

COVID-19 HOTSPOT ANALYSIS AND PREDICTION

Shanmuga Sundari N¹, Anandhavalli D², Karthikeyan A³, Ramachandran R⁴, VikashRaj S⁵

¹⁻⁵Department of Information Technology, Velammal College of Engineering and Technology, Madurai-625009.

Abstract - *The flare-up of Corona Virus Disease 2019 (COVID-19) in Wuhan has significantly impacted the economy and society globally. Countries are in a state of prevention and control of this pandemic. In this study, the improvement pattern investigation of the aggregate confirmed cases, total passing, and total relieved cases was directed based on data from Wuhan, Hubei Province, China from January 23, 2020 to April 6, 2020 utilizing a relapse calculation, neural net model. The developed machine learning model can work in real-time and can adequately foresee the quantity of positive cases. Key measures and ideas have been advanced thinking about the impact of lockdown. This paper assembles prescient models that can anticipate the quantity of positive cases with higher exactness. Regression based, Support Vector Machine, and Holt Linear Regression models have been based on the information from around the world. The model is found to be effective and will be able to analyze the most hotspot are with the predicted number of positive cases in the future with minimal error.*

Keywords — Machine Learning, Support Vector Machine, Linear Regression Model, Analysis.

1. INTRODUCTION

Infectious diseases are brought about by different microbes that can be communicated from one individual to another, creature to creature, or individual to creature. They can be sent differently, and the speed of transmission is quick. Early finding of irresistible sicknesses is significant, and prevention and control are paramount. At the finish of 2019, the new (COVID-19) spread broadly in China, and an enormous number of individuals became infected. At present, the homegrown flare-up has been viably controlled, while the new Covid is spreading quickly in other areas. Presently, Europe has gotten the focal point of the current flare-up pneumonia. The new Covid has made an incredible danger the wellbeing and security of individuals everywhere on the world because of its astonishing spreading force and likely damage. The infection, named the 2019 Novel Corona infection (2019-nCoV) by the World Health Organization (WHO) on

January 12, 2020, the infection can be available asymptotically, making tainted individual seem sound, or can show with different indications in which uninfected individual is by all accounts experiencing a typical respiratory sickness. It additionally has a long idleness period. As per report from Lancet, the middle dormancy season of the new crown viruses roughly 20 days, coming to as long as 37 days for certain patients.

2. LITERATURE SURVEY

There have been a few analysts adding to the territories of gauging the pandemic. Remuzzi and Remuzzi have talked about the genuine effect that Coronavirus2 (SARS-CoC-2) have caused to China and its further spread to Italy. They fabricated a prescient model that will help comprehend the patient's raise. Furthermore, this thusly would assist the clinical offices with taking choices. Lobby, Gani, Hughes, and Leachalso contemplated the spread of the H5N1 flu infection in birds and have anticipated utilizing relapse investigation the circumstance of the commonness of the pandemic wave. In the new past, there is research considering China information to anticipate the quantity of cases. Hu, Ge, Jin, and Xiong have proposed a fake intelligence-based technique for real-time anticipating of COVID-19 to assess the size, lengths, and finishing season of COVID-19 across China. The information that was utilized was from the time of January 11 to February 27, 2020, distributed by the World Health Organization (WHO). It was deduced in their work that the spread would lessen by April. Anastassopoulou, Russo, Tsakris, and Siettos stated that the openly accessible information from the epidemiological information of Hubei, China, an assessment on case casualty and case recuperate recuperation proportions were advanced. Their investigation uncovers that there would be a drop in the quantity of cases by January 26, 2020. It also has a long latency period. According to a report from Lancet, the median latency time of the new corona viruses approximately 20 days, reaching up to 37 days for some patients. [1]

S.NO	Algorithm	Accuracy
1.	Linear Regression	66.3%
2.	Support Vector Machine	72.8%
3.	Holt Linear Regression Model	89.4%

Table. 1. Comparison of efficiency with various machine learning algorithms

3. PROPOSED WORK

In this section, the proposed system for covid-19(confirmed/recovered/death) in all prediction is described. As a first step it would be of use to look at the growth and impact of Corona throughout the world and also a comparison between the recent outbreaks like Severe Acute Respiratory Syndrome (SARS) and Ebola was taken into consideration to give a fair understanding of how rapid and contagious is the growth of COVID-19.

Machine-learning has demonstrated to be priceless in anticipating hazards in numerous circles. With clinical danger explicitly, Machine-learning is conceivably fascinating in three key manners.

- Infection risk: What is the danger of a particular individual or gathering getting COVID-19?
- Severity risk: What is the danger of a particular individual or gathering creating extreme COVID-19 side effects or intricacies that would require hospitalization or escalated care?
- Outcome risk: What is the danger that a particular treatment will be inadequate for someone in particular or gathering, and how probably would they say they are to kick the bucket?

Machine learning can possibly help anticipate each of the three dangers. Despite the fact that it's still too soon for much COVID-19-explicit AI examination to have been led and distributed, early analyses are promising. Moreover, we can see how Machine learning is utilized in related regions and envision how it could assist with hazard forecast for COVID19.[2]

Phase I: Data Collection

The process of measuring the overall collection of data and gathering the information with the help of software. There are many techniques and procedures to collect the

data. We will be collecting the quantitative data which is structured and categorized. Observing the wellbeing circumstance, patterns, progress and execution of wellbeing frameworks requires information from numerous sources on a wide assortment of wellbeing themes. Before any kind of machine learning analysis, data collection is required. However, validity of the dataset is a must otherwise there is no point in analyzing the data. Hence, we need to be careful about the source of the data and check its validity beforehand.

1	Date	Country/# Province/ Lat	Long	Confirmed/Recovered/Deaths
2	22-01-2020	Afghanistan	33	65 0 0
3	23-01-2020	Afghanistan	33	65 0 0
4	24-01-2020	Afghanistan	33	65 0 0
5	25-01-2020	Afghanistan	33	65 0 0
6	26-01-2020	Afghanistan	33	65 0 0
7	27-01-2020	Afghanistan	33	65 0 0
8	28-01-2020	Afghanistan	33	65 0 0
9	29-01-2020	Afghanistan	33	65 0 0
10	30-01-2020	Afghanistan	33	65 0 0
11	31-01-2020	Afghanistan	33	65 0 0
12	01-02-2020	Afghanistan	33	65 0 0
13	02-02-2020	Afghanistan	33	65 0 0
14	03-02-2020	Afghanistan	33	65 0 0
15	04-02-2020	Afghanistan	33	65 0 0
16	05-02-2020	Afghanistan	33	65 0 0
17	06-02-2020	Afghanistan	33	65 0 0
18	07-02-2020	Afghanistan	33	65 0 0
19	08-02-2020	Afghanistan	33	65 0 0
20	09-02-2020	Afghanistan	33	65 0 0
21	10-02-2020	Afghanistan	33	65 0 0
22	11-02-2020	Afghanistan	33	65 0 0
23	12-02-2020	Afghanistan	33	65 0 0

Fig.1.Dataset for Analysis

1	Sno	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
2	1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1	0	0
3	2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14	0	0
4	3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6	0	0
5	4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1	0	0
6	5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0	0	0
7	6	01/22/2020	Guangdong	Mainland China	1/22/2020 17:00	26	0	0
8	7	01/22/2020	Guangxi	Mainland China	1/22/2020 17:00	2	0	0
9	8	01/22/2020	Guizhou	Mainland China	1/22/2020 17:00	1	0	0
10	9	01/22/2020	Hainan	Mainland China	1/22/2020 17:00	4	0	0
11	10	01/22/2020	Hebei	Mainland China	1/22/2020 17:00	1	0	0
12	11	01/22/2020	Heilongjiang	Mainland China	1/22/2020 17:00	0	0	0
13	12	01/22/2020	Henan	Mainland China	1/22/2020 17:00	5	0	0
14	13	01/22/2020	Hong Kong	Hong Kong	1/22/2020 17:00	0	0	0
15	14	01/22/2020	Hubei	Mainland China	1/22/2020 17:00	444	17	28
16	15	01/22/2020	Hunan	Mainland China	1/22/2020 17:00	4	0	0
17	16	01/22/2020	Inner Mongolia	Mainland China	1/22/2020 17:00	0	0	0
18	17	01/22/2020	Jiangsu	Mainland China	1/22/2020 17:00	1	0	0
19	18	01/22/2020	Jiangxi	Mainland China	1/22/2020 17:00	2	0	0
20	19	01/22/2020	Jilin	Mainland China	1/22/2020 17:00	0	0	0
21	20	01/22/2020	Liaoning	Mainland China	1/22/2020 17:00	2	0	0
22	21	01/22/2020	Macau	Macau	1/22/2020 17:00	1	0	0
23	22	01/22/2020	Ningxia	Mainland China	1/22/2020 17:00	1	0	0

Fig.2.Dataset for Prediction

Phase II: Analysis

Here is the manner by which a few countries saw the pandemic attack, America has been the most exceedingly awful influenced country by Covid-19 and its count will before long touch the bleak sign of 20 million cases. The nation has been hit by a large number of rushes of the infection.[3]

pandas, choropleth, matplotlib, plotly.express these are the built-in functions we are used. plotly.express is a python visualization library, it's a wrapper for plotly.py that exposes a simple syntax for complex charts. choropleth it comes from plotly.express and it is used for rendering the view in a map.[5]

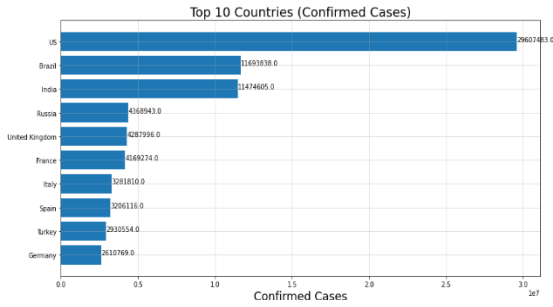


Fig.3. Top 10 Confirmed Cases

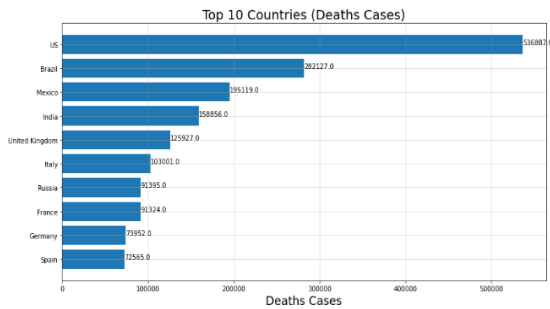


Fig.4. Top 10 Deaths Cases

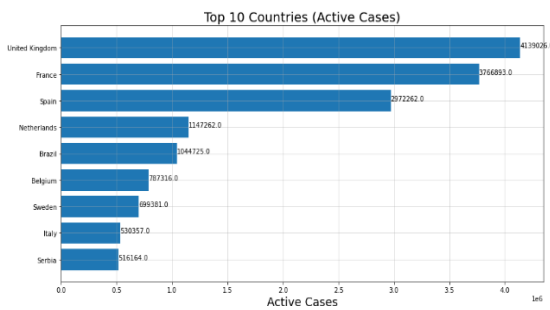


Fig.5. Top 10 Active Cases

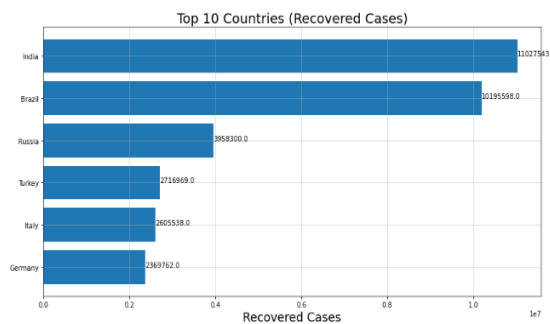


Fig.6. Top 10 Recovered Cases

Phase III: Prediction

The present study has aimed to predict COVID- 19 using recovery rate and active cases load rate on basis of available data.[6]in addition with pandas, matplotlib library functions like sklearn.svm, sklearn.linear model,

statsmodel.tsa.api and sklearn metrics we are used for prediction. numpy is an open source numerical python library. numpy has a multidimensional array and matrix data structures.

This will be used to perform mathematical operations on arrays. The class sklearn, linear model and linear regression are used to perform linear to make predictions from globally. Finally, Support vector machine algorithm we are used.

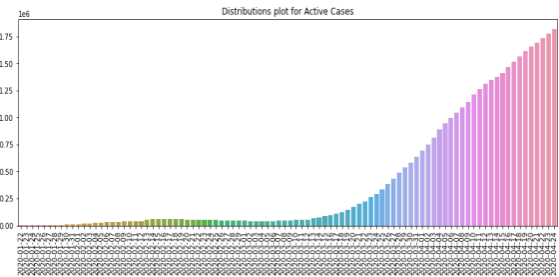


Fig.7. Graphical Representation for Day by Day Active Cases

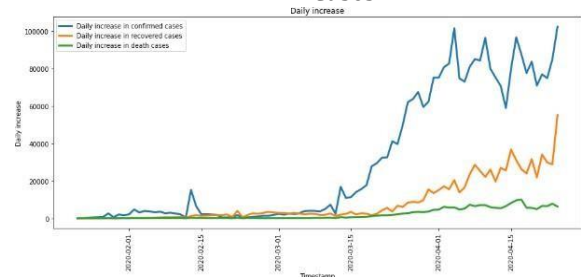


Fig.8. Graphical Representation for Day by Day Confirmed, Recovered and Death Cases

4. RESULTS AND DISCUSSION

The rise of COVID-19 has been a significant weight. During the previous 2 months, in any case, huge advancement has been made in the conclusion and treatment of this infection. In this examination, AI was utilized to anticipate and investigate the improvement pattern and development scope of COVID-19. Hence, we propose a two-fold methodology, where zones with high cases need to limit receptiveness and have exacting measures and open up the economy in places where the cases are smoothed. The opening up of the exercises ought to be focused on with key measures on the hold. Besides, there ought to screen of the quantity of positive

cases at miniature level, ideally at Panchayat level, in close time spans.

This will empower the public authority to make choices on additional course of moves concerning whether to open up the economy further or to bring back the limitations.

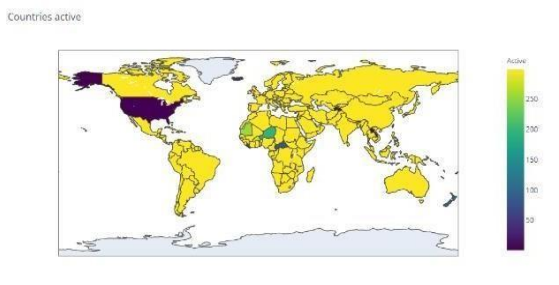


Fig.9. World Map Graphical Representation from 100-300cases

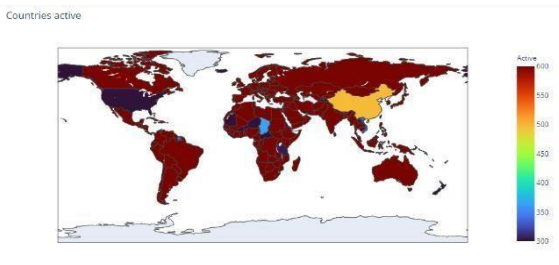
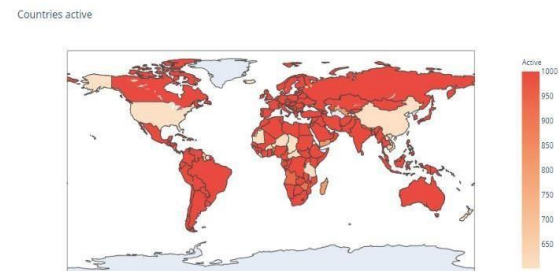


Fig.10. World Map Graphical Representation from 300-600cases



600cases

Fig.11. World Map Graphical Representation from 600-1000cases

Dates	LR	SVR	Holts	Linear Model Prediction
0 2020-04-25	1560529	3322586		2855246
1 2020-04-26	1582219	3500761		2933902
2 2020-04-27	1603909	3686599		3012558
3 2020-04-28	1625599	3880344		3091214
4 2020-04-29	1647289	4082245		3169870

Fig.12. Prediction based on different Algorithms

The exploratory outcomes on information from the United States checked the strength of the models, albeit the expectation aftereffect of LSTM was more awful than that of Wuhan. In future exploration, we will improve the models dependent on the previously mentioned issues and keep on improving the generalizability of the models. [7]

5. CONCLUSION

In this paper, through analyzing the existing data of Hubei epidemic situation, the corresponding model is established, and then the reproduction is completed. Here, we contemplated the principle factors influencing the spread of COVID-19, for example, the number of basic recoveries, the brooding time frame and the normal number of long stretches of fix. Furthermore, we anticipated the evolution trend of the current scourge information, and tracked down that overwhelming controls would significantly affect the plague. In addition, as indicated by the current information abroad, we additionally make strong expectations of the plague advancement patterns in South Korea, Italy, and Iran, calling attention to the potential episodes and the relating control time, and following the earliest transmission dates of nations. At long last, we trust that this article can make a few commitments to the world's reaction to this epidemic and give a few references for future research.

6. REFERENCES

- [1] Ritanjali Majhi, Rahul Thangeda, Renu Prasad, Sugasi, NirajKumar "Analysis and prediction of COVID-19 trajectory : A machine learning approach" <https://onlinelibrary.wiley.com/doi/epdf/10.1002/pa.2537>
- [2] How to fight COVID-19 with machine learning 9 ways machine learning is helping us fight the viral pandemic. By Markus Schmitt <https://www.datarevenue.com/en-blog/machine-learning-covid-19>
- [3] Vikas Kulshreshtha & N.K.Garg "Predicting the New Cases of Corona virus [COVID-19] in India by Using Time Series Analysis as Machine Learning Model in python" <https://link.springer.com/article/10.1007/s40031-021-00546-0>
- [4] Amit Kumar Gupta, Vijander Singh, Priya Mathur& Carlos M. Travieso Gonzalez "Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario" <https://doi.org/10.1080/09720502.2020.1833458>
- [5] Vijander Singh , Ramesh Chandra Poonia , Sandeep Kumar , Pranav Dass , Pankaj Agarwal , Vaibhav Bhatnagar &

Linesh Raja "Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine.

<https://doi.org/10.1080/09720529.2020.1784535>

[6] Prediction and analysis of Corona Virus Disease 2019 Yan Hao, Ting Xu, Hongping Hu, Peng Wang, Yanping Bai. Published: October 5, 2020

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0239960>

[7] Smita Rath, Alakananda Tripathy and Alok Ranjan Tripathy "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model"

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7395225/>

BIOGRAPHIES



I'm Karthikeyan A did my B.Tech in the stream of Information Technology in Velammal College of Engineering and Technology. I have the strong knowledge in technical & logical thinking and interested to work with web and mobile developing.



I'm Ramachandran R did my B.Tech in the stream of Information Technology in Velammal College of Engineering and Technology. I have the strong knowledge in analytical and logical thing and interested work with mobile developing.



Myself Vikash Raj R did my degree B-Tech in the stream of Information Technology and I'm currently working in Maveric Systems as Test Associate through the framework of using selenium.