Data based Analysis, Modeling and Forecasting of Novel Coronavirus in infected regions using Extreme Learning Machine algorithm

A. Saranya¹, S. Aruna², M. Maheswari³

¹Assistant Professor, Department of Software Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.
²Assistant Professor, Department of Software Engineering, SRM Institute of Science and Technology, Kattankulathur
³Assistant Professor, Department of Software Engineering, SRM Institute of Science and Technology, Kattankulathur

E-mail: saranyaa2@srmist.edu.in, arunas@srmist.edu.in, maheswam@srmist.edu.in

Abstract. In current development in science and technology, Machine learning algorithms play an essential role for prediction, classification, data analysis and data visualization. With this efficient algorithm, we can solve many real-world problems in all domains like education, healthcare, banking, geographical analysis, etc., in the current scenario; much research work is going on with the new virus's infection called the corona. This Corona virus is a comprehensive unit of virus this cause illness in humans or animals, now in East Asian countries, this virus affected more people. In India, the first case was found in January month, originated from China. The entire world is focusing on the disease, and day by day, the infection and death rate is increasing. In this, we intended to focus on the spread of this deadly disease and to demonstrate which countries are the most affected by doing statistical analysis. On December 2019, As of 10 February 2020, China reported overall of 40,235 cases 909 deaths, evoking local and foreign terror. Here we provide estimates of the major epidemiological parameters, Based on the epidemiological data available to the public for Hubei, China, 11 January to 10 February 2020. In particular, we give an estimate Fatality and case recovery rates, along with their 90 per cent confidence levels as the epidemic progresses. For this work implementation, Extreme Learning Machine algorithm used. ELM is a feed-forward network and its learning rate also fast when compare to normal neural network. No need to provide any weights and bias values. This algorithm will give a promising result with the best accuracy.

Keywords: Machine Learning algorithm, prediction, classification, statistical analysis, Extreme machine learning algorithm

1. Introduction:

Corona-viruses are a wide range of viruses which can source disease in natural world or individual. Corona viruses’ symptoms are common cold to severe diseases such as respiratory syndrome is well-known to cause human respiratory infections. Novel Corona-virus discovered causes a COVID-19 Corona-virus disease. Many individuals who have the infection get COVID-19. The bad health extend principally From individual to person, by tiny nose or mouth beads that are stripped off when a person with COVID-19 hacks, sniffles or talks. Individuals get this virus infection if they take these beads in. So we have to stay away from others to stop spreading this virus. Such beads for example surface like tables and objects like door handles which are touched by infected person. Person may get contaminated by rubbing their eyes, nose, or mouth after touching the objects or surface which are used by infected person. World Health Organization advised to wash our hands.
for minimum 20 minutes with soap and water. [2] Six modes are proposed by UCCSF, this is self preservation mode, COVID containment mode, economic survival mode, herd immunity mode and Darwinian mode. It has traversed approximately 180,000 confirmed infections worldwide as of March 17, 2020, including 7426 deaths. On 11 March the WHO declared it to be a pandemic. The size of the outbreak has exceeded diseases caused by two other significant corona viruses to date. The overall objective of this work is to focus on the spread of the deadly disease. To demonstrate which countries are the most affected by doing statistical analysis. Machine learning algorithms play an essential role for prediction, classification, data analysis and data visualization. To achieve this proposed work Extreme Learning Machine algorithm ELM algorithm.

1.1 About ELM

[1] Normal feed forward neural network has lac of fast learning method; to solve this problem ELM is proposed. It has the ability to set the hidden layer and randomly assigned input weights. It needs a single iteration for doing the learning process. It overcomes the drawback of the feedforward neural network also improves the learning ability and overcome the Overfitting problem.

Figure:1 Yearly publication using ELM algorithm

This Graph shows that from 2014 onwards research works for doing predictions and data analysis; this ELM shows excellent results of accuracy.

1.2 About Linear Regression

The simplest regression models comprised of a single response variable Y and a single predictor X. STATGRAPHICS will suit several functional types. They will list the models in decreasing R-squared order. When outliers identified, resistant methods may be used instead of least squares to satisfy the models.

Multiple Regression

The multiple regression method suits a model with various predictor variables X1, X2, ... in relation to an answer variable Y. The user can include all predictor variables in the match, or ask the program to pick a subset with only relevant predictors using stepwise regression. Around the same time, the Box-Cox approach and the Cochrane-Orcutt test can be used to tackle non-normality.

To test our model, we’ll use Root Mean Square Error (RMSE) and determination of coefficient (R2 score).

RMSE is square root of the remaining square sum total.
RMSE calculated by the formula

\[ RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (h(x^i) - y^i)^2} \]

R2 score or decision coefficient describes how the use of the least square test will reduce the overall variance of the dependent variable. \( R^2 \) is determined by

\[ R^2 = 1 - \frac{SS_r}{SS_i} \]

Were ss is the total sum of errors if the mean of the observed value as the predicted value.

1.3 About Neural Network:
This neural network is composed of several layers. Each layer is made up of nodes where the computation will take place. A node combines the input, coefficients set and weights. The input weights are multiplied and then summed, the sum given to the node which is called as an activation function. By using Neural Network is used for many application developments.

![Sample Neural Network](image)

**Figure:2 Sample Neural Network**

2. Proposed Work:
This proposed paper consists of several statistical data analysis to identify the spread of viruses throughout India and all over the world. For predicting along with ELM, Linear regression and Multiple regression. Now, India is in critical condition due to drastic increases in the rate of infection and also death. Compared to China India, U.S.A and Italy are facing more difficult because of this virus infection, and all Asian countries have been panic about the rise in death, diseases, and also affecting all nations' economic growth.[4] The impact of this pandemic will last for a long time, affecting many human lives and affecting all global development activities, including technical, financial and investment activities. To handle the situation, our government giving guidance with the help of Public health programs which are an equally essential branch of the health system to protect communities' health and keep them involved in disease prevention.
and leading safe and prosperous lives. Even though the virus spread is increasing drastically, also increase the number of a confirmed case, both India and Tamilnadu. This research work was focusing on analysis which is places are most affected by this disease spread, checking the death rate by doing the statistical analysis. In previous work, by implementing the Arima Model, they predicted that the infection confirmed person in India would go to 27lakhs for India total population.

3. Dataset details

In this paper, two datasets consider for the analysis was taken from Kaggle. The first dataset consists of 40 states of virus infection details. It consists of 9 attributes which are test taken time, date, state. In that state, patients belong to India and foreign countries, no of death, no of confirmed cases and death rate. The second dataset consists of virus infection details for Tamilnadu. All these two datasets collected from January to June and total records are more than 3000.

From the COVID 19 all-state record dataset first we analyzed that what is the total confirmation infected people in a particular state.

![Figure:3 Dataset Statistical analysis](image)

From this Graph, the total confirmed cases, Maharashtra is the most affected place in India and Tamilnadu in second place. This research work is focusing on statistical analysis to demonstrate which are places are most affected by this virus. These insights generated with analytical techniques for forecasting. The prediction and forecasting mostly made on time series data. A linear regression model is used to correct the model for better prediction.

| Dep. Variable: Name of State / UT | R-squared: 0.032 |
|-----------------------------------|------------------|
| Model: OLS                        | Adj. R-squared: 0.032 |
Method: Least Squares  
Date: Mon, 15 Jun 2020  
Time: 10:35:58  
No. Observations: 2916  
No. Model: 1  
Covariance Type: nonrobust  

| coef   | std err | t     | P>|t|     | 0.025      | 0.975       |
|--------|---------|-------|-------|-----------|-------------|
| const  | 18.6114 | 0.211 | 88.165| 0.000     | 18.197      |
| 19.025 |
| Total Confirmed cases | -0.0003 | 2.77e-05 | -9.818 | 0.000 |
| 0.000 |

By applying the linear regression the r^2 value shows the accuracy for the model, so here the accuracy value is 32% to improve the accuracy Multiple regression by considering attributes total confirmed cases, death, cured concerning date and place.

OLS Regression Results

| coef   | std err | t     | P>|t|     | 0.025      | 0.975       |
|--------|---------|-------|-------|-----------|-------------|
| const  | 18.7477 | 0.209 | 89.761| 0.000     | 18.338      |
| x1     | -0.0011 | 0.000 | -7.681| 0.000     | -0.001      |
| 0.001  |
| x2     | 0.0214  | 0.002 | 8.967 | 0.000     | 0.017       |
| 0.026  |
| x3     | 0.0002  | 0.000 | 0.933 | 0.351     | -0.000      |
| 0.001  |
By using Multiple regression $r^2$ value 58%

![Graph showing data visualization for statistical analysis](image)

**Figure: 4 Data Visualization for statistical analysis**

This graph represents the infection spread in various places and to understand the dataset with respect to other information. The table 1 shows the India and Tamilnadu total no of affected case and total number of death and also the death rate with respect to the total population of the individual country and state.

| Date     | India          | Tamil Nadu     |
|----------|----------------|----------------|
| 23/4/2020 | affected cases:31324 | affected cases:2058 |
|          | Death:1008     | Death:25       |
| 23/6/2020 | affected cases:440450 | affected cases:62087 |
|          | Death:14015   | Death:794      |

**Table:1 India and Tamilnadu Infection Details**

Infection and death rate on 23/6/202 with respect to total population.
1) US: Affected 0.72%
   - Death 0.037%
2) Italy: Affected 0.39%
   - Death 0.057%
3) India: Affected 0.032%
   - Death 0.001%
4) Tamil Nadu: Affected 0.086%
   - Death 0.001%

Figure 5: Graph showing Death rate vs Death and Cured

Graph showing that there is a constant death rate even though drastic increase in infection and death in India and Tamil Nadu.

Some of the statistical analysis of infection spread, death, cured and death rate are:

| State/UnionTerritory                  | Confirmed  |
|--------------------------------------|------------|
| Maharashtra                         | 2419367    |
| Tamil Nadu                           | 849167     |
| Delhi                                | 780823     |
| Gujarat                              | 662180     |
| Rajasthan                            | 352485     |
| Uttar Pradesh                        | 325722     |
| Madhya Pradesh                       | 324595     |
| West Bengal                          | 211879     |
| Andhra Pradesh                       | 162672     |
| Cases being reassigned to states     | 155434     |
| Karnataka                            | 129828     |
| Bihar                                | 125858     |
| State/UnionTerritory                  | Confirmed |
|--------------------------------------|-----------|
| Telengana                            | 103803    |
| Jammu and Kashmir                    | 103168    |
| Haryana                              | 101344    |
| Punjab                               | 99816     |
| Odisha                               | 69244     |
| Kerala                               | 62213     |
| Assam                                | 46937     |
| Uttarakhand                          | 27248     |
| Jharkhand                            | 25908     |
| Chhattisgarh                         | 20803     |
| Telangana                            | 18515     |
| Tripura                              | 15535     |
| Chandigarh                           | 11306     |
| Himachal Pradesh                     | 10756     |
| Goa                                  | 5444      |
| Ladakh                               | 4514      |
| Manipur                              | 4056      |
| Puducherry                           | 2594      |
| Andaman and Nicobar Islands          | 2060      |
| Nagaland                             | 1677      |
| Meghalaya                            | 1169      |
| Mizoram                              | 842       |
| Arunachal Pradesh                    | 811       |
| Sikkim                               | 273       |
| Dadar Nagar Haveli                   | 186       |
| Unassigned                           | 161       |
| Dadra and Nagar Haveli and Daman and Diu | 131 |
| Daman & Diu                          | 2         |

Figure: 6 State wise Infection Details

The redder the cell, the bigger the value. So, the darker cells represent a higher number of affected cases, and the lighter ones show otherwise.
Worldwide COVID-19 Cases

Figure: 7 plotting date wise confirmed cases

Figure: 8 Confirmed forecast data frame
Figure: 9 Death rate forecast data frame

Figure: 10 Correlation-Heat map
The heat map is a data analysis software that uses colors which is used as a data visualization tool for the given dataset. Heat map shows the correlation which describes how the features are related to each of the target variables and help us to increase the accuracy for predicting the output.

By implementing ELM algorithm getting the accuracy as more than 85% to improve accuracy by increase the number of hidden units in our model. But in practice, it will take a very long time an

It will run out of RAM. For this neural network, there are total of six inputs they are state/ Union Territory, s.no, cured, deaths, confirmed case and a bias value. In a hidden layer total of 5 neurons with a sigmoid activation function. In sigmoid function is used to reach the optimal value.
4. Conclusion:

According to the study from the famous university, India has witnessed more than 30000 cases between March 24 to April 14 if the government has not issued the lockdown. Now, as per the statistical analysis in India, no of infection cases is increasing. As per the research analysis, compared to the total population of China, USA, Italy, India and Tamil Nadu having the constant death rate even though infection and death are increasing this because of India and Tamil Nadu weather condition. Apart from ICMR, India did not have many laboratories that could provide diagnostic support. So the Indian government is taking the effort to stop spreading the viruses. India COVID-19 containment strategy has aligned with WHO’s strategic preparedness and response plan for COVID ’19. So the lesson learnt during this Covid19 is caution should be the keyword during the days to come.

Reference:

[1] A Review of Advances in Extreme Learning Machine Techniques and Its Applications Oyekale Abel Alade Faculty of Computing, Universiti Teknologi Malaysia,Skudai, Johor, Malaysia kafeabel@gmail.com, (aselamat,roselina)@utm.my2 Computer Science Department, Federal Polytechnic, Bida, Nigeria 2018.

[2] UCCSF: From Coronavirus Infected World to Coronavirus Immune World Authors M Indrabalan & Soumya Awasthi Research Report and Strategy Paper published by VIVEKANANDA INTERNATIONAL FOUNDATION Website : www.vifindia.org 2019.

[3] A Comprehensive Analysis of COVID-19 Outbreak situation in India- medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20058477;this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

[4] Impact of COVID-19 pandemic on health system & Sustainable Development Goal 3 Indian J Med Res 151, May 2020, pp 395-399 DOI: 10.4103/ijmr.IJMR_1920_20

[5] A comparative analysis on linear regression and support vector regression Kavitha,varuna,ramya November 2016 DOI: 10.1109/GET.2016.7916627 Conference: 2016 Online International Conference on Green Engineering and Technologies (IC-GET). 

[6] Extreme Learning Machine: A Review Article in International Journal of Applied Engineering Research - January 2018.

[7] A Review of Advances in Extreme Learning Machine Techniques and Its Applications May 2018.

[8] Trends in extreme learning machines: A review- Elsevier, Neural Network, 2014.

[9] Y. Chen et al., ”Extreme learning machine-based device displacement free activity recognition model”, Soft Comput., vol. 16, no. 9, pp. 1617-1625, 2012.
[10] Extreme learning machine: A new learning scheme of feedforward neural networks-Neural Networks, 2004.
Proceedings. 2004 IEEE International Joint Conference on, Volume:

[11] Lessons learnt during the first 100 days of COVID-19 pandemic in India Indian J Med Res 151, May 2020, pp 387-391 DOI: 10.4103/ijmr.IJMR_1925_20

[12] Statistics and Machine Learning in Python Edouard Duchesnay, Tommy Löfstedt, Feki Younes Nov 13 2019.

[13] Laboratory surveillance for SARS-CoV-2 in India: Performance of testing & descriptive epidemiology of detected COVID-19, January 22 - April 30, 2020 Indian J Med Res 151, May 2020, pp 424-437 DOI: 10.4103/ijmr.IJMR_1896_20.

[14] Gupta, R., Pandey, G., Chaudhary, P., & Pal, S. K. (2020). SEIR and Regression Model based COVID-19 outbreak predictions in India. medRxiv. Accessed from https://www.medrxiv.org/content/10.1101/2020.04.01.20049825v1 on 5th April 2020

[15] A Survey of Scholarly Data Visualization: JIAYING LIU1, TAO TANG2, WEI WANG1
Received December 18, 2017, accepted February 22, 2018, date of publication March 12, 2018, date of current version April 23, 2018.