *et al.* (2020), a new model for automatic COVID-19 detection using raw chest X-ray images was presented. The proposed study is developed to provide accurate diagnostics for binary classification (COVID vs. No-Findings) and multi-class classification (COVID vs. No-Findings vs. pneumonia). The model produced a classification accuracy of 98.08% for binary classes and 87.02% for multi-class cases. The DarkNet model was used in their study as a classifier for the You Only Look Once (YOLO) real-time object detection system. They implemented 17 convolutional layers and introduced different filtering on each layer. Their model can be employed to assist radiologists in validating their initial screening, and can also be employed via the cloud to immediately screen patients.

Cheema *et al.*(2020) designed an intelligent computing paradigm called the Levenberg-Marquardt artificial neural networks (LMANNs) as an attempt to solve the mathematical model SEIPAHRF (Susceptible, Exposed, Infected and symptomatic, super Propagation, infected but Asymptomatic, Hospitalized, Recovered, Fatality)of Coronavirus disease propagation via human to human interaction. The model is represented with systems of nonlinear ordinary differential equations represented with susceptible, exposed,

symptomatic and infectious, super spreaders, infection but asymptomatic, hospitalized, recovery, and fatality classes. The reference dataset of the COVID-19 model is then generated by exploiting the strength of the explicit Runge-Kutta numerical method for metropolitans of China and Pakistan. The created dataset is arbitrarily used for training, validation, and testing processes for each cyclic update in Levenberg-Marquardt's back-propagation for the numerical treatment of the dynamics of the COVID-19 model. The effectiveness and reliable performance of the design LMANNs are endorsed based on assessments of achieved accuracy in terms of mean squared error-based merit function, error histograms, and regression studies. In a separate study, Singh *et al.* (2020) designed and implemented another deep convolutional neural network approach for the data. They tuned the hyper-parameter of CNN using Multi-objective adaptive differential evolution (MADE). Considering the benchmark COVID-19 dataset, extensive experiments were performed. In comparison, it was found that the proposed technique outperforms the competitive machine learning models in terms of various performance metrics.

Li *et al.* (2020) proposed an automated measure of COVID-19 pulmonary disease severity on chest radiographs, for longitudinal disease tracking and outcome prediction based on a convolutional Siamese neural network-based algorithm, which was trained to output a measure of pulmonary disease severity on CXRs. Severity scores based on Siamese neural networks automatically measure radiographic COVID-19 pulmonary disease severity that can be used to track disease change and predict deaths. Before this paper, studies used manually annotated features from chest imaging to predict outcomes, such as mortality, need for intensive care, and other adverse events. However, barriers to the adoption of these systems include inter-rater reliability and the learning curve for users. In this study, raters assessing longitudinal change showed only moderate inter-rater agreement. The automated Siamese neural network-based approach addresses these challenges.

Ardakani *et al.* (2020) suggested a rapid and valid method for COVID-19 diagnosis using multiple AI techniques and then compared them. Ten well known convolutional neural networks were used to distinguish infection of COVID-19 from non-COVID-19 groups: AlexNet, VGG-16, VGG-19, SqueezeNet, GoogleNet, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101 and Xception. Among all the networks, the best performance was achieved by ResNet-101 and Xception. Although the Xception network gave the best performance, it did not have the best sensitivity. In contrast, ResNet-101 could diagnose COVID-19 infection with the highest sensitivity and implied lower specificity compare to the Xception network. The authors suggest that a method with the highest sensitivity to diagnose all patients with COVID19 is desirable. In this regard, the ResNet-101 has an advantage over other networks, especially Xception, due to its highest sensitivity and AUC. ResNet-101 is trained based on residual learning. This kind of learning can facilitate the training of networks by considering the layer inputs as a reference. In addition, residual networks are optimized easier, and the accuracy can be improved as the depth increases, as confirmed by He *et al.* (2016). This specific residual learning can lead to better training and provide a robust model. Hence, the best performance can be obtained using ResNet-101.

Cough analysis is another important area in clinical diagnosis. Bansal *et al.* (2020) proposed a CNN-based audio classifier using the open cough dataset. The dataset is labeled manually into cough categories with the final labeling of COVID and Non-COVID classes. Two approaches proposed in this paper are based on Mel-frequency Cepstral Coefficients (MFCC) features and spectrogram images as input to the CNN network. They found that the MFCC approach produced 70.58% test accuracy with 81% sensitivity and is better than the spectrogram-based approach. The authors opine that the methodology of their

paper is based on a novel CNN architecture with MFCC input and that being a contactless methodology, it serves as a first-hand assessment of the Covid infection for masses with a Smartphone along with other symptoms. Similar work has been seen in the paper by Hassan *et al.* (2020), where they highlight the importance of speech signal processing in the process of early screening and diagnosing the COVID-19 virus by utilizing the LSTM for analyzing the acoustic features of cough, breathing, and voice of the patients. Their result shows that the accuracy of the voice test compared to both coughing and breathing sound is lower.

Haque *et al.* (2020) proposed another CNN model to detect COVID-19 patients from chest X-ray images. This model is evaluated with a comparative analysis of two other CNN models. The architecture of this CNN is such that it has four convolutional layers, the first one being a 2D convolutional layer with  $3 \times 3$  kernels and a Rectified Linear Unit (ReLU) activation function. The next three layers are 2D convolutional layers along with the ReLU activation function and Max pooling. Max pooling accumulates the features of the convolutional layer by convolving filters over it. In each of the three layers, a  $2 \times 2$  Max pooling layer is added after the convolutional layer to avoid overfitting and to make the model computationally efficient. Then, the output of the convolutional layers is converted to a long 1D feature vector by a flattened layer. This output from the flattened layer is fed to the fully connected layer with dropout. In a fully connected layer, every input neuron is connected to every activation unit of the next layer. All the input features are passed through the ReLU activation function and this layer categorizes the images to the assigned labels. The Sigmoid activation function makes the classification decision depending on the classification label of the neurons. Finally, in the output layer, it is declared if the input X-ray image is COVID-19 positive or normal. The proposed model performs with an accuracy of 97.56% and a precision of 95.34%.

Govindarajan and Swaminathan (2020) aimed to differentiate COVID-19 conditions from healthy subjects in chest radiographs using simplified end-to-end CNN model and occlusion sensitivity maps. The images were considered from publicly available datasets. In terms of architecture, each convolutional layer was made of a varied number of filters as 8, 16, 32, 64, 128, 256, and 512 with varying filter sizes  $3 \times 3$  and  $5 \times 5$ . The convolutions were all zero-padded to preserve the input resolution. Further, each convolutional 2D layer was followed by a batch normalization layer and ReLU activation layers. An epsilon value of  $1e-5$  was chosen for batch normalization. Further, a max-pooling layer was followed with a stride of 2 and a pool size of 2. Resultant feature maps were fed to a fully connected layer, followed by the Softmax layer followed by yet another classification layer. In this work, the output size of 2, corresponding to the two classes was fixed. The weights were initialized using the Glorot method on all layers where each weight was initialized from  $N(0, k)$ ,  $k < \infty$ . Finally, the classification scores are obtained to generate visualization-based occlusion sensitivity maps. Significant biomarkers representing critical image features were extracted from the CNN by experimentally investigating cross-validation methods and hyperparameter settings. Occlusion Sensitivity Maps work when the occluded images are passed through the CNN network and Euclidean distances are calculated for each iteration. The differences between the occluded distances and nonoccluded distances are computed. This difference increases once a patch occludes an area in the image relevant to the network, thus creating a heat map. The pathological regions correspond to higher probability and a drop in the value indicates that the pathological locations have been occluded. In this study, a mask size of 15 and a stride value of 10 pixels are heuristically chosen. Thus, COVID-19 specific CXR biomarkers can be localized providing an approximate visual diagnosis. The performance of this network was evaluated using standard metrics. Results describe that the simplified CNN model with

optimized parameters can extract significant features with a sensitivity of 97.35% and F-measure of 96.71% to detect COVID-19 images.

The primary innovation of the method proposed by Hartono (2020) for developing a transmission dynamics predictor that takes advantage of the time difference among many countries with respect to the transmission of this disease is that the proposed method only requires the transmission similarities between countries in the publicly available data for this current disease. For this purpose, two models have been used, the topical autoencoder (TA), a hierarchical neural network having a low-dimensional topological hidden layer that is a simplified version of the soft-supervised topological autoencoder (STA) proposed in Hartono (2020), in which the basic mathematical properties were proposed in Hartono *et al.* (2015), and a vanilla LSTM.

Hawas (2020) used RNNs to predict the daily infections in Brazil. His article introduces long-term time-series predictions for the virus's daily infections in Brazil by training forecasting models on limited raw data. The prediction data was generated by training 4200 recurrent neural networks (54 to 84 day's validation periods) on raw data. Militante *et al.* (2020) tried to find out if the patient has these diseases; experts conduct physical exams and diagnose their patients through Chest X-ray, ultrasound, or biopsy of lungs. The study employs a flexible and efficient approach of deep learning applying the model of CNN in predicting and detecting a patient unaffected and affected with the disease employing a chest X-ray image. The study utilized a collected dataset of 20,000 images using a 224x224 image resolution with a 32 batch size is applied to prove the performance of the CNN model being trained. Their trained model gives a 95% accuracy rate. Also, they can try to detect and predict COVID-19, bacterial, and viral-pneumonia diseases based on chest X-ray images.

The ANN-based models proposed in Niazkar and Niazkar (2020) were utilized to estimate the confirmed cases of COVID-19 in China, Japan, Singapore, Iran, Italy, South Africa, and the United States of America. These models exploit historical records of confirmed cases, while their main difference is the number of days that they assume to have an impact on the estimation process. Their results show that the ANN-based model that takes into account the previous 14 days outperforms the other ones. This comparison reveals the importance of considering the maximum incubation period in predicting the COVID-19 outbreak.

Sridhar *et al.* (2020) have analyzed the similarity in features between the novel coronavirus 2019 and various other lung diseases such as Pneumonia, Pneumothorax, Atelectasis, Pleural Thickening, etc. based on Chest X-ray scans in the posteroanterior view for various diseases. CNN using the Residual Network (ResNet) is built to identify similar regions in the chest X-rays of COVID-19 and various lung diseases. The regions of similarity are visualized using class activation maps. A total of eleven conditions affecting the lungs are studied and compared to COVID-19. Their results show that Atelectasis, Consolidation, Emphysema, and Pneumonia are most similar to COVID-19 of the eleven diseases considered.

Toraman *et al.* (2020) have used some deep learning models like CNN, DNN, etc. which are mostly used in radiology images to determine the positive cases. They studied a novel artificial neural network, Convolutional Capsule Neural Network (CapsNet) for the detection of COVID-19, and further proposed using chest X-ray images with capsule networks. The proposed approach is designed to provide fast and accurate diagnostics for

COVID-19 diseases with binary classification (COVID-19, and No-Findings), and multi-class classification (COVID-19, and No-Findings, and Pneumonia). Their proposed method achieved an accuracy of 97.24% and 84.22% for binary class and multiclass respectively.

## 2.2. Other important models and methodologies

Li *et al.* (2020) introduced a statistical disease transmission model using case symptom onset data to estimate the transmissibility of the early phase outbreak in china and provide sensitivity analyses with various assumptions of diseases. They fitted the transmission model to several publicly available sources of the outbreak data until 11, Feb 2020, and estimated lockdown intervention efficacy of Wuhan city. The estimated  $R_0$  was between 2.7 and 4.2 from plausible distribution assumptions of the incubation period and relative infectivity. They finalized on the most agreed  $R_0$  value of 2.2. Then they compared different modelling approaches. On comparison they found that, there was no negative correlation in the spatial distribution of Covid-19 incidence, and there were no high-value clusters.

In their article, Martin *et al.* (2020) have summarized some of the most important pitfalls, provided intuitive clinical examples (mostly related to other types of infections), and referred to methodological articles explaining the mathematical frameworks of some of the most common neural networks. Their article also describes the common errors which may occur in the presence of competing risks and time-dependent covariates, or when conditioning on future events. They have used Kaplan–Meier survival plots, Cox or logistic regression models as well for comparative purposes.

Wang *et al.* (2020) study publicly released data for 1212 COVID-19 patients in Henan, China using various statistical probabilistic models and network analysis. The incubation period was statistically estimated, and they proposed a state transition diagram that could explore the epidemic stages of emerging infectious diseases. They found that although the quarantine measures were gradually working, strong measures still might have been required for a while, since 7.45% of patients may have had very long incubation periods. They also found that migrant workers and/or college students were at high risk. A related paper by Sivapalan *et al.* (2020) describes the detailed statistical analysis plan for the evaluation of primary and secondary endpoints of the proPAC-COVID study. The purpose of this paper is to provide primary publication of study results to prevent selective reporting of outcomes, data-driven analysis, and to increase transparency. The proPAC-COVID is a multi-center, randomized, placebo-controlled, double-blinded clinical trial. Data is analyzed using intention-to-treat (ITT) principles, and main analyses will also be subject to modified ITT analysis and per-protocol analysis. The paper by Talmoudi *et al.* (2020) describes the transmission dynamics of the outbreak and the impact of intervention measures that are critical to planning responses to future outbreaks and providing timely information to guide policymakers' decisions. They estimated the serial interval and temporal reproduction number of SARS-CoV-2 in Tunisia. They used the maximum likelihood approach to estimate the dynamics of  $R_t$ . They collected data of investigations and contact tracing between March 1, 2020, and May 5, 2020, as well as illness onset data during the period February 29-May 5, 2020 from the National Observatory of New and Emerging Diseases of Tunisia.

The findings of Shruthi and Ramani (2020) signify that the statistical volatility transmission differs post food price crisis. Moreover, risk transmission materializes as an additional element of the statistical interrelations among agricultural and energy markets. The major contributions of the article are:

- The various qualities of volatility spillover impacts among each commodity market and crude oil after and before the COVID-19 are evaluated systematically, and the impact of crude oil fluctuations in the prices on different commodity markets is being compared.
- The statistical conditional correlation between each commodity market and the crude oil market is significantly high. The time differing qualities of correlations after and before the COVID-19 are also significant.

Phucharoen *et al.* (2020) have analyzed Thailand's COVID-19 data using the probit model to investigate the risk factors for transmission from confirmed COVID-19 cases to their high-risk contacts. The analysis further focused on the impact of quarantine measures in state-provided facilities on contacts' probability of infection and interpreted that the quarantine policy, which mandated individual isolation in the state provided facilities for all high-risk contacts, diminished the contact's chance of infection from the confirmed cases, especially in the epicenter districts. A comparative study paper by Nayak *et al.* (2020) has made some statistical analysis based on Gaussian modeling, ANOVA test, and probabilistic model approach. After ANOVA they conclude that the recovery rate for all the countries is significantly higher than the mortality rate except for the UK where the mortality rate is significantly higher than the recovery rate. Gaussian modeling applied here was able to predict the original peak values of confirmed cases of countries. Using the probabilistic model approach, the authors were able to predict that there is around a 5% probability that a person in India will be tested positive for COVID-19 on 100 tests.

Milano and Cannataro (2020) analyzed different Italian COVID-19 data at the regional level for the period 24th Feb to 29th Mar 2020. They built multiple similarity matrices using statistical tests and then they mapped those similarity matrices into networks where nodes represent Italian regions. Then the network-based analysis of Italian COVID-19 data can elegantly show how regions form communities, *i.e.*, how they join and leave them according to time and how community consistency changes over time and with respect to the different available data. They found that the regions where the epidemic had a greater impact, such as the Lombardi, Veneto, Piemonte and Emilia, had a different behaviour with respect to other regions. This was evident in the community detection in which the regions most affected by the epidemic form individual communities or they are part of the same community. In addition, this study also led to identifying similar behaviours of regions that are geographically distant but that together form community. Another interesting finding is given by Hamzah *et al.* (2020), where the real-time data query is done and visualized; the queried data is used for Susceptible-Exposed-Infectious-Recovered (SEIR) predictive modeling. They utilized SEIR modeling to forecast COVID-19 outbreaks within and outside China based on daily observations. They also analyzed queried news, and classified the news into negative and positive sentiments, to understand the influence of the news on people's behavior both politically and economically.

Li *et al.* (2020) collected available data with respect to all confirmed Covid-19 cases in China and analyzed the spatial patterns of these cases by a Geographical Information Systems (GIS) approach applying ArcGIS software to create a spatial database, after which they applied the ArcGIS geographic statistical analysis module to investigate the spatial distribution trends. Also, Faes *et al.* (2020) investigated the time from symptom onset to diagnosis and hospitalization or the length of stay in the hospital and whether there are differences in the population. The distribution of different event times of different patients



groups is estimated accounting for interval censoring and right truncation of the time intervals.

An interval-censored likelihood framework was adopted in the paper by Zhao *et al.* (2020) to fit three different distributions including Gamma, Weibull, and Lognormal that govern the serial interval of COVID-19. They select the distribution according to the AIC (corrected) for a small sample size. They found that the lognormal distribution performed slightly better than the other two distributions in terms of AIC. Another model-based paper by Livadiotis (2020) focuses on the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19 for US and Italian regions. Also, the researcher has derived growth rates for regions characterized by a readable exponential growth phase and has plotted them against the environmental temperatures averaged within the same regions. This is used to derive the relationship between temperature and growth rate and has evaluated its statistical confidence.

Using simulated epidemic data, O'Driscoll *et al.* (2020) have assessed the performance of seven commonly used statistical methods to estimate  $R_0$  as they would be applied in real-time outbreak analysis scenario, fitting to an increasing number of data points over time and with varying levels of random noises in the data. Method comparison was also conducted on empirical outbreak data, using Zika surveillance data. On early analysis of the SARS-CoV-2 epidemic in China, they estimated a declining trend in the reproduction number from 7.93 (95% CI: 5.00-12.00) on the 29th December 2019 to 2.60 (95% CI: 0.57-5.17) on the 18th January 2020 using the Wallinga and Teunis method. Most methods considered here frequently over-estimate  $R_0$  in the early stages of epidemic growth on simulated data.

### 3. Open Research Areas and the Future of the Pandemic

Over the last few years, deep neural networks have proven superior over most other models in terms of accuracy and forecasting ability. However, there are arguments against their use, including the fact that they do not play any role in identifying key relationships between variables, among other critiques mentioned in the work of Zador (2019).

One more critique in the case of neural networks may be in the way it is used by researchers worldwide. The fact that the architecture of the ANN is a defining feature of the model itself provides an opportunity for researchers to employ known combinations of varied architecture with considerable ease and, regrettably in some cases, even without proper reasoning, the primary defence of which extends to higher accuracy and architecture novelty. This may be the primary reason behind why so many CNN's have been discussed in this paper, most of which are concerned with the classification of patients into infected/non-infected or COVID-19/Pneumonia/Other Pulmonary Conditions, based on CT-Scan or X-ray Images.

From the works of Shams *et al.* (2020), we can see that they plan to use a Region of Interest feature extractor instead of the whole image to extract more details for medical images. Yasar and Ceylan (2020) planned to analyze the effects of using the obtained results through direct transfer learning in pipeline classification algorithms, on the study results. Ozturk *et al.* (2020), as well as Toraman *et al.* (2020), intend to validate their model by incorporating more images. The resultant model may be placed in a cloud to provide instant diagnosis and help in the rehabilitation of affected patients quickly. Haque *et al.* (2020) have postulated that in addition to improvements in the model, clinical testing shall prove to be helpful in diagnosis. It is also stated in the work by Militante *et al.* (2020) that further studies

may further improve the CNN architecture by hyper-parameter tuning and transfer learning combinations.

Aside from the suggestions put forward by the authors cited in this section, one may engage in trying to combine the results from CT-Image-based classification and X-Ray based classification. One may even think of developing an ensemble learning algorithm, combining the two models, thereby introducing more information as well as complexity. Keep in mind that the choice of architecture may be the key here. One may even think of developing a unanimous CNN with multiple input gates, to accommodate both types of images as input. Another alternative would be to accommodate bivariate data in the CNN, obviously, after pixel-correction of any one of the two images.

As far as transmission dynamics are concerned, we have already seen several RNNs and Time Series Models which have been developed during the pandemic, trying to estimate the same with great accuracy. We have found that a lot of focus is being spent on vanilla LSTMs, and on consolidating them with various other supervised and unsupervised learning pipelines. Seldom have we seen any set of researcher(s) publish advanced versions of LSTMs, although, it may have happened that advanced LSTMs might not have reaped the desired accuracy. A viable alternative approach to this problem might lie in developing Hybridized Time Series Models and seeing if it provides greater accuracy.

Aside from the Neural Network realm, Milano and Cannataro (2020) plan to extend their study considering the evolution of the communities at greater time intervals to demonstrate a new pattern of regions with respect to COVID-19 data. Shruthi and Ramani (2020) state that future investigations may expand their writing in any event in two different ways. Primarily, inquiries about whether profit by the adaptability of auxiliary vector auto regression models, force hypothetical limitations to recognize effects of vitality and money showcase on agrarian markets. Furthermore, multivariate unpredictability overflow examination which gives space to apply causality in fluctuation tests and drive reaction investigation dependent on multivariate models may give new experiences.

We can see that, not many researchers have focused on panel data modelling when addressing the pandemic itself; however, owing to the increasing availability of data, as time passes, the prospect might be even more attractive to the point that, a lot of statisticians may be working on it right now. Network or Random Graph Models have also been approached by a handful of researchers, but all of them have mentioned explicitly, that lack of data may be a contributing factor to somewhat poor results. Attempting to adopt such methodologies might prove fruitful now, although the reliability of most data sources that are used to generate similarity matrices may be an issue of concern. The standard SIR/SEIR model has been reproduced quite frequently during this pandemic to trace transmission dynamics. But it is also noteworthy, that most papers pertaining to this model have restricted themselves to simply adding intervening events as model enhancements. One might look into combining SEIR models with Latent Variable Models and other advanced models to gain better insight.

In overview, the statistician today, has the massive responsibility to, firstly, seamlessly combine conventional statistical techniques and machine learning and deep learning tools and techniques, and to act as the go-to-entity for validation of statistical as well as ML and DL oriented models/analyses. They must also ensure that modern ML and DL techniques are treated with a strong statistical acumen.

### **Acknowledgments**

The authors thankfully acknowledge the support from the referee and the Chair Editor.

## References

- Apostolopoulos, I. D. and Mpesiana, T. A. (2020). Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 1-6.
- Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N. and Mohammadi, A. (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Computers in Biology and Medicine*, 103795.
- Bansal, V., Pahwa, G. and Kannan, N. (2020). Cough Classification for COVID-19 based on audio mfcc features using convolutional neural networks. *IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, 604-608.
- Chatterjee, A., Gerdes, M. W. and Martinez, S. G. (2020). Statistical explorations and univariate timeseries analysis on COVID-19 datasets to understand the trend of disease spreading and death. *Sensors*, **20**, 11-39.
- Cheema, T. N., Raja, M. A., Ahmad, I., Naz, S., Ilyas, H. and Shoaib, M. (2020). Intelligent computing with Levenberg-Marquardt artificial neural networks for a nonlinear system of COVID-19 epidemic model for future generation disease control. *The European Physical Journal Plus*, **11**, 1-35.
- Faes, C., Abrams, S., Beckhoven, D. V., Meyfroidt, G., Vlieghe, E., Hens, N. and Surveillance, B. C. (2020). Time between symptom onset, hospitalisation and recovery or death: Statistical analysis of Belgian COVID-19 patients. *International Journal of Environmental Research and Public Health*, **17**, 7560.
- Govindarajan, S. and Swaminathan, R. (2020). Differentiation of COVID-19 conditions in planar chest radiographs using optimized convolutional neural networks. *Applied Intelligence*, 1-12.
- Hamzah, F. A., Lau, C. H., Nazri, H., Ligot, D. V., Lee, G., Tan, C. L., . . . Emmaanue, R. (2020). Corona Tracker: Worldwide COVID-19 outbreak Data Analysis and Prediction. *Bulletin of the World Health Organisation*, **1**, 32.
- Haque, K. F., Haque, F. F. and Gandy, L. (2020). Automatic detection of COVID-19 from chest X-ray images with convolutional neural networks. *2020 International Conference on Computing, Electronics and Communications Engineering (iCCECE)*, 125-130.
- Hartono, P. (2020). Mixing autoencoder with classifier: Conceptual data visualization. *IEEE Access*, **8**, 105301-105310.
- Hartono, P. (2020). Similarity maps and pairwise predictions for transmission dynamics of COVID-19 with neural networks. *Informatics in Medicine Unlocked*, **20**, 100386.
- Hartono, P., Hollensen, P. and Trappenberg, T. (2015). Learning-regulated context relevant topographical map. *IEEE Transactions on Neural Networks and Learning Systems*, **26**, 2323-2335.
- Hassan, A., Shahin, I. and Alsabek, M. B. (2020). COVID-19 detection system using recurrent neural networks. *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, 1-5.
- Hawas, M. (2020). Generated time-series prediction data of COVID-19's daily infections in Brazil by using recurrent neural networks. *Data in Brief*, **32**, 106175.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
- Li, H. H. (2020). Statistical and network-based analysis of Italian COVID-19 data: Communities detection and temporal evolution. *International Journal of Environmental Research and Public Health*, 1-62.

- Li, H., Li, H., Ding, Z., Hu, Z., Chen, F., Wang, K., . . . Shen, H. (2020). Spatial statistical analysis of Coronavirus Disease 2019 (Covid-19) in China. *Geospatial Health*, **15**, 867-875.
- Li, H., Li, H., Ding, Z., Hu, Z., Chen, F., Wang, K., . . . Shen, H. (2020). Statistical and network-based analysis of Italian COVID-19 data: Communities detection and temporal evolution. *International Journal of Environmental Research and Public Health*, **15**, 1-62.
- Li, M. D., Arun, N. T., Gidwani, M., Chang, K., Deng, F., Little, B. P., . . . Kalpathy-Cramer, J. (2020). Automated assessment and tracking of COVID-19 pulmonary disease severity on chest radiographs using convolutional siamese neural networks. *Radiology: Artificial Intelligence*, **2**, 200079.
- Livadiotis, G. (2020). Statistical analysis of the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19. *Plus One*, **15**, 5-26.
- Martin, W., Jerome, L., Maja, v. C., Lars, B., Marlon, G., Derek, H., . . . Klaus, K. (2020). Statistical Analysis of Clinical COVID-19 Data: A concise overview of lessons learned, common errors and how to avoid them. *Clinical Epidemiology*, **12**, 925-928.
- Milano, M. and Cannataro, M. (2020). Statistical and network-based analysis of Italian COVID-19 data: Communities detection and temporal evolution. *International Journal of Environmental Research and Public Health*, **17**, 4182-4194.
- Militante, S. V., Dionisio, N. V. and Sibbaluca, B. G. (2020). Pneumonia and COVID-19 detection using convolutional neural networks. *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 1-6.
- Nayak, S. R., Arora, V., Sinha, U. and Poonia, R. C. (2020). A statistical analysis of COVID-19 using Gaussian and probabilistic model. *Journal of Interdisciplinary Mathematics*, 1-15.
- Niazkar, H. R. and Niazkar, M. (2020). Application of artificial neural networks to predict the COVID-19 outbreak. *Global health research and policy*, **5**, 1-11.
- O'Driscoll, M., Harry, C., Donnelly, C. A., Cori, A. and Dorigatti, I. (2020). A comparative analysis of statistical methods to estimate the reproduction number in emerging epidemics with implications for the current COVID-19 pandemic. *medRxiv*, 1-27.
- Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O. and Acharya, U. R. (2020). Automated detection of Covid-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, 103792.
- Phucharoen, C., Sangkaew, N. and Stosic, K. (2020). The characteristics of COVID-19 transmission from case to high-risk contact, statistical analysis from contact tracing data. *EClinicalMedicine*, 1-11.
- Pirouz, B., Nejad, H. J., Violini, G. and Pirouz, B. (2020). The role of artificial intelligence, MLR and statistical analysis in investigations about the correlation of swab tests and stress on health care systems by COVID-19. *Information, MDPI*, **11**, 454-474.
- Shams, M. Y., Elzeki, O. M., Elfattah, M. A., Medhat, T. and Hassanien, A. E. (2020). Why are Generative Adversarial Networks Vital for Deep Neural Networks? A case study on COVID-19 chest X-Ray images. *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*, Springer, Cham, 147-162.
- Shruthi, M. and Ramani, D. (2020). Statistical analysis of the impact of COVID 19 on India commodity markets. *Materials Today: Proceedings*, 1-6.
- Singh, D., Kumar, V., Yadav, V. and Kaur, M. (2020). Deep convolutional neural networks based classification model for COVID-19 infected patients using chest X-ray images. *International Journal of Pattern Recognition and Artificial Intelligence*, 1-19.
- Sivapalan, P., Ulrik, C. S., Lappere, T. S., Eklöf, J. V., Shaker, S. B., Bødtger, U. C., . . . Biering-Sørensen, T. (2020). Proactive prophylaxis with azithromycin and

- hydroxychloroquine in hospitalized patients with COVID-19(proPAC-COVID): a statistical analysis plan. *Trials*, **21**, 1-9.
- Sridhar, S., Seetharaman, R., Mootha, S. and Rayan, A. X. (2020). A study on co-occurrence of various lung diseases and COVID-19 by observing chest X-ray Similarity using deep convolutional neural networks. 2020 *Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*. *IEEE*, 1-8.
- Talmoudi, K., Safer, M., Letaief, H., Hchaichi, A., Harizi, C., Dhaouadi, S., . . . Bougat, S. (2020). Estimating transmission dynamics and serial interval of the first wave of COVID-19 infections under different control measures: a statistical analysis in Tunisia from February 29 to May 5, 2020. *BMC Infectious Diseases*, **20**, 1-8.
- Toraman, S., Alakus, T. B. and Turkoglu, I. (2020). Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. *Chaos, Solitons and Fractals*, **140**, 110122.
- Wang, P., Lu, J.-a., Jina, Y., Zhu, M., Wang, L. and Chen, S. (2020). Statistical and network analysis of 1212 COVID-19 patients in Henan, China. *International Journal of Infectious Diseases*, 391-398.
- Yasar, H. and Ceylan, M. (2020). A new deep learning pipeline to detect Covid-19 on chest X-ray images using local binary pattern, dual-tree complex wavelet transform, and convolutional neural networks. *Applied Intelligence*, **51**, 2740-2763.
- Zhao, S., Gao, D., Zhuang, Z., Chong, M. K., Cai, Y., Ran, J., . . . Wang, M. H. (2020). Estimating the serial Interval of the Novel Coronavirus Disease (COVID-19): A Statistical Analysis Using the Public Data in Hong Kong From January 16 to February 15, 2020. *Transboundary and Emerging Diseases*, 1-17.
- Zador, A.M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications*, **10(1)**, 1-7.