International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

A Data Analysis of Global Pandemic Coronavirus using Machine Learning Algorithms

Sakshi Jain

B.E, Department of Computer Science, P.D.M University, Haryana, India

Abstract: ***The world is experiencing so many disturbing thing at social, political, natural and health level. The outbreak of COVID-19 Coronavirus, namely SARS-CoV-2, has become a global issue and declared as pandemic by world health organization. Many countries have implemented restrictions on population movement to slow the spread of severe acute respiratory syndrome coronavirus 2 and prevent health systems from becoming overwhelmed; some countries have instituted full or partial lockdowns. There are several outbreak prediction models for COVID-19 are being used around the world to make informed decisions and enforce relevant control measures. We try to concentrate on COVID-19 predictions using Machine Learning Techniques.The covid-19 virus second mutate has been found which increase infection rate more rapidly and increase the number of patient. The Data such as the number of persons are infected by the virus in the world, total number of persons infected by this virus in India, number of persons infected by this virus per day, the number of persons deaths, number of persons per day, number of persons recover due to this virus is used as an input and these features are modeled for analysis and prediction.***

Keywords: Forex Reserve, Foreign currency, Machine Learning, Arima, Predictive Analysis, Fbprophet, Forecasting Model, Central bank (RBI)

1. Introduction

Access to accurate outbreak prediction models is essential to obtain insights into the likely spread and consequences of infectious diseases.

After the first case of coronavirus (Covid-19) come out from Wuhan, China. Sudan, Africa suspected two caes of coronavirus infection and started coming from everywhere from the world.

World Health Organization has been declared COVID-19 as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020.

People face various health issues but recently we found that the coronavirus has start mutating itself and become more complected to inspect through RT-PCR test.

The first case of coronavirus reported:
According to unpublished repots from government of china, the first case of coronavirus traced back to 17 December 2019, a 55 year old Hubei province citizen. There were four men and five women reported to be infected in November, but none of them were "patient zero".

2. Why it is called COVID 19?

Coronaviruses are a family of viruses that can cause illnesses such as the common cold, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS).

'CO' stands for corona, 'VI' for virus, and 'D' for disease. Viruses are named based on their genetic structure to facilitate the development of diagnostic tests, vaccines and medicines.

Coronaviruses, named for the crown-like spikes on their surfaces, are a large family of viruses that are common in people and many different species of animals, including camels, cattle, cats, and bats.

Animal coronaviruses rarely infect people and then spread between people. This occurred with two earlier coronaviruses, MERS-CoV and SARS-CoV.

SARS-CoV-2 virus is a beta coronavirus, like MERS-CoV and SARS-CoV. These viruses have their origins in bats.

The virus itself is named SARS-CoV-2. Each part of the name is an abbreviation: SARS stands for “severe acute respiratory syndrome”

3. Symptoms

Signs and symptoms of coronavirus disease 2019 (COVID-19) may appear two to 14 days after exposure.
Common signs and symptoms can include:

i. Fever
ii. Cough
iii. Tiredness

Other symptoms can include:
Shortness of breath or difficulty breathing

i. Muscle aches
ii. Chills
iii. iv. Sore throat
iv. v. Runny nose
v. Headache
vi. Chest pain
vii. Pink eye (conjunctivitis)
viii. Nausea
ix. Vomiting
x. Diarrhea
xi. Rash

This list is not all inclusive. Children have similar symptoms to adults and generally have mild illness.

4. Literature Review

1) **PaperTitle**: COVID-19 Outbreak Prediction with Machine Learning.

Author: Sina F. Ardabili, Amir Mosavi, PedramGhamisi, Filip Ferdinand, Annamaria R. Varkonyi-Koczy, Uwe Reuter, TimonRabczuk, Peter M. Atkinson

Objective: analysis of machine learning and soft computing models to predict the COVID-19 outbreak

Conclusion: Although the most difficult prediction is to estimate the maximum number of infected patients, estimation of the individual mortality rate (n(deaths)/n(infected)) is also essential. The mortality rate isparticularly important to accurately estimate the number of patients and the required beds in intensive care unit.

2) **Title Paper:** Machine learning-based prediction of COVID-19 diagnosis based on symptoms

Author: Yazeed Zoabi, Shira Deri-Rozov and Noam Shomron

Objective: Prediction models that combine several features to estimate the risk of infection have been developed.

These aim to assist medical staff worldwide in triaging patients, especially in the context of limited healthcare resources.

Conclusion: Close contact with an individual confirmed to have COVID-19 was also an important feature, thus corroborating the disease's high transmissibility15 and highlighting the importance of social distancing.

3) **Title Paper:** Intelligent system for COVID-19 prognosis: a state-of-the-art survey

Author: Janmenjoy Nayak & Bighnaraj Naik & Paidi Dinesh & Kanithi Vakula& B. Kameswara Rao

Objective: Analyzing the impact of data types and the nature of data along with challenges in processing the data for and to focus on some future challenges in COVID-19

Conclusion: The COVID-19 had a major impact on the global economy. These are extremely unsure times for the economic
markets, with nations around the world anguish the threatening effects of the epidemic.

4) **Title Paper: Predictive Analytics of COVID-19 using Information,Communication and Technologies**

Author: Parikshit N. Mahalle, etal.

Objective: Analysis of various predictive analytics methods

Conclusion: Simulation result of this model shows that the confirmed COVID-19 infected cases would be 1.6 million and 2.3 million by end of May and June respectively.

5) **Title Paper:** Deep Learning applications for COVID-19

Author: Connor Shorten, Taghi M. Khoshgoftaar and BorkoFurht

Objective: Precision Diagnostics, Protein Structure Prediction, and Drug Repurposing and utilized in Spread Forecasting for Epidemiology.

Conclusion: Explored across data domains in Natural Language Processing, Computer Vision, Life Sciences, and Epidemiology.

5. Research Methodology

ARIMA: An autoregressive integrated moving average, or ARIMA, is very popular and widely used a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.

Briefly, Autoregression (AR) refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.

Integrated (I) represents the differencing of raw observations to allow for the time series to become stationary, i.e., data values are replaced by the difference between the data values and the previous values.

Moving average (MA) incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

A linear regression model is constructed including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e. to remove trend and seasonal structures that negatively affect the regression model.

A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model.

Adopting an ARIMA model for a time series assumes that the underlying process that generated the observations is an ARIMA process. This may seem obvious, but helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model.

Prophet Model
Prophet is one of the newest buzzwords in the field of time series forecasting.

Prophet has many parameters that could be set, like growth, changepoints, holidays, interval_width and weekly_seasonality. Growth is used to specify a trend, linear or logistic. Other parameters are very intuitive.
Prophet has its implementation in Python and R.

HowProphetworks:
At its core, the Prophet procedure is an additive regression

model with four main components:

A piecewise linear or logistic growth curve trend. Prophet automatically detects changes in trends by selecting change points from the data.

6. Limitations

The major limitations of the study are

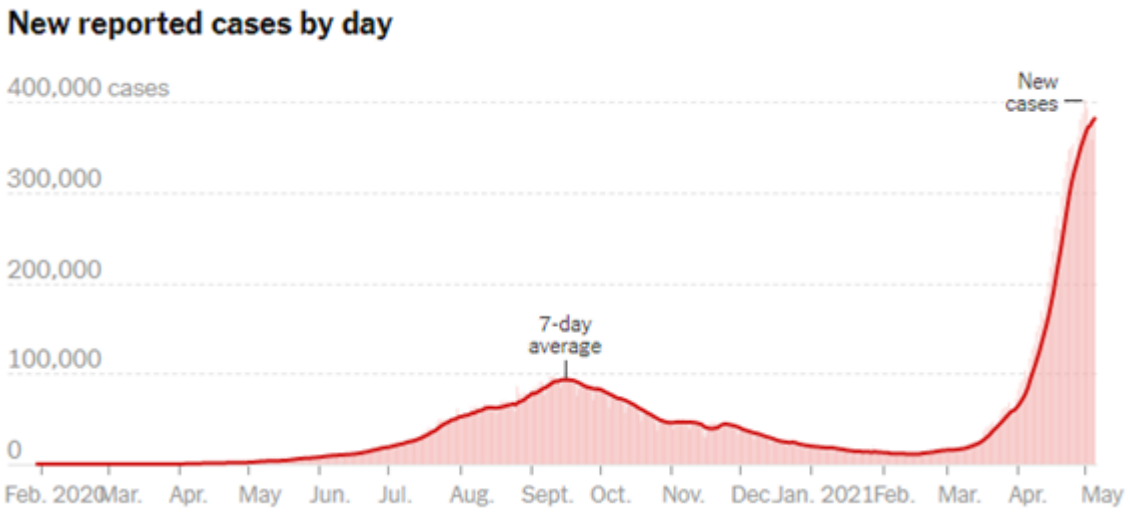
- The study is made only in consideration with India and not applicable to any part of the globe.
- The study fully depends on the secondary data, which has its own limitations.

7. Data Analysis and Interpretation

[43]

	Date	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	No. of countries
0	2020-01-22 00:00:00	555	17	28	510	0	0	0	3.060000	5.050000	60.710000	6
1	2020-01-23 00:00:00	654	18	30	606	99	1	2	2.750000	4.590000	60.000000	8
2	2020-01-24 00:00:00	941	26	36	879	287	8	6	2.760000	3.830000	72.220000	9
3	2020-01-25 00:00:00	1434	42	39	1353	493	16	3	2.930000	2.720000	107.690000	11
4	2020-01-26 00:00:00	2118	56	52	2010	684	14	13	2.640000	2.460000	107.690000	13
5	2020-01-27 00:00:00	2927	82	61	2784	809	26	9	2.800000	2.080000	134.430000	16
6	2020-01-28 00:00:00	5578	131	107	5340	2651	49	46	2.350000	1.920000	122.430000	16
7	2020-01-29 00:00:00	6166	133	125	5908	588	2	18	2.160000	2.030000	106.400000	18
8	2020-01-30 00:00:00	8234	171	141	7922	2068	38	16	2.080000	1.710000	121.280000	20
9	2020-01-31 00:00:00	9927	213	219	9495	1693	42	78	2.150000	2.210000	97.260000	24
10	2020-02-01 00:00:00	12038	259	281	11498	2111	46	62	2.150000	2.330000	92.170000	25
11	2020-02-02 00:00:00	16787	362	459	15966	4749	103	178	2.160000	2.730000	78.870000	25

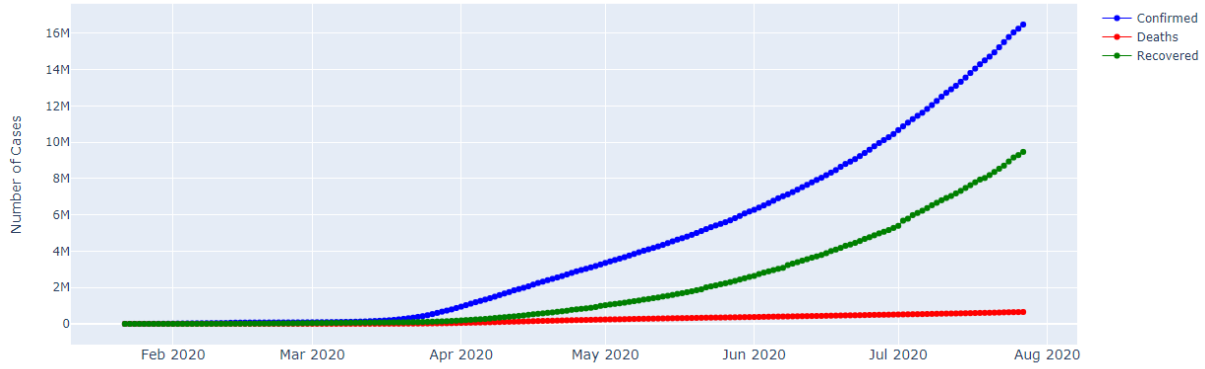
Covid cases starting data



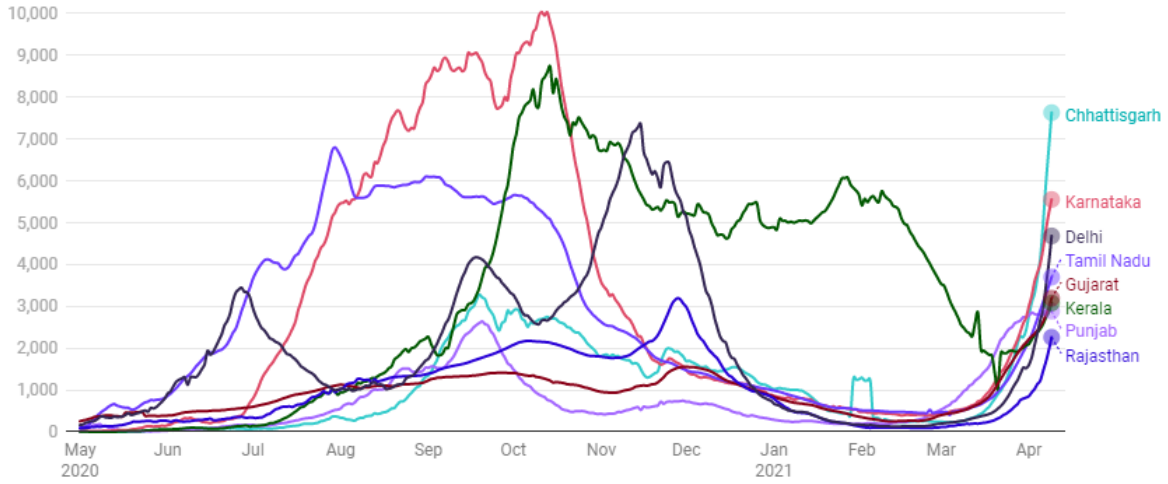
The Seven days average and previous six days data

	UID	iso2	iso3	code3	FIPS	Admin2	Province_State	Country_Region	Lat	Long_	Combined_Key	Date	Confirmed	Deaths
0	16	AS	ASM	16	60.0	NaN	American Samoa	US	-14.271000	-170.132000	American Samoa, US	1/22/20	0.0	0.0
1	316	GU	GUM	316	66.0	NaN	Guam	US	13.444300	144.793700	Guam, US	1/22/20	0.0	0.0
2	580	MP	MNP	580	69.0	NaN	Northern Mariana Islands	US	15.097900	145.673900	Northern Mariana Islands, US	1/22/20	0.0	0.0
3	63072001	PR	PRI	630	72001.0	Adjuntas	Puerto Rico	US	18.180117	-66.754367	Adjuntas, Puerto Rico, US	1/22/20	0.0	0.0
4	63072003	PR	PRI	630	72003.0	Aguada	Puerto Rico	US	18.360255	-67.175131	Aguada, Puerto Rico, US	1/22/20	0.0	0.0

World Covid start cases



Growth of confirmed, death and recovered



Covid second wave in states

8. Observation

The number of cases is increase rapidly in the second wave of coronavirus and its become more dangerous due to its mutation. A big rise of confirmed cases from the month of april and it reach to its peak in mid of may. The mortality rate was its peak in between last april to min may after that the number of active patients will start decreasing which indicate the recovery of patients increased.

Top 5 Countries, highly affected by corona virus.

1) United State
2) India
3) Brazil
4) France
5) Turkey

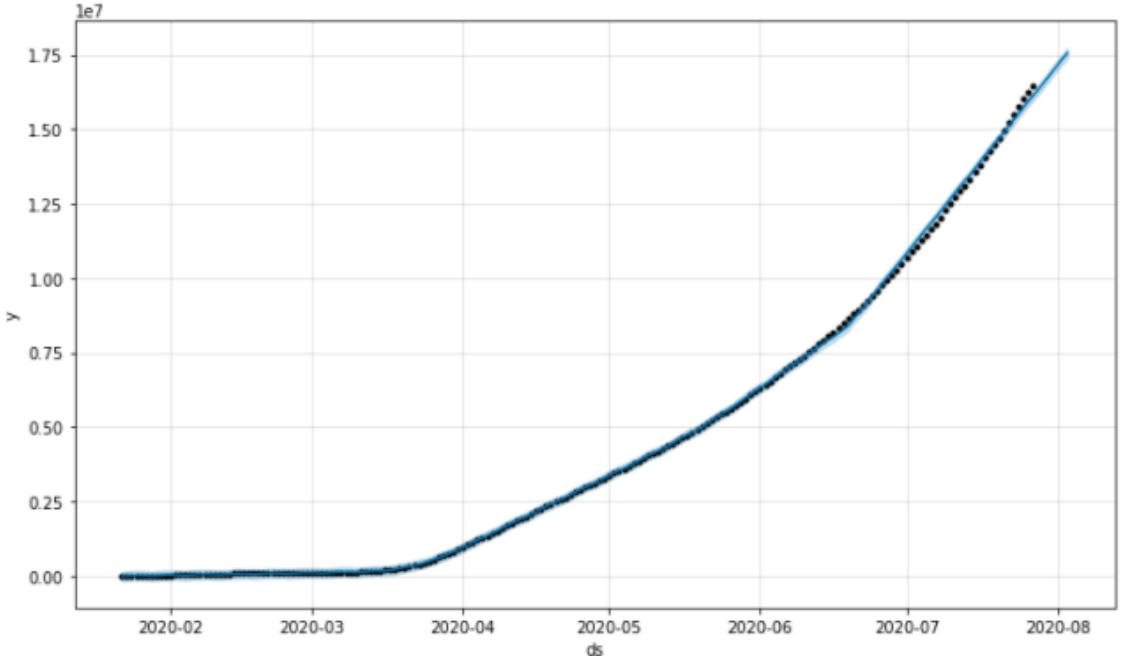
	ds	yhat	yhat_lower	yhat_upper
190	2020-07-30	1.674593e+07	1.658745e+07	1.690256e+07
191	2020-07-31	1.696120e+07	1.679334e+07	1.712960e+07
192	2020-08-01	1.716899e+07	1.698455e+07	1.734716e+07
193	2020-08-02	1.736665e+07	1.718182e+07	1.754349e+07
194	2020-08-03	1.756134e+07	1.736239e+07	1.774413e+07

Number of cases with upper and lower limites

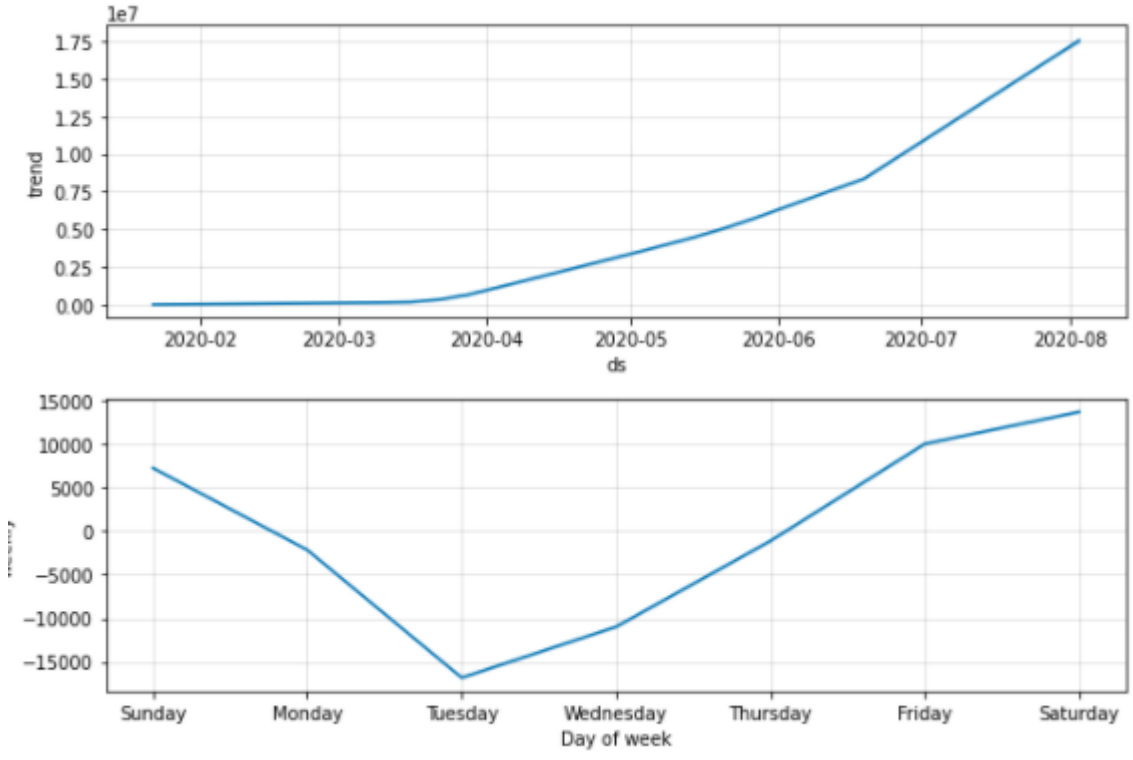
	Date	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	No. of countries
0	2020-01-22	555	17	28	510	0	0	0	3.06	5.05	60.71	6
1	2020-01-23	654	18	30	606	99	1	2	2.75	4.59	60.00	8
2	2020-01-24	941	26	36	879	287	8	6	2.76	3.83	72.22	9
3	2020-01-25	1434	42	39	1353	493	16	3	2.93	2.72	107.69	11
4	2020-01-26	2118	56	52	2010	684	14	13	2.64	2.46	107.69	13
...	...	...	...	...	...	...	...	...	...	...	...	...
183	2020-07-23	15510481	633506	8710969	6166006	282756	9966	169714	4.08	56.16	7.27	187
184	2020-07-24	15791645	639650	8939705	6212290	281164	6144	228736	4.05	56.61	7.16	187
185	2020-07-25	16047190	644517	9158743	6243930	255545	4867	219038	4.02	57.07	7.04	187
186	2020-07-26	16251796	648621	9293464	6309711	204606	4104	134721	3.99	57.18	6.98	187
187	2020-07-27	16480485	654036	9468087	6358362	228693	5415	174623	3.97	57.45	6.91	187

188 rows × 12 columns

Confirmed cases model forcast



The prediction model the graph shows limit of increase



The forecast of growth of cases

9. Conclusion

At this time the Covid situation is very serious. Its start mutating and become more dangerous then before. It is important to remember that to recover from this pandamic situation we need to follow the guidelines very strictly. And vaccinate our self.

References

[1] Kim, T.H.; Hong, K.J.; Shin, S.D.; Park, G.J.; Kim, S.; Hong, N. Forecasting respiratory infectious outbreaks using ED-based syndromic surveillance for febrile ED visits in a Metropolitan City. Am. J. Emerg. Med. 2019, 37, 183–188.

[2] Zhong, L.; Mu, L.; Li, J.; Wang, J.; Yin, Z.; Liu, D. Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model. IEEE Access 2020, 8, 51761–51769.

[3] Wu, J.T.; Leung, K.; Leung, G.M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. Lancet 2020, 395, 689–697.

[4] Reis, J.; Yamana, T.K.; Kandula, S.; Shaman, J. Superensemble forecast of respiratory syncytial virus outbreaks at national, regional, and state levels in the United States. Epidemics 2019, 26, 1–8. [CrossRef] Algorithms 2020, 13, 249 33 of 36

[5] Burke, R.M.; Shah, M.P.; Wikswo, M.E.; Barclay, L.; Kambhampati, A.; Marsh, Z.; Cannon, J.L.; Parashar, U.D.; Vinjé, J.; Hall, A.J. The Norovirus Epidemiologic Triad: Predictors of Severe Outcomes in US Norovirus Outbreaks, 2009–2016. J. Infect. Dis. 2019, 219, 1364–1372. [CrossRef] 30. Carlson, C.J.; Dougherty, E.R.; Boots, M.; Getz, W.M.; Ryan, S.J. Consensus and conflict among ecological forecasts of Zika virus outbreaks in the United States. Sci. Rep. 2018, 8, 4921.

[6] Reported Cases and Deaths by Country, Territory, or Conveyance. Available online: https://www. worldometers.info/coronavirus/#countries (accessed on 30 September 2020).

[7] Back, T. Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms; Oxford University Press: Oxford, UK, 1996.

[8] Deb, K. Multi-Objective Optimization Using Evolutionary Algorithms; John Wiley & Sons: Hoboken, NJ, USA, 2001; Volume 16.

[9] Zitzler, E.; Deb, K.; Thiele, L. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. Evol. Comput. 2000, 8, 173–195.

[10] Ghamisi, P.; Benediktsson, J.A. Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization. IEEE Geosci. Remote. Sens. Lett. 2014, 12, 309–313.

[11] Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. Advances in engineering software 2014, 69, 46- 61.

[12] Ardabili, S.; Mosavi, A.; Varkonyi-Koczy, A. Building Energy information: demand and consumption prediction with Machine Learning models for sustainable and smart cities. Engineering for Sustainable Future, Lecture Notes in Networks and Systems, Springer Nature Switzerland.: 2019. 72. Ardabili;, S.; Mosavi;, A.; Varkonyi-Koczy;, A. Systematic Review of Deep Learning and Machine Learning Models in Biofuels Research,. Engineering for Sustainable Future, Lecture Notes in Networks and Systems, Springer Nature Switzerland 2019.

[13] Gundoshmian, T.M.; Ardabili, S.; Mosavi, A.; Varkonyi-Koczy, A.R. Prediction of combine harvester performance using hybrid machine learning modeling and response surface methodology. 2019.

[14] Nosratabadi, S., Mosavi, A., Duan, P. and Ghamisi, P., 2020. Data Science in Economics. arXiv preprint arXiv:2003.13422.

[15] Hamilton, J.D. Time series analysis; Princeton university press Princeton, NJ: 1994; Vol. 2.

[16] Ardabili, S.; Mosavi, A.; Mahmoudi, A.; Gundoshmian, T.M.; Nosratabadi, S.; Varkonyi-Koczy, A.R. Modelling Temperature Variation of Mushroom Growing Hall Using Artificial Neural Networks. Engineering for Sustainable Future, Lecture Notes in Networks and Systems, Springer Nature Switzerland. 2019

[17] Xu B, Kraemer MU, Gutierrez B, Mekaru S, Sewalk K, Loskill A, Li S (2020) Open access epidemiological data from the COVID19 outbreak. Lancet Infect Dis 20(5):534

[18] Frazer JS, Shard A, Herdman J (2020) Involvement of the opensource community in combating the worldwide COVID-19 pandemic: a review. J Med EngTechnol 44(4):169–176

[19] Yang W, Yan F (2020) Patients with RT-PCR-confirmedCOVID19 and normal chest CT. Radiology 295(2):E3–E3

[20] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Xia L (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 296(2):200642.

[21] Farooq J, Bazaz MA (2020) A novel adaptive deep learning model of Covid-19 with focus on mortality reduction strategies. Chaos, Solitons Fractals 138:110148

[22] Loey M, Smarandache F, Khalifa M (2020) Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. Symmetry 12(4):651

[23] Pathak Y, Shukla PK, Tiwari A, Stalin S, Singh S, Shukla PK (2020) Deep transfer learning based classification model for COVID-19 disease. IRBM, Elsevier, Amsterdam

[24] Johns Hopkins CSSE, Novel Coronavirus (COVID-19) Cases, 2020, https://github.com/CSSEGISandData/COVID-19.

[25] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT press, 2016.

[26] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[27] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” 2017, http://arxiv.org/abs/1710.05941.

[28] COVID-19 MLP, Riteh AI and Robotics Group, 2020,

https:// github.com/RitehAIandRobot/COVID-19-MLP.
[29] Chan D, Rao R, Huang F, Canny J. t-SNE-CUDA: GPU-Accelerated t-SNE and its applications to modern data.
[30] Narang S, Rafel C, Lee K, Roberts A, Fiedel N, Malkan K. WT5?! training text-to-text models to explain their predictions; 2020. arXiv:2004.14546.
[31] Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples; 2014. arXiv:1412.6572
[32] Ahamed S, Samad M. Information Mining for COVID-19 research from a large volume of scientifc literature; 2020. arXiv:2004.0208.
[33] Rao R, Bhattacharya N, Thomas N, Duan Y, Chen X, Canny J, Abbeel P, Song YS. Evaluating protein transfer learning with TAPE; 2019. arXiv:1906:08230.
[34] Gao W, Mahajan S, Sulam J, Gray J. Deep learning in protein structural modeling and design