

Some Aspects of Modeling in the Study of COVID-19 Data

SYED KHALID MUSTAFA¹, VYACHESLAV LYASHENKO², N. AMEER AHAMAD³, MOHAMMAD REHAN, AJMAL⁴, M. AYAZ AHMAD⁵

¹Department of Chemistry, Faculty of Science, University of Tabuk, 71491, Saudi Arabia

²Department of Informatics, Kharkiv National University of Radio-Electronics, Ukraine

³Department of Mathematics, Faculty of Science, University of Tabuk, 71491, Saudi Arabia

⁴Department of Biochemistry, Faculty of Science, University of Tabuk, 71491, Saudi Arabia

⁵Department of Physics, Faculty of Science, University of Tabuk, 71491, Saudi Arabia

*Corresponding Author

Email ID: lyashenko.vyacheslav@gmail.com and mayaz.alig@gmail.com

Received: 09.10.20, Revised: 27.11.20, Accepted: 19.12.20

ABSTRACT

Human society or civilization is constantly faced with various tests. One such test is the COVID-19 pandemic, which is associated with the spread of a special type of coronavirus. Therefore, humanity is forced to find adequate measures to protect against the COVID-19 pandemic. For this purpose it is needed to know the reliable statistics. Additionally, it is important to know the features of the spread of COVID-19, to be able to make effective forecasts. This can be possible only by using various models. Different models take into account with specific features for the future prediction of the spread of the pandemic. These basic aspects are discussed in this work, which helps people for making better decisions for future strategies.

Keywords: Modeling, Forecasting, Analysis, Data, COVID-19 pandemic.

INTRODUCTION

The modern development of mankind is facing a new global threat, the pandemic of COVID-19. This pandemic originated from a new type of virus, which has various characteristics [1]. Therefore, it is important to have reliable statistics that can reveal the dynamics of the spread of the COVID-19 pandemic. For such analysis, various methods can be used, approaches, models [2], [3]. Such research tools compel to establish the relationship between these factors like the origin, spread, and development of a pandemic. The current study allows a better understanding of the nature of the phenomenon that is being investigated.

It is advisable to reveal the dynamics of the spread of the COVID-19 pandemic by using modeling. Simulation process is the better one to reveals the dynamics of the COVID-19 pandemic. Further it is potentially to predict individual stages in the development of an epidemic, which is characterized by outbreaks or attenuation of the disease. In order for the modeling process to be effective, it is important to take various aspects into account in the study of data on COVID-19. These are topics which are discussed here.

BRIEF OVERVIEW OF RELATED WORKS

There are several workers or researchers address various issues in the analysis of COVID-19 data. The spread of the COVID-19 pandemic and the

mortality rate of the disease are measured by using a meta-analysis of various COVID-19 data [4]. For the study of the dynamics of data on COVID-19 to consider the asymptomatic cases of the corresponding disease, a new model is substantiated and developed [5]. At the same time, the other workers examined the key issues of the impact of COVID-19 on economic development [6]. That's important for identifying funding sources in the response to the COVID-19 pandemic. J. Kucharski and et al. (2020) suggested in their work "mathematical modeling studies", and reported early dynamics of transmission and control of COVID-19. The period from January 2020 to February 2020 is considered for the modeling process of COVID – 19 transmissions [7].

Assessment of COVID-19 infected population size, and the dynamics of transmission of the epidemic in February 2020 is described [8]. Various mathematical models such as SUQC, SIDR, SRID, SIR, SEIR, SIS, and ABM are considered by a number of researchers [9-15]. These models are typically used to predict the spread of COVID-19.

Thus, there are various approaches to modeling the spread of the COVID-19 pandemic and its impact on various spheres of human life. However, it is important to consider some aspects of modeling in the study of data on COVID-19.

COLLECTION AND PROCESSING OF DATA AS AN ELEMENT OF BUILDING A MODEL

The collection and processing of a lot of data concerning a worldwide pandemic is required. Therefore, the data concerning COVID-19 for collecting and analyzing information for the prediction of the transmission, a new models and methods are used. Particularly, scientists identify four major types of large amounts of data analysis that are relevant today [7] [16].

(1) **Descriptive analysis** – used to describe existing and pre COVID-19 situation; represented graphically. The analysis (time, place, and person) should give a preliminary idea of the clinical spectrum and transmissions of disease. Input variables for the analysis are as:

- Summary of individual details of the infected patient;
- Checking/ scrutinizing the transmission effect on age and gender;
- Output variables are indicators of transmission of the disease as well as the Indicators of infection and mortality.

Advantages – ease of data collection, minimum time; the prevalence of computer programs for processing and analyzing the collected information, ease of interpretation. Disadvantages– a complete list of parameters should be identified earlier.

(2) **Predictive analysis** – a data analysis tool that helps to understand the happening in the future. Therefore, mathematical methods are used, such as, regression and forecasting models. The predictive analyses consider the transmission of infections between the regions, as well as the severity of quarantine measures in a particular.

Predictive analysis is used to formulate mathematical forecast models (data from official sources are only analyzed), as well as to increase the accuracy of the prediction model. Input variables: infected and mortality rates. Output variables – is basically the predictions about the actions of people in response to emerging events. The advantage is the simplicity of data collection, the prevalence of computer programs for processing, and analysis of collected information. Disadvantages: difficulty incorrect selection of input variables, quantity, and balance. One cannot completely rely on predictive analysis; sometimes a large amount of data is not enough.

(3) **Prescriptive analysis (PA)** – not applied to determine COVID-19 transmission, but to determine effective strategic and operational decisions. PA analytics used at global and regional levels.

Input variables: infected and mortality rates; data on pharmaceutical and non-pharmaceutical

measures. For example, the role of non-pharmaceutical measures:

- Isolation of people with symptoms of the virus;
- Voluntary home quarantine;
- Social distance of people over 70;
- Social distance of the entire population;
- Closure of schools and universities.

The output variables are strategic predictions regarding preventive measures applied in the society.

The advantage is the simplicity of data collection, prevalence of computer programs for processing, and analysis of collected information.

Disadvantages: it does not provide always a guaranteed optimal solution, but provides a consistent solution.

(4) The established analysis results shows that the comparative analysis (CA) is one of the most popular and fundamental methods of analysis, which involves identifying certain characteristics of a study object and comparing it in these parameters with other objects as well as with itself in different periods of time.

CA is used to assess the dynamics of COVID-19 transmission. A dynamic model of stochastic propagation of COVID-19 and a plurality of publicly available data sets of cases of COVID-19 outbreak.

Input variables: time of outbreaks, indicators of infected and fatal cases. The output variables are indicators of infected, fatal cases and time of possible outbreaks, as well as growth rates of COVID-19.

The advantage is the simplicity of data collection, the prevalence of computer programs for processing, and analysis of collected information. Disadvantages: application of these methods have difficulties when using criteria with non-numeric ratings (verbal ratings, yes-no ratings).

FEATURES OF THE SPREAD OF COVID-19 AS A BASIS FOR BUILDING MODELS

The time and location for the COVID-19 outbreak are caused by the rapid transmission of the virus, especially among the floating population. Additionally, the incubation period of the virus directly affects the rate of transmission.

As analyzing COVID-19 transmission data, two main areas of transmission are important:

- People having weak immunity are riskier to be infected;
- Populations be isolated or have limited movement.

However, it should be noted that the propagation rate of COVID-19 is independent of air temperature [17], but the effect of climatic and seasonal factors cannot yet be completely ruled out.

On analysis and the calculated result for the propagation of COVID-19 often satisfies Gaussian transmission.

Such as, if used Gaussian transmission to simulate, the main thrust is given to the average value, however, standard deviation has little effect, and the obtained result has the best agreement with objective fact variance to simulate [18].

With the help of Gaussian transmission, an optimal transmission that simulates the mortality rate due to COVID-19 can be obtained. It involves three parameters, height, peak location, and width. The streaming data can be used to infer function value, slope, and inflection location as a minimum set of constraints to estimate the subsequent trajectories [19].

For the calculation of the COVID-19 serial intervals, the Gaussian transmission is more suggestive than estimated gamma or Weibull transmissions, since they are limited to positive values only. So the Gaussian transmission method is used to obtain and estimate the average serial interval for COVID-19 [20].

A BRIEF OVERVIEW OF METHODS USED IN MODELS FOR ANALYSIS AND PREDICTION OF COVID-19 TRANSMISSION

Important methods are highlighted here, such as:

(1) The Bootstrap method:

For the estimation of the confidence interval, the date of incidence peak and for a total number of deaths after the end of the epidemic examined.

To obtain the peak date of the incidence curve, a sample of bootstrap used for the analysis and prediction of COVID-19 transmission. First, determine the confidence interval of incidence peak, and Calculate accordingly by using an algorithm implemented in programming language R [21]. The bootstrap method is used within the framework of the SRID model; the confidence intervals are the most important parameters of the epidemic process, which can be estimated, by avoiding the initial errors in statistical data.

(2) The Least-squares method:

For the solutions of least squares problem, a sliding window with a one-day step is used [11]. This method is applied to solve various problems, based on minimizing the sum of squared deviations of some functions from desired variables. It can be used to "solve" overdetermined systems of equations (when the number of equations exceeds the number of unknowns).

The least-squares method is used within the framework of the SIRD model; it is feasible by knowing the value of optimal infection rate.

Including, the optimal basic reproduction number (RO) and recovery rate can be calculated with the help of the said method.

(3) A method based on an RFID platform:

The method is based on radio frequency identification (RFID) devices that evaluate mutual proximity by putting an RFID tag. The distance of RFID devices should be from 1 to 2 meters.

For the scientific study, participants carry RFID badges on their chests, and contacts are recorded when participants look at each other, the body acts as a screen for RF signals that are only sensitive to proximity. Additionally, for the detection of neighboring devices, the RFID tags send collected local contact information to a number of receivers installed in the environment that transmit this information over the local network to the computer system used to monitor the stored data.

Proximity scanning is performed at random times, and each tags sends information to recipients every few seconds. Time is roughly expressed at a 20-second interval, during which you can evaluate proximity to face with a confidence of more than 99%.

The RFID method is also enough to track the dynamics of social interaction and the assessment of epidemic transmission effects. This method is mainly used to measure contact networks, in other words, the SEIR and SIDR models, it allows to store large data; evaluating the tag data at any location. The RFID method provides environmental resistance, which affects the assessment of epidemic transmission.

(4) Other method based on deterministic model for knowing the transmission structure in the particular time periods. The deterministic model uses real-time data of confirmed cases of patient who are getting treatment and the patient who lost their life during the treatment, which provide the up-to-date dissemination forecasts. The main condition of this method is strictly based on non-pharmaceutical measures [21].

This method is mainly used for SIR, which allows us to simulate a set of scenarios to assess the impact of various approaches to mitigate social distance measures. The method provides a quantitative assessment of those patients who are critically ill and getting critical treatment in the hospital, as well as the fatal cases.

(5) The optimization method for forecasting, FPASSA model is used often. The developed FPASSA-ANFIS model is used as a forecasting method for COVID-19. FASSA-ANFIS uses advanced FPA to train the ANFIS adaptive neuro-fuzzy inference model by optimizing its parameters. FASSA-ANFIS contains five layers as

a classic ANFIS model. FASSA-ANFIS begins by formatting input data in form of time series [22]. The method used to design neural networks is to use fuzzy systems or fuzzy structures.

FEATURES OF SIMULATION COVID-19

The current analysis is based on the mathematical models and methods, the key points are highlighted here, which will provide to understand their features clearly. The SUQC model, which operates with four variables that are associated with functions and are used to suggest the four possible types of people:

- (1) Number of susceptible people having very low immune in the population;
- (2) Number of people infected and not quarantined, which may be either asymptomatic or symptomatic (infectious);
- (3) Number of infected quarantined;
- (4) Number of infected confirmed cases.

Further, the SUQC differs from SEIR in the following aspects:

- (1) Infected people are classified as:
 - (a) People who are "Not quarantine";
 - (b) People who are quarantined;
 - (c) People with a confirmed diagnosis.

And only "not quarantined" can infect susceptible people, and in the SEIR model, all infected people are transmittable.

(2) Quarantine and speed of propagation, SUQC is used to determined by modeling the "effects" of quarantine and control measures;

(3) Parameters confirmed by infected persons (observation data), the total number of infected, and parameter virus transmission rate depend on medical resources and sensitivity of diagnostic methods.

The SRID model is based on a forecast of epidemic dynamics in two scenarios: without increasing estimates of infected people number, people recovered number and taking these amendments into account.

The SIDR model, in contrast to the above, takes into account the average number of secondary cases resulting from the introduction of infection into a fully susceptible population during the period of infection.

The SIR model is characterized in that way:

- (1) Population is taken into account;
- (2) Reflects a reassessment or the prediction of the number of infected, which are compensated for unregistered or undetected.
- (3) It can be used, if good medical support is guaranteed.

The SEIR model for modeling transmission of infectious diseases applied in the study population.

The SEIR model differs from the above in that:

(1) The present model, emphasizes on fact that there were no births, no deaths, no critical conditions: susceptible (S), open (E), infectious (I), or recovered (R). The model is individual and stochastic.

(2) Model taking into account: average duration and number of contacts, as well as the average path of the person moving.

(3) Model based on a dynamic network of contacts: homogeneous (HOM), heterogeneous (HET), and dynamic (DYN) networks.

(4) Model based on 2 scenarios: a very short latency and an infectivity scenario; short-latency, short infectivity scenario.

SEIR is a typical example taking into account the incubation period.

The SIS model is based on characterized in that:

(1) Model divides the population into two categories of susceptible S and infected I;

(2) In this model, external factors that arise when social and economic interactions transmit infectious diseases are analyzed.

Agent-based model: This model is differentiated as:

(1) In applying this model to compare several intervention strategies, including restrictions on international air travel, case isolation, home quarantine, social distance with different levels of compliance, and school closures.

(2) Age-related infections, range of reproductive numbers, age-related, and socio-contextual transmission rate. Real-time forecasting is based on phenomenological models that have been tested during previous outbreaks to create and evaluate short-term forecasts of a total number of confirmed recorded cases.

This model provides 5, 10, and 15-day forecasts for five consecutive days – a distinctive feature of this model.

A real-time predictive method based on artificial intelligence (AI) [23] using this model to predict confirmed cases of COVID-19. A Multi - stage forecasting.

Benchmarking is based on the COVID-19 proliferation dataset. However, there are problems with such an analysis, especially in real-time. The difference between this modeling approach is that one can take into account delays in the onset of symptoms and uncertainty by explicitly including delays resulting from the natural history of infection and reporting processes [7], [20].

The FPASSA-ANFIS model is highly capable of confirmed cases predicting numbers within ten days. In addition, FPASSA-ANFIS is superior to other forecasting models in terms of RMSE (Root Mean Squared Relative Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), RMSRE (Root Mean Squared Relative Error), and R2 (Coefficient of Determination). **Real-time prediction:** Different existing susceptibility, susceptible, infected, and remote (SEIR) -based virus transmission models that are well suited for pandemic modeling with the enough historical data, the SEIR-based model also uses short historical data to find a trend. The parameters of advanced SEIR models are practiced, using data from a brief past window to accurately predict trends. This differential epidemic model is essentially a window-based, time-varying SEIR model. Since the SEIR model is still a physical model of disease transmission, its short-term prognosis (for example, for one month) can still be very useful for politicians to make a decision on reducing the incidence and changing business policies in real-time [24].

CONCLUSION

The present work indicates that the current existing models are applied in different directions. Each model addressed specific features of the development of the COVID-19 pandemic, which is important for making forecasts and future predictions of the pandemic. The current work provides a comparative analysis that highlighted the features of modeling. This analysis is required for the study with a more precise specification of the features of existing models. This is a prerequisite for the further development of the most effective models for assessing and predicting the development of the COVID-19 pandemic.

ACKNOWLEDGEMENTS

The author (Mohammad Ayaz Ahmad) would like to acknowledge the keen support in financial assistance for this work of the Vice Presidency / Studies and Scientific Research/Deanship of Scientific Research on behalf of the University of Tabuk, Kingdom of Saudi Arabia and Ministry of Higher Education, K.S.A. Under the research grant no. S-0263-1436/dated 15-03-1436, along with some recent publications based on COVID-19 [6] and [25]-[35]. And also highly acknowledge the Department of Informatics, Kharkiv National University of RadioElectronics, Ukraine for numerous help and support to complete this article.

REFERENCES

1. Spinelli and G. Pellino. (2020). "COVID-19 pandemic: perspectives on an unfolding crisis", *The British journal of surgery*, vol. 107, no. 7, pp. 785–787, 2020.
2. P. Verma, R. Gupta, S. Varma, A. Singh and A. Kumar. (2020). "Sentimental analysis on COVID-19 twitter data using convolutional neural networks (CNN)", *Journal of critical reviews*, vol. 7, no. 18, pp. 530-537.
3. N. Nawaz, A. M. Gomes, M. A.Saldeen. (2020). "Artificial intelligence (AI) applications for library services and resources in COVID-19 pandemic", *Journal of critical reviews*, vol. 7, no. 18, pp. 1951-1955.
4. L.Q. Li and et al. (2020). "COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis", *Journal of medical virology*, vol. 92, no. 6, pp. 577-583.
5. Y. Bai and et al. (2020). "Presumed asymptomatic carrier transmission of COVID-19", *Jama*, vol. 323, no. 14, pp. 1406-1407.
6. S. K. Mustafa and et al. (2020). "Using Wavelet Analysis to Assess the Impact of COVID-19 on Changes in the Price of Basic Energy Resources", *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 7, pp. 2907-2912, 2020.
7. J. Kucharski and et al. (2020). "Early dynamics of transmission and control of COVID-19: a mathematical modelling study", *The lancet infectious diseases*, vol. 20, no 5, pp. 553-558.
8. J. T. Wu and et al. (2020). "Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China", *Nature Medicine*, vol. 26, no. 4, pp. 506-510.
9. L. Geng and et al. (2015). "Multi-scale fractional-order sparse representation for image denoising", *International Conference on Neural Information Processing*, pp. 462-470, 2015.
10. M. Namihira. (2020). "A COVID-19 Model", *Espy Analytics*, pp. 1-12.
11. J. Stehlé and et al. (2011). "Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees", *BMC medicine*, vol. 9, no. 1, pp. 87.
12. Simha, R. V. Prasad and S. Narayana. (2020). "A simple stochastic SIR model for COVID 19 infection dynamics for Karnataka: Learning from Europe", *arXiv preprint arXiv:2003.11920*.
13. Z. Bethune and A. Korinek. (2020). "COVID-19 Infection Externalities: Pursuing Herd Immunity or Containment?", *Covid Economics: Vetted and Real-Time Papers*, vol. 1, no. 11, pp. 1.
14. S. L. Chang and et al. (2020). "Modelling transmission and control of the COVID-19 pandemic in Australia", *arXiv preprint arXiv:2003.10218*.
15. D. Kai and et al. (2020). "Universal masking is urgent in the covid-19 pandemic: Seir and agent based models, empirical validation, policy

- recommendations”, arXiv preprint arXiv:2004.13553.
16. V. Surveillances. (2020). “The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)–China, 2020”, *China CDC Weekly*, vol. 2, no. 8, pp. 113-122.
 17. E. Shim, A. Tariq, W. Choi, Y. Lee and G. Chowell. (2020). “Transmission potential and severity of COVID-19 in South Korea”, *International Journal of Infectious Diseases*, vol. 93, pp. 339-344.
 18. L. Li and et al. (2020). “Propagation analysis and prediction of the COVID-19”, *Infectious Disease Modelling*, vol. 5, pp. 282-292.
 19. T. W. Lee, J. E. Park and D. Hung. (2020). “Gaussian Statistics and Data-Assimilated Model of Mortality due to COVID-19: China, USA, Italy, Spain, UK, Iran, and the World Total”, medRxiv, 2020.
 20. S. E. Moore and E. Okyere. (2020). “Controlling the transmission dynamics of COVID-19”, arXiv preprint arXiv:2004.00443.
 21. M. J. Keeling and et al. (2020). “Predictions of COVID-19 dynamics in the UK: short-term forecasting and analysis of potential exit strategies”, medRxiv.
 22. M. A. Al-Qaness, A. A. Ewees, H. Fan and M. Abd El Aziz. (2020). “Optimization method for forecasting confirmed cases of COVID-19 in China”, *Journal of Clinical Medicine*, vol. 9, no. 3, pp. 674.
 23. Z. Hu, Q. Ge, L. Jin and M. Xiong. (2020). “Artificial intelligence forecasting of covid-19 in China”, arXiv preprint arXiv:2002.07112.
 24. S. X. Tan and L. Chen. (2020). “Real-Time Differential Epidemic Analysis and Prediction for COVID-19 Pandemic”, arXiv preprint arXiv:2004.06888.
 25. S. Sotnik, S. K. Mustafa, M. A. Ahmad, V. Lyashenko and O. Zeleniy. (2020). “Some Features of Route Planning as the Basis in a Mobile Robot”, *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5 pp. 2074-2079.
 26. V. Lyashenko, S. K. Mustafa, N. Belova and M. Ayaz Ahmad. (2019). “Some Features in Calculation of Mold Details for Plastic Products”, *International Journal of Emerging Trends in Engineering Research*, vol. 7, no. 11, pp. 720-724.
 27. M. Ayaz, T. Sinelnikova, S. K. Mustafa and V. Lyashenko. (2020). “Features of the Construction and Control of the Navigation System of a Mobile Robot”, *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 4, pp. 1445-1449.
 28. M. Ayaz, I. Tvoroshenko, J. H. Baker and V. Lyashenko. (2019). “Computational Complexity of the Accessory Function Setting Mechanism in Fuzzy Intellectual Systems”, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 5, pp. 2370-2377.
 29. Syed Khalid Mustafa, M. Ayaz Ahmad, Svitlana Sotnik, Oleksandr Zeleniy, Vyacheslav Lyashenko, Othman Alzahrani, “Brief review of the Mathematical Models for Analyzing and Forecasting Transmission of COVID-19”, *Journal of Critical Reviews*, Vol. 7 (19), (2020), pp. 4206-4210. <https://doi.org/10.31838/jcr.07.19.605>
 30. Syed Khalid Mustafa and M. Ayaz Ahmad, et. al., COVID-19 and Immune Function – “A Significant” Zinc, *Oriental Journal of Chemistry (OJC)*, Vol. 36(6), 1026-1036 (2020).
 31. Valentina Pavlova, Biljana B. Popov, Vesna K. Hristova et al., “Chemical characterization of VARUMIN (1 and 2) oral solutions as potential therapeutic beverages” *International Journal of Pharmaceutical Research*, Vol 12, Supp. Issue 2, 4454-4460, (2020).
 32. Mircea Resteiu, Remus Dobra, Dragos Pasculescu, M. Ayaz Ahmad, “Quality Engineering Tools Focused on Designing Remote Temperature Measurements for Inaccessible Locations by Using Light Components Parameterization of the Heated Materials”, *IOP Conference Series Materials Science and Engineering*, Vol. 133, (2016), pp. 012059.
 33. Maria Loredana Boca, Remus Dobra, Dragos Pasculescu, M. Ayaz Ahmad, “Analysis and simulation of industrial distillation processes using a graphical system design model”, *Proc. SPIE*, Vol. VIII (10010), (2016), pp. 100102U-1-8, doi:10.1117/12.2245971
 34. M. Ayaz Ahmad, Jalal Hasan Baker, Mir Hashim Rasool, Shafiq Ahmad, R. Dobra, D. Pasculescu & Charles Roberto Telles, “Fluctuations in produced charged particle multiplicities in relativistic nuclear collisions for simulated events”, *IOP Conf. Series: Journal of Physics: Conf. Series* 1258 (2019) 012010 (1-9), doi:10.1088/1742-6596/1258/1/012010
 35. M. Ayaz Ahmad, Shafiq Ahmad, “Study of angular distribution and KNO scaling in the collisions of ²⁸Si with emulsion nuclei at 14.6A GeV”, *Ukrainian J. of Physics* Vol. 57, No. 12, (2012), 1205-1213.