BRAIN. Broad Research in Artificial Intelligence and Neuroscience
ISSN: 2068-0473 | e-ISSN: 2067-3957
Covered in: Web of Science (WOS); PubMed.gov; IndexCopernicus; The Linguist List; Google Academic; Ulrichs; getCITED; Genamics JournalSeek; J-Gate; SHERPA/RoMEO; Dayang Journal System; Public Knowledge Project; BIUM; NewJour; ArticleReach Direct; Link+; CSB; CiteSeerX; Socolar; KVK; WorldCat; CrossRef; Ideas RePeC; Econpapers; Socionet.

2020, Volume 11, Issue 4, pages: 168-184 | https://doi.org/10.18662/brain/11.4/147

Predicting COVID-19 Incidence Using Data Mining Techniques: A case study of Pakistan

Saba NOOR[1],
Waseem AKRAM[2]*,
Touseef AHMED[3],
Qurat-ul-Ain[4]

[1] Department of Computer Science, Lahore Garrison University, Lahore, Pakistan

[2] Department of Computer Science, COMSATS University Islamabad, Pakistan, imwaseem.khan@yahoo.com

[3] Department of Computer Science, Lahore Garrison University, Lahore, Pakistan

[4] Department of Computer Science, Lahore Garrison University, Lahore, Pakistan

Abstract: *The Outbreak of Coronavirus (COVID-19) came to the world in early December 2019. The early cases of coronavirus were reported in Wuhan City, Hubei Province, China. Till May 18, 2020, 198 countries have been affected by this life-threatening disease. The most common and known traits of COVID-19 are tiredness, fever, and dry cough. In this paper, we have discussed the Predictive data mining approach for COVID-19 predictions. In Predictive data mining, a model is developed and trained using supervised learning and then it predicts the behavior of provided data. Predictive data mining is a renowned technique known to many health organizations for the classification and prediction of diseases such as Heart disease and various types of cancers etc. There are several factors for comparing the model's accuracy, scalability, and interpretability. This predictive model is compared to the basics of its accuracy. In this proposed approach, we have used WEKA as it provides a vast collection of many machine learning algorithms. The main objective of this paper is to forecast the possible future incidence of corona cases in Pakistan. This study concludes that the number of corona cases will increase swiftly. If the government take proactive steps and strictly implement precautionary measures, then Pakistan may be able to overcome this pandemic.*

Keywords: *WEKA; Predictive data mining; COVID-19.*

How to cite: Noor, S., Akram, W., Ahmed, T., & Qurat-ul-Ain (2020). Predicting COVID-19 Incidence Using Data Mining Techniques: A case study of Pakistan. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 11*(4), 168-184. https://doi.org/10.18662/brain/11.4/147

Introduction

The disease of COVID-19 was originated in Wuhan city of province Hubei, China. Chunyun, the days of mass migration for the yearly Spring Festival. To limit the spread of COVID-19, Chinese authorities adopted an extraordinary approach on 23 January 2020. These guidelines comprised of national-wide quarantine, Limited and strict traveling policies, and vast surveillance of covid-19 alleged cases.

The Covid-19 was confirmed to reach Pakistan on February 26, 2020, when a student returning from Iran tested positive. By March 18, Cases of COVID-19 has been reported all across the country, as of June 04, 2020, there have been about 85264 confirmed cases with 28923 recoveries and 1688 deaths in the country.

The aim of this is study mainly centered on forecasting the breakout trends in Pakistan base on the breakout pattern in China, Spain as the density of population is dense in these countries. We wanted to illustrate, how these precautionary measures restricted the outbreak.

Current Condition in Pakistan

According to the Ministry of Health, Government of Pakistan, the total number of confirmed cases is 85264 and 1770 deaths on Thursday, June 06, 2020. Punjab province is most affected with confirmed cases (31104), then the province of Sindh with confirmed cases (32910), province of Khyber Pakhtunkhwa (11373), province of Baluchistan (5224), Gilgit Baltistan (824), Federal city (3054) and Kashmir with total cases of 285. The results are shown in Table 1. The overall Covid-19 case history of Pakistan is as in Figure1.

Table 1 : Results of Confirmed, Deaths and Recovered cases in different province of Pakistan (covid.gov.pk)

Sr.	Province	Confirmed cases	Deaths	Recovered Cases
01	Punjab	31104	607	7712
02	Sindh	32901	555	16022
03	Khyber Pakhtunkhwa	11373	500	3150
04	Baluchistan	5224	51	2021
05	Islamabad	3544	38	518
06	Azad Jammu & Kashmir	285	7	173
07	Gilgit Baltistan	824	12	532

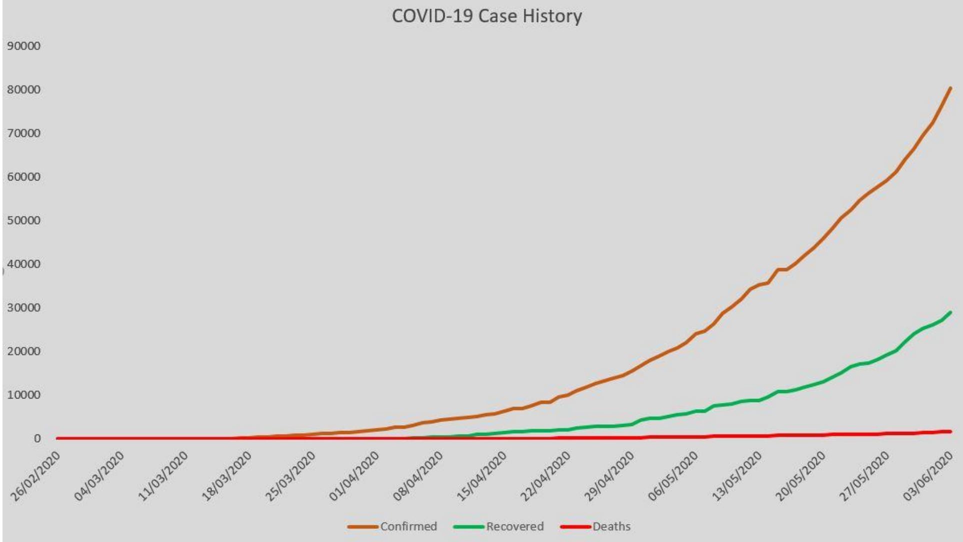

Figure 1: COVID-19 Case History (covid.gov.pk)

Method

The predictive model is based on a time-series cumulative dataset of coronavirus confirmed, recovered, and mortalities.

Definitions

S. Zhang et al. (2020) defines corona disease as Coronavirus disease (COVID-19) is a transferrable disease spread by a recently discovered coronavirus. A positive case of COVID-19 infection was defined as a case with a positive result for viral nucleic acid testing in respiratory specimens. A suspected case can be defined as a case with symptoms of COVID-19 infection but not confirmed by viral nucleic acid testing.

Dataset

Datasets are collected from Humandata.org. That track Global-time series data. The extracted data for the model is updated to June 04, 2020.

Model Development

Forecasting the COVID-19 incidence in Pakistan, the Linear regression-based approach is mainly focused on comparing the performance of Model RMSE, and MAE is proposed.

Literature Review

Twenty kinds of literature were reviewed for this study. The objective of this literature review is to discover how different models perform according to the given scenarios. In (Yang et al., 2020) presented a study, predicting the epidemic trend of Coronavirus in China. They used a Modified SEIR model with an AI approach trained in the late 2003 SARS dataset, to predict the outbreak. This study concludes that the breakout trend will start to decline by end of April. In (S. Zhang et al., 2020) proposed a study to estimate the reproductive cases of COVID-19 to predict daily cases on Diamond Princess cruise ship. They used serial interval distribution of existing daily incidence and estimate reproductive numbers of COVID-19 based on approximately Poisson distribution. The outcome of this study states that the number of new cases will gradually increase and cumulative COVID-19 cases may reach 1514 in the next ten days. The paper by (Binti Hamzah et al., 2020) developed an online tracker for daily statistics and analysis of corona cases, this paper aims to forecast the active confirmed, recovered cases of COVID-19 within and outside of China. This study uses Susceptible-Exposed-Infectious-Recovered (SEIR) as a predictive model. His study concludes that the peak of the outbreak will reach in late May 2020 with cases exceeding 76000 and start to decline in early July 2020. Authors in Qasim et al., (2020) use a mathematical model, sequence mean weight (TSMW) to predict COVID-19 cases across Pakistan. This model finds out that the count of patients may reach 77,905 with at least 8285 confirmed cases and 1382 death in the next 45 days, till 29th April 2020. Autoregressive–moving-average model (ARMA) is a hypothesis-based testing model firstly proposed in 1951; it is mainly used for un-stationary time-series data. This same heuristic can be used in Simulation modeling (L. Li et al., 2020) in which an earlier digital prototype is developed to analyze the performance of this model before deploying it. In Simulation modeling, the heuristic can be used for digital prototyping. This study represented a baseline of the transmission process of COVID-19 by using a new model based on Gaussian distribution theory. This paper finds out the key factors of virus spread, such as the incubation period of the virus, reproductive number, and daily infections. The study (Y. Li et al., 2020) developed a dynamic time series model to forecast the short term trend of COVID-19 spread inside China. The model is based on different mathematical formulas. This study concludes that in China, total cases of coronavirus may reach to 36,343 after one week (February 8,2020). Wuhan will peak its confirmed cases on March 2020. After which the infection rate will start decline throughout China. This study ignore some factors that can impact the result, factors such as birth rate or natural deaths Artificial intelligence with another predicting model can provide more realistic figures as (Yang et al., 2020). Table 2 below provides summary on the literature review.

Table 2: A detailed comparative analysis table of different techniques

SR #	Title	Published	Technique	Advantages	Disadvantages	Remarks
1	Yang et al. (2020)	2020	Artificial-intelligence and (SEIR)	It's used an AI-based model trained on past SARS dataset for more effective predictions	This study did not take into consideration of phase-adjusted protective measures and Realtime changing parameters, which can disturb the prediction accuracy	Using Mathematical tools for Epidemic prediction but it considers homogeneous Population
2	S. Zhang et al. (2020)	2020	Heuristic models	provide mortality and disease spread	The heuristic models define only one phase of the breakout but fail when the disease grows toward another stage.	Useful for finding mortality and disease spread.
3	Binti Hamzah et al. (2020)	2020	Susceptible-Exposed-Infectious-Recovered Model (SEIR)	assess the decline in efficient contacts when the completion of the acute and extreme closure of the society	Since Outbreak in not successfully contained, so it Does not provide an entirely accurate prediction	Also, Provide political as well as economic expected Predictions
4	Qasim et al. (2020)	2020	Mathematical derivation Model	Model Can Validate new future data	The lower bound remains close to actual data for the same situation	provide a generic prediction of expected COVID-19 cases

5	L. Li et al. (2020)	2020	Simulation Model	provide transmission of the virus using Gaussian distribution, Accurate results	The model can provide Accurate results for Developed countries, but for undeveloped countries, the simulation propagation may not be as accurate	use Simulation of propagation which can be reliable for factors of Spread
6	Petropoulos & Makridakis (2020)	2020	S-Curve model and univariate time series model	Past patterns like precautionary measures will remain effective as data is accurate	It does not provide Long term prediction of cases as well as cumulative cases predictions	Using time-series real-time data, predictions can be used for Govt.
7	Ardabili et al. (2020)	2020	SIR Model	Provide generalized Ability of ML models and accuracy for different lead times	SIR model cannot provide promising results, where independent individual for social distancing	Use the ML model as well SIR model and provide differentiation of accuracy
8	Yan et al. (2020)	2020	Supervised B oost-classifier	Represents a simple clinical test for precisely qualify death risk	the model will not remain the same and start varying for different data	Identify high death risk patients in the early stages of the disease
9	Wynants et al. (2020)	2020	PROBATE	support medical decision making	Update Coronavirus predictive models are available	This Paper evaluates other prediction models
10	Y. Li et al. (2020)	2020	SEIQDR, TS Modelama model	Dynamic Model	Does not consider natural deaths and birth	Useful in developed countries
11	Qin et al. (2020)	2020	ARMA	More useful as SMI data is more user relative	Big data from Unauthentic source	It can be useful for undeveloped countries as SMI is more user relative

12	Tiwari et al. (2020)	2020	Time series forecasting Method	Predicts daily cumulative cases effectively	Ignores Social-economic factors	The model can be implemented by using WHO data rather than Govt. provided
13	Fong et al. (2020)	2020	Machine learning and Multiple regression	Forecast with relatively lowest prediction error	Small data can be used for a fully observable situation not for partial observable	Prediction using small data requires an accurate data set
14	Stübinger & Schneider (2020)	2020	Dynamic Time Warping	Forecast the breakout of COVID-!9 using a lead-lag structure	Cannot predict the long term spread	Produce results based on Differentiation of different countries
15	Chakraborty & Ghosh (2020)	2020	Wavelet-based Forecasting	This model is most suitable for nonstationary data.	Not provide an accurate result for stationary data	Lack of accuracy as compared to ARMA model
16	Avery et al. (2020)	2020	Phenomenological modeling	Interpreting the limited data	Require Govt. provided accurate facts and figures	Can guide in for economic policymaking during this epidemic
17	Janies et al. (2008)	2020	SDT (Demarcation of sequence e characters)	Provide interpretation of Zoonotic potential of Coronavirus	It lacks Out grouping and rooting criteria	Provide bio medic details of COVID-19
18	G. Zhang et al. (2020)	2020	Improved SEIR	Dynamically predicts results	Insufficient test cases	can predict small scale predictions
19	Wu et al. (2020)	2020	EIR-metapopulation model	Provide support validity for the forecast.	Ignore the traveling factor of disease spread	It can also nowcast current situation that happening

20	Murray (2020)	2020	Statistical Model	Explicitly supports age structure variation	Not dynamic as more accurate data will be a need as no. of patients increases	Can help for better management of health resource in case of instant increase of cases

Statement of the Problem

On February 26, 2020, Pakistan registered its first Covid-19 case, and on March 25, 2020, Pakistan confirmed its first death in Lahore due to Covid-19.

Since from February 26, Covid-19 Outbreak keep its spread in Pakistan, and as the government of Pakistan lifted most of the lockdown,

It is essential to predict what Pakistan affords this ease. Moreover, it can be able to overcome this pandemic, or it will become another America or Wuhan!

Objectives of the Study

The main objective of this study is to predict Coronavirus incidence and its trends across the different regions of the country by using an Efficient and Strong Model.

- Analysing the current trend of Covid-19 in Pakistan.
- Developing a reliable predictive model
- Predicting the coronavirus cases using Linear Regression.

Methodology

Based on the gathered outbreak data, this model tried to discover the transmission rule of the coronavirus, forecast the breakout situation. The dataset of Coronavirus is collected from humadata.org and verified through figures provided by the Ministry of Health of Pakistan. To carry out this prediction, Weka, a tool of data mining which was developed by The University of Waikato, New Zealand. Weka applies different algorithms on datasets and provides results. There are four major phases of this model. In the first phase data pre-processing and data, transformation is carried out. The second phase of this study comprises of model training, in which Linear Regression as a forecasting algorithm is used. During the training of the model, cumulative confirmed, recovered and mortalities cases area fed as the dependent variable and time-series data variable as the independent variable.

Linear regression plots straight lines on scatter graph, so the possibility of outliers is minimum, but it has been observed that in daily cases, total daily cases decline or incline against the plotted curves causing outliers. In such regard, it is better to use mode instead of median as it provides a Real-time accurate average. The third phase of models validates the accuracy of the model, in which RSME and MAE are considered. The fourth and last phases provide results and forecasts for the next 58 days.

Figure 2 represents a detailed overview of Proposed Methodology with all these four phases of the model as:

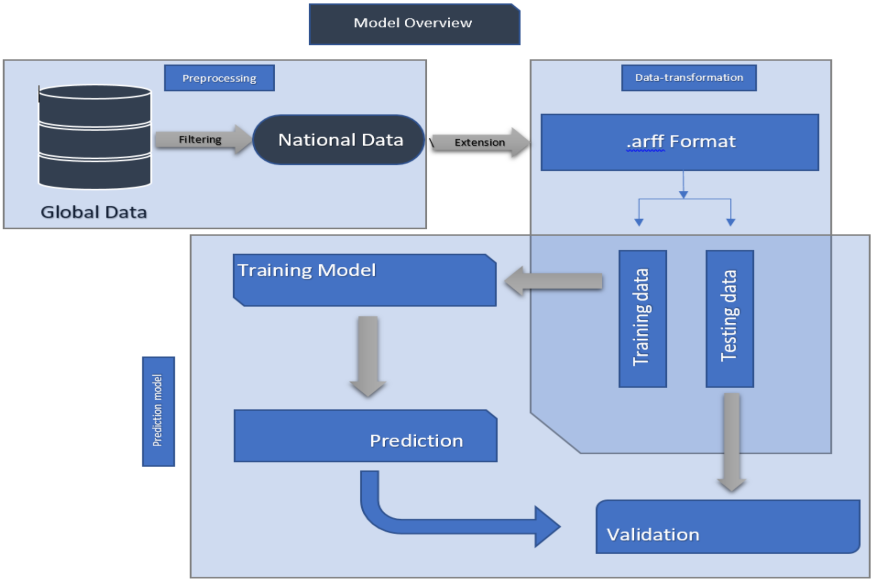

Figure 2: A detail overview of Proposed Methodology

Data-Pre-Processing: *The* Required data for this model is filtered from the global pool time-series dataset. The attributes of this dataset are confirmed cases, recovered cases, and deaths to date.

Data-Transformation: Data pre-processing provides the national data set required by the model. However, as Weka is being used as a data mining tool for this model, the default format for data is. arff, so after extracting meaningful data for this model, data transformation is applied in which data is transformed from CSV to arff.

Training and evaluation of data: For training and testing of this model, 80% of data are used for training and 20% for evaluation.

Model Training: The main objective of training the model is for learning so it can generate and predict. This model uses Linear regression for forecasting. 80% of data is feed to the model, and 20% of data is used for its performance and efficiency evaluation.

Validation of Model: To evaluate the accuracy of the Model RMSE is mainly targeted.

Results

This is a dynamic predictive model that can predict data changes, and the model can predict cases on a daily or weekly basis. This model works with three datasets and predicts cumulative, confirmed cases of patients infected with Covid-19, death toll, and recovered cases. As the number of confirmed cases increases rapidly in the last three weeks and from Figure 3 generated by this model concludes that the confirmed cases in Pakistan are expected to increase rapidly. This model expects that the number of coronaviruses confirmed cases till July 31, 2020, will get a rapid rate with no effective policy to encourage masses for social distancing and other safety measures.

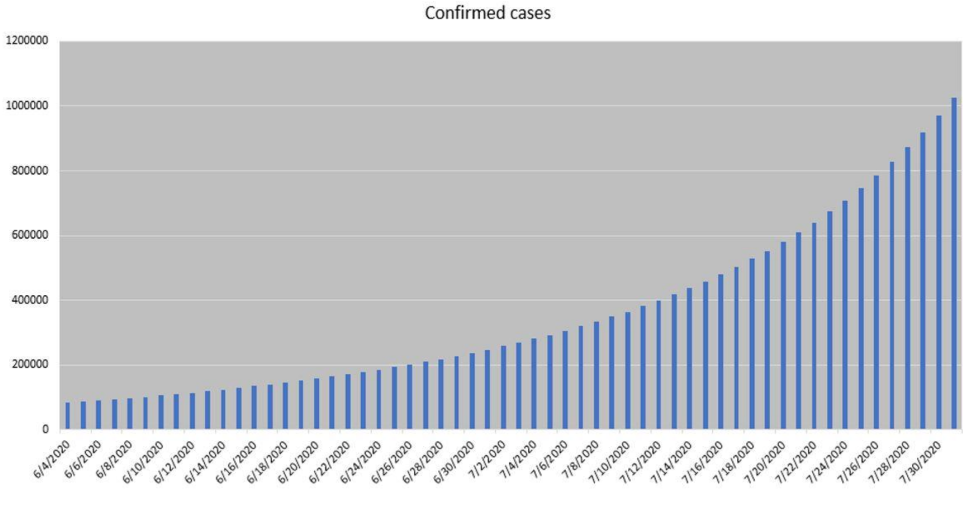

Figure 3: Confirmed cases

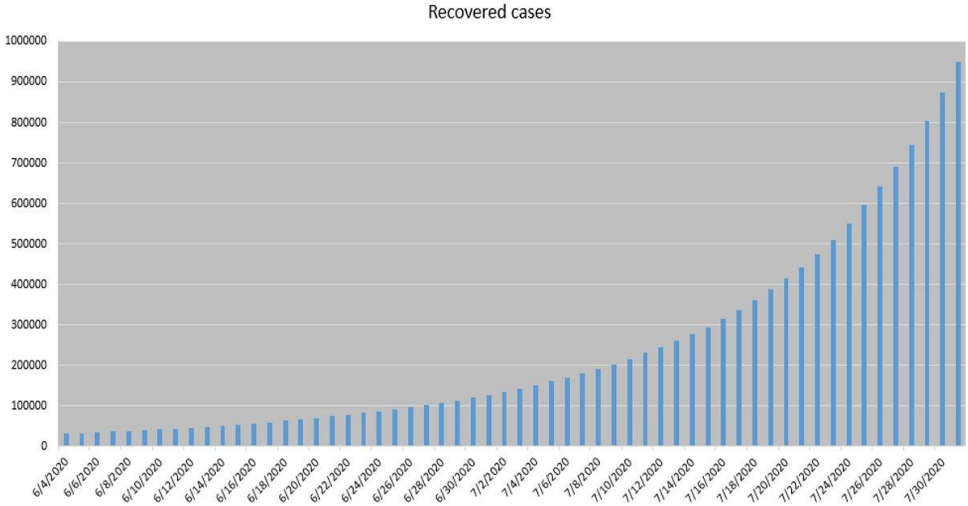

Figure 4: Recovered cases

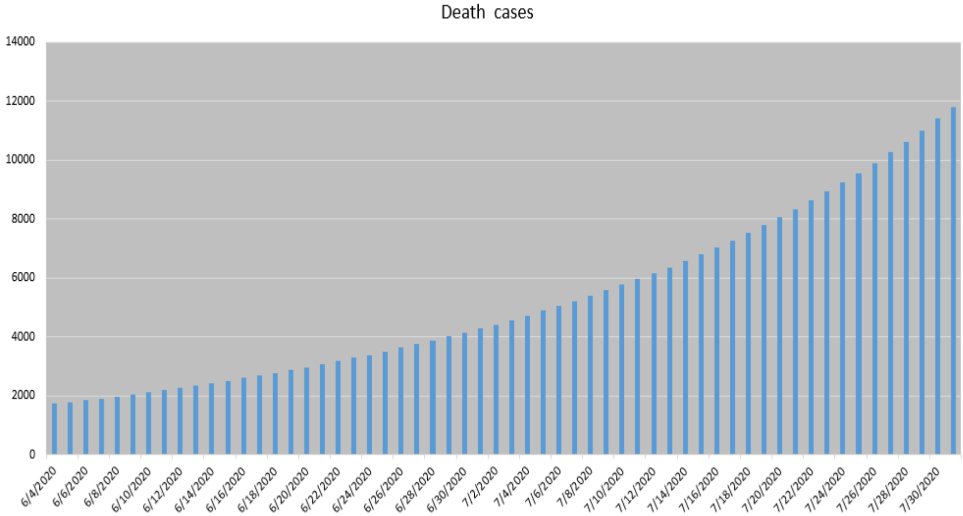

Figure 5: Death cases

This model also predicts recovered cases and deaths for this period. The recovery rate will also swiftly increase with nearly 90,000 recoveries, in figure 4. From Figure 5, the death toll for this period is expected to remain at 12,000.

Evaluation of results

Evaluation of Model is performed using very well-known measures of accuracy, mean absolute error (MAE), and Root means square error (RMSE). The 58-day evaluation of predicted cases is represented in Figures 6, 7, and 8.

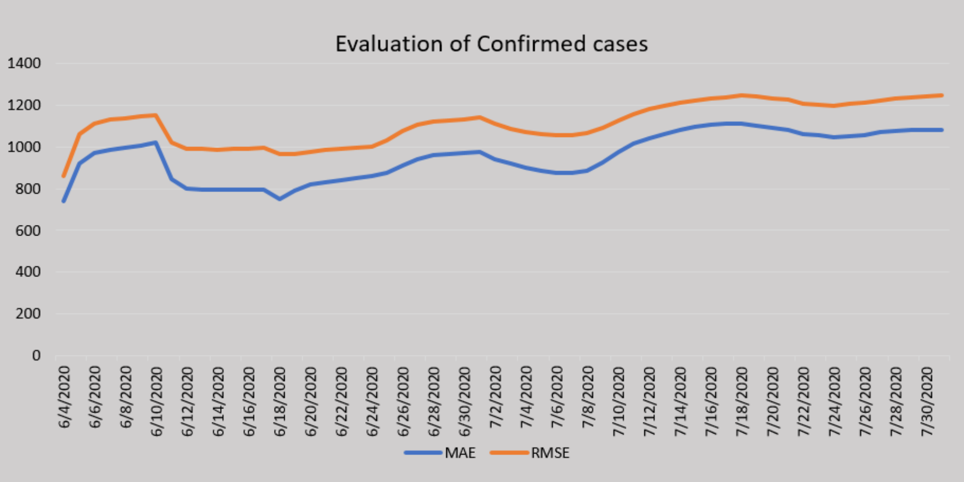

Figure 6: Evaluation of Confirmed cases

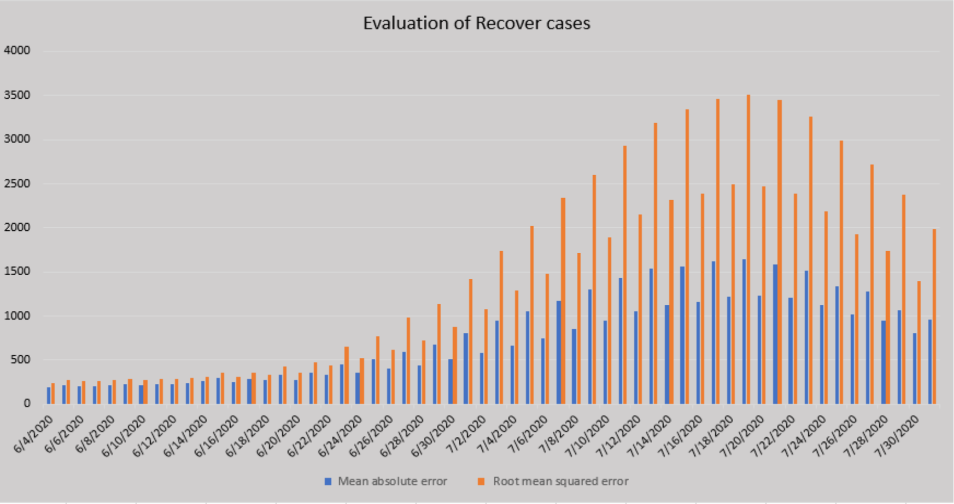

Figure 7: Evaluation of Recover cases

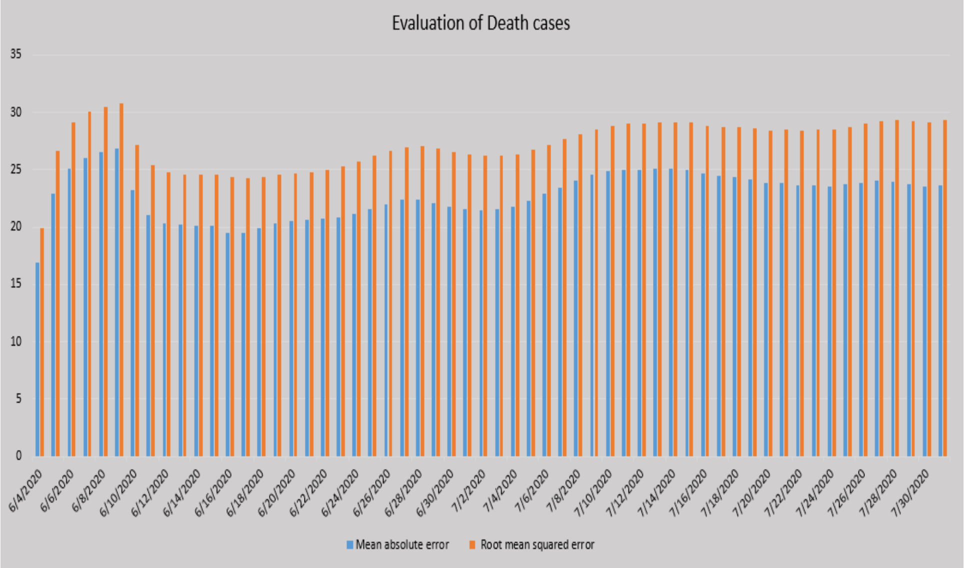

Figure 8: Evaluation of Death cases

Discussion and Limitations

In this paper, we have reviewed many datamining kinds of literature and their techniques. Susceptible-Exposed-Infectious-Recovered Model (SIER) is a mathematical model that describes circumstances in which an individual with an infectious disease becomes a source of infection for others. As from the name, this model has four stages, with parameter β (beta), which controls the rate of spread, α (alpha) incubation rate, and γ (gamma), which is the recovery rate. This technique base on the SIR model. This technique forecast future events Just like COVID-19, which spreads through close contact of infected masses.

In early 2020, one of the biggest pandemics of the 21st century came to light with many mortalities and infected cases. The rate of transfer of this virus was very fast in both developed or underdeveloped countries, and it was essential to predict and analyze this rapid spread; till now, many researchers have proposed many techniques and models. The forecast has relatively the lowest prediction error as it used machine learning algorithms for the prediction of corona cases. Machine learning is one of the significant developments of the last ten years. It is an application of Artificial intelligence in which we train the machine by providing available data, and machines can then use artificial neural networks upon some pattern to provide results. Similarly, one of the most critical applications of AI is deep learning; it is a more advanced version of machine learning. The concepts of deep learning were proposed in the early 2000s, but the breakthrough in deep learning came after the winter of AI in 2010.

As from the name, this model has four stages, with parameter β (beta), which controls the rate of spread, α (alpha) incubation rate, and γ (gamma), which is the recovery rate. This technique base on the SIR model. This technique forecast future events Just like COVID-19, which spreads through close contact of infected masses.

The Heuristic model use some early profit estimation techniques, find the best cost-effective solutions; completeness is not guaranteed at some point where backtracking is not possible or not be efficient.

This study is not considering any Social and economic factors such as Educational, economic, or relational beliefs. These factors may affect the spread.

Conclusion

The pandemic of coronavirus came to the world in early January 2019 and till June 04, 2020, almost the whole world is affected by it. This virus is closely related to bat coronaviruses causing COVID-19 disease. As explained earlier there are many known symptoms of this disease such as tiredness, fever, and dry cough. The disease of COVID-19 spread exponentially causing a rapid infection rate. This infection rate is even faster in the Third world and highly populated countries like India, Bangladesh Pakistan, etc, and the current situation of Pakistan is not satisfactory as the infection rate continuously rising, with very limited finical and medical resources. Pakistan must take proactive measures to gain control over this pandemic and this study can help policymakers to take comprehensive action as well as necessary future needs in the health sector. We have carried out this study to find out the future trend of the situation with the help of the data mining technique of Linear regression, with the help of three different cumulative data sets of recovered, deceased, confirmed cases, and our proposed methodology. This model finds out that the infection rate will gradually increase but at the same time, it is also observed that the recovery rate will increase rapidly as compared to the death rate. In the future, we will analyze this rapid rate of recovery as compared to the death and some other social factors like public awareness and personal belief of the public regarding reality and severeness of the Coronavirus.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Ethical Approval

Approval was not required.

Reference

Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., Rabczuk, T., & Atkinson, P. M. (2020). COVID-19 Outbreak Prediction with Machine Learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3580188

Avery, C., Bossert, W., Clark, A., Ellison, G., & Ellison, S. F. (2020). Policy Implications of Models of the Spread of Coronavirus: Perspectives and Opportunities for Economists. *National Bureau of Economic Research*. https://doi.org/10.3386/w27007

Binti Hamzah, F. A., Lau, C. H., Nazri, H., Ligot, D. C., Lee, G., Tan, C. L., & et al. (2020). CoronaTracker: World-wide Covid-19 outbreak data analysis and prediction. *Bulletin of the World Health Organization*, *March*, Submitted.

Chakraborty, T., & Ghosh, I. (2020). Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons and Fractals*, *135*. https://doi.org/10.1016/j.chaos.2020.109850

Fong, S. J., Li, G., Dey, N., Gonzalez-Crespo, R., & Herrera-Viedma, E. (2020). Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak. *International Journal of Interactive Multimedia and Artificial Intelligence*, *6*(1), 132. https://doi.org/10.9781/ijimai.2020.02.002

Janies, D., Habib, F., Alexandrov, B., Hill, A., & Pol, D. (2008). *Cladistics*. *24*, 111–130.

Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., Wang, D., Chen, G., Zhang, J., Peng, H., & Shao, Y. (2020). Propagation analysis and prediction of the COVID-19. *Infectious Disease Modelling*, *5*, 282–292. https://doi.org/10.1016/j.idm.2020.03.002

Li, Y., Wang, B., Peng, R., Zhou, C., Zhan, Y., Liu, Z., Jiang, X., & Zhao, B. (2020). Mathematical Modeling and Epidemic Prediction of COVID-19 and Its Significance to Epidemic Prevention and. *Annals of Infectious Disease and Epidemiology*, *5*(1), 1052.

Murray, C. J. (2020). *Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months*. *114*. https://doi.org/10.1101/2020.03.27.20043752

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS ONE*, *15*(3), 1–8. https://doi.org/10.1371/journal.pone.0231236

Qasim, M., Ahmad, W., Zhang, S., Yasir, M., & Azhar, M. (2020). *Data model to predict prevalence of COVID-19 in Pakistan*. https://doi.org/10.1101/2020.04.06.20055244

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*(7). https://doi.org/10.3390/ijerph17072365

Stübinger, J., & Schneider, L. (2020). Epidemiology of Coronavirus COVID-19: Forecasting the Future Incidence in Different Countries. *Healthcare*, *8*(2), 99. https://doi.org/10.3390/healthcare8020099

Tiwari, S., Kumar, S., & Guleria, K. (2020). Outbreak trends of CoronaVirus (COVID-19) in India: A Prediction. *Disaster Medicine and Public Health Preparedness*, *May*. https://doi.org/10.1017/dmp.2020.115

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

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., Bonten, M. M. J., Damen, J. A. A., Debray, T. P. A., De Vos, M., Dhiman, P., Haller, M. C., Harhay, M. O., Henckaerts, L., Kreuzberger, N., Lohmann, A., Luijken, K., Ma, J., Andaur Navarro, C. L., … Van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *The BMJ*, *369*. https://doi.org/10.1136/bmj.m1328

Yan, L., Zhang, H. T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., Huang, X., Xiao, Y., Cao, H., Chen, Y., Ren, T., Wang, F., Xiao, Y., Huang, S., Tan, X., … Yuan, Y. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence*, *2*(5), 283–288. https://doi.org/10.1038/s42256-020-0180-7

Yang, Z., Zeng, Z., Wang, K., Wong, S. S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., … He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3), 165–174. https://doi.org/10.21037/jtd.2020.02.64

Zhang, G., Pang, H., Xue, Y., Zhou, Y., & Wang, R. (2020). *Forecasting and Analysis of Time Variation of Parameters of COVID-19 Infection in China Using An Improved SEIR Model*. 1–6. https://doi.org/10.21203/rs.3.rs-16159/v1

Zhang, S., Diao, M. Y., Yu, W., Pei, L., Lin, Z., & Chen, D. (2020). Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *International Journal of Infectious Diseases*, *93*, 201–204. https://doi.org/10.1016/j.ijid.2020.02.033