Covid-19 Prediction in India using Machine Learning

Sarfraj Alam[1], Vipul Kumar[1], Sweta Singh[1], Sweta Joshi[1], Madhu Kirola[2]

[1]Student, [2]Assistant Professor,

[1,2]Computer Science and Engineering, Uttaranchal Institute of Technology, Dehradun, Uttarakhand, India

ABSTRACT

Various computational models are used around the world to predict the number of infected individuals and the death rate of the COVID-19 outbreak [3]. Machine learning-based models are important to take proper actions. Due to the ample of uncertainty and crucial data, the aerodynamic models have been challenged regarding higher accuracy for long-term prediction of this disease [1]. By researching the COVID19 problem, it is observed that lockdown and isolation are important techniques for preventing the spread of COVID-19 [2]. In India, public health and the economical condition are impacted by COVID-19, our goal is to visualize the spread of this disease [5]. Machine Learning Algorithms are used in various applications for detecting adverse risk factors. Three ML algorithms we are using that is Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest Classifier (RFC). These machine learning models are predicting the total number of recovered patients as per the date of each state in India [8].

KEYWORDS: *Data Analysis and Visualization, Logistic Regression, Support Vector Machine, Random Forest Classifier*

INTRODUCTION:

Various drastic respiratory syndrome coronavirus 2, also called SARS-CoV-2, is reported as a strain of virus inducing the respiratory disease of COVID-19 The coronavirus disease is confirmed to be extremely contagious by the global nations and the World Health Organization (WHO). The COVID-19 pandemic has become an international concern that has recognized as a public health emergency all over the world. Epidemiological models have been broadly used by media and officials to assess the outbreak, recognize the peak beforehand, and also anticipate the mortality rate. Outburst prediction models have manifested to be elemental to furnish observations into the harm brought by COVID-19 [3]. Moreover, the prediction models are taken as a scientific report to construct new plans and to assess the conditions of curfew. The COVID-19 pandemic has been broadcasted to be exceedingly aggressive to spread [1]. Towards the end of 2019, the new coronavirus spread widely in China, and a huge mass get infected. At present, the household upsurge has been productively managed, although the new coronavirus is escalating speedily in other areas. Presently, Europe has become the base of the ongoing outburst of new pneumonia. On March 11, the World Health Organization declared new pneumonia outburst a "global pandemic" [5]. This disease has provoked huge impact to the health and economical condition of the people. The research on the household and international endemic and the upcoming evolution shift has been subject undergoing intense study [6]. While no model can precisely foresee the rates of infection and mortality, there have been efforts made to contemplate and inspect the strengths and drawbacks of severe studies and models conferred regarding the coronavirus.

Background:

We investigate the matter of modeling the latest coronavirus (Covid19) spread the constrained or practical scenarios in India. It is monitored that lock-down and isolation are the crucial steps to control the spread of this disease. We study the result of those prevention techniques mathematically and build a model to predict the new cases or total infected cases. This prediction is required to organize reports and to take further steps to stop preventing this disease. [2]. COVID-19 has been proved a global pandemic and a couple of tests are using various ML based models to manage the reports of the various cases at global level. Here, we showcase our model that would be able to predict the arrival of new cases across India. Various classification methods are used for making an crucial report on COVID-19 across India [4].

Machine learning-based models have showcase the importance to predict the perioperative output to enhance the decision making on the longer-term period. Various prediction techniques are used to handle speculate problems. The study of COVID-19 shows the potential of ML based models to predict the number of upcoming patients suffering from COVID-19 [1].

Computational techniques for machine learning-based models have shown their importance for monitoring the potential threats for valueable decisions. Machine learning algorithms utilized in various applications show the inauspicious risk factors. The study shows the ability of model to estimate the number of individuals who are affected by this disease. This machine learning based model has different kinds of outcomes like the number of newly infected cases, death rates, and the recovered cases estimates in the next 10 days across India [8].

Proposed Methodology:

We extracted Indian country data about the COVID-19 disease from the Kaggle website Data was extracted on 4 September 2020. Using Jupyter Notebook and Python libraries we merged and filtered the final analysis of the dataset. In this, we use Pandas, NumPy, matplpotlib, seaborn and, sklearn library. From the list of unique states, we get the names of the top five states having maximum confirmed cases, totally cured cases, and total new cases. Then we plot the facet grid for total confirmed cases concerning the state. We plot one feature map which helps us to do the feature selection. And based on feature selection, we divide the dataset into training and test data.
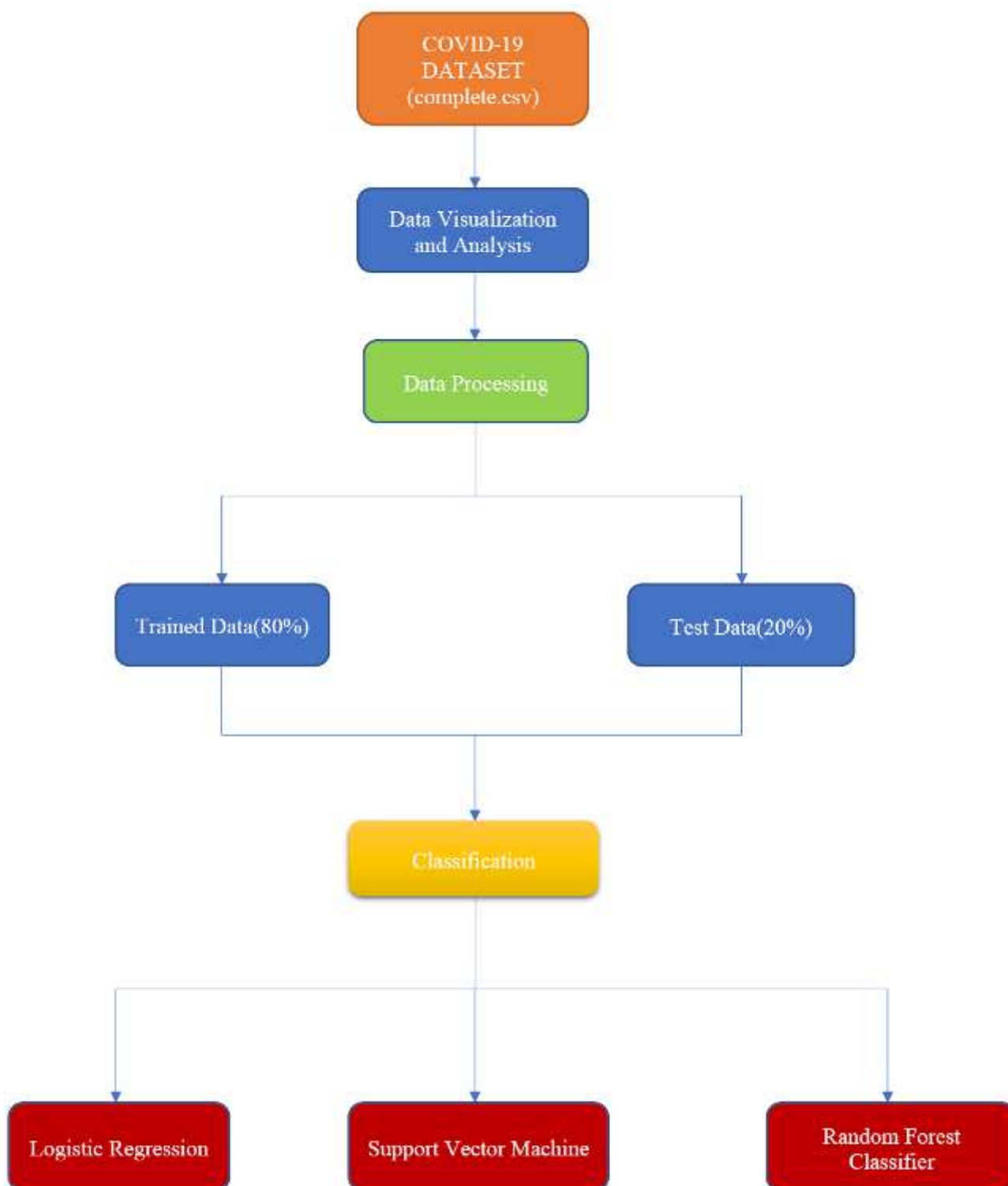
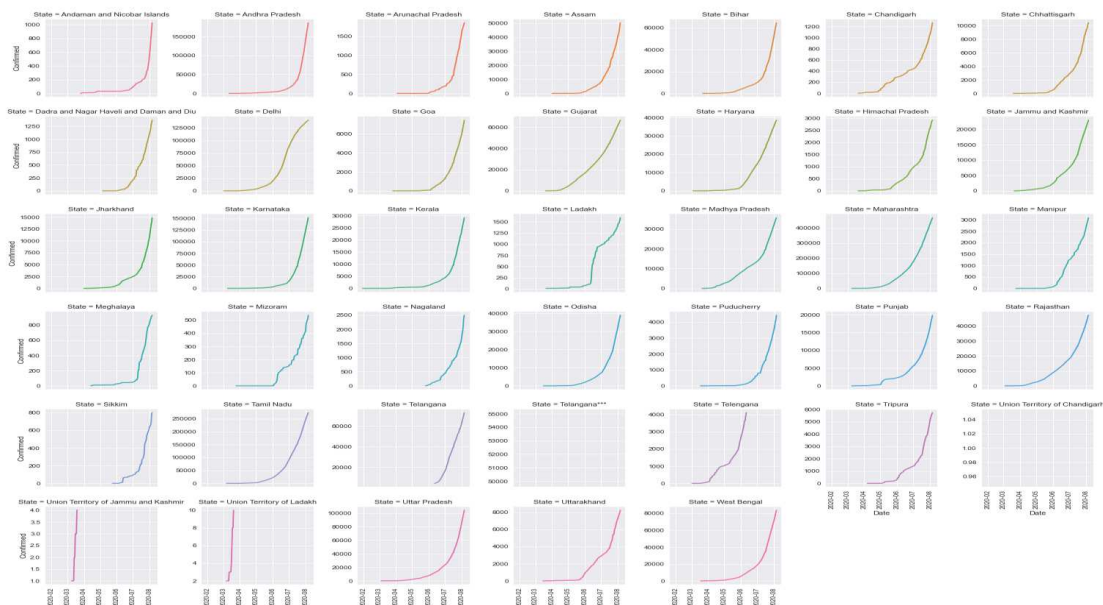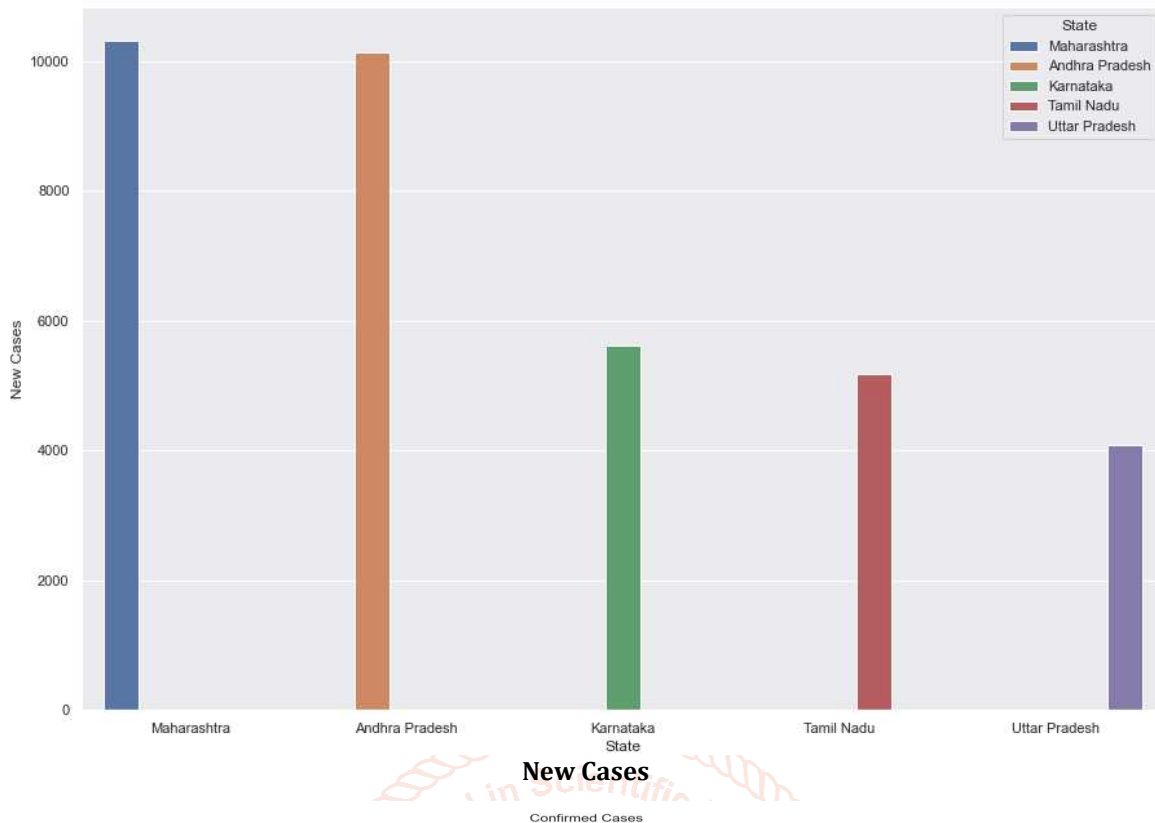Here are some visualizations for total confirmed cases, cured cases, new cases, and facetgrid for each state (Confirmed Case) –
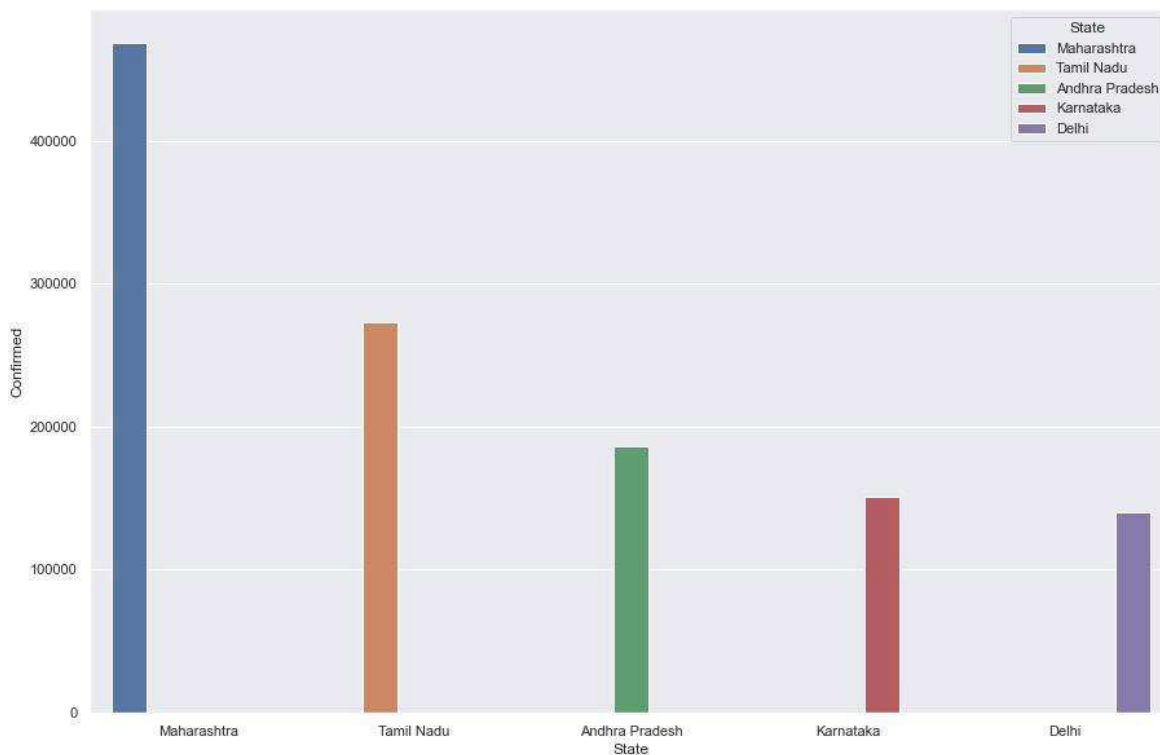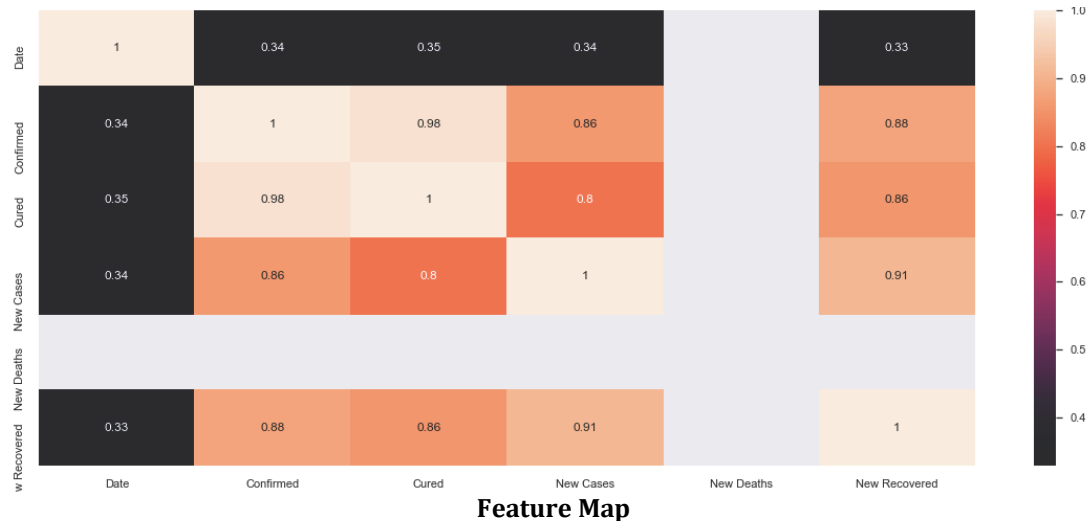

Confirmed Cases
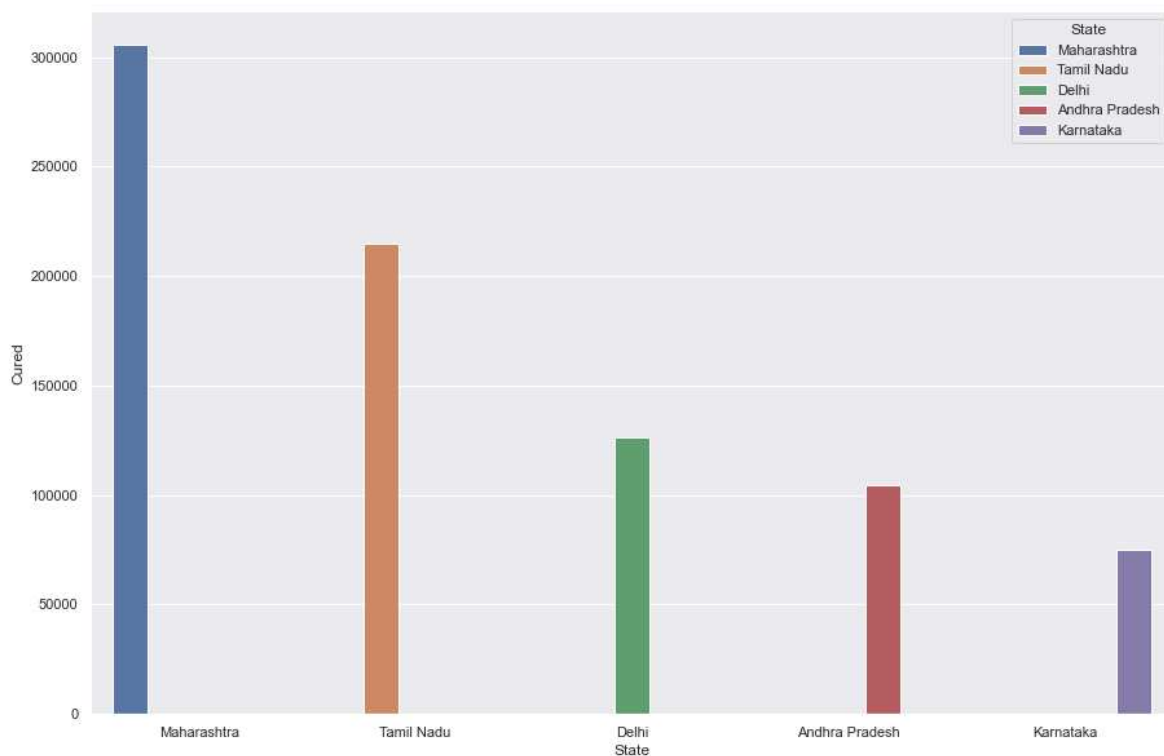
Cured Cases

New Cases

Facetgrid of Each State (Confirmed Cases)

Feature Map

We use three machine learning algorithms to build the model:

Logistic Regression: It is an classification based algorithm which is used to predict the unique values. It predict the output in the form of binary variable Basically it predicts whether the particular event is going to happen or not. Its output falls in then range between 0 and 1 [6].

Support Vector Machine: It is an classification based method in which we plot item as an point in N-dimensional world (N denotes total number of features the dataset contains). Each point has two co-ordinates that is X and Y and this co-ordinates are called as support vectors [8].

Random Forest Classifier: The random forest classifier is a powerful algorithm for disease prediction and it is a collection of decision trees. It takes the votes from all the decision trees depending on the parameters we give and it combines all the votes of decision trees to produce a single output. These decision trees contains two different stages that is ensemble learning and feature engineering. First, it maps from the training data that is uniquely selected, which ensures that each tree predicts some values using machine learing protocols. After this, it collects all the predicted values to produce the best unique output on the basis of votes [7].

Result Analysis: In this study, we build a computational model that will predict the total number of recovered patients that are suffering from COVID-19. The dataset we used contains the record about total confirmed cases, total deaths, cured patients, new cases, and total recovered patients per state wise. Three machine learning algorithms are used to predict the number of recovered patients.

S.NO	Algorithm	Accuracy
1.	Logistic Regression	36.74%
2.	Support Vector Machine	36.74%
3.	Random Forest Classifier	36.10%

Conclusion: We have looked into the issue of Covid19 spread in India in a big picture. A mathematical model has been accepted, which accompanies the actual data course of Covid19 spread in India. As we do lots of data visualization to see in the particular region how many cases are arising day by day and after that making use of the machine learning algorithms we will train our model in such a way that it will predict up to maximum accuracy it can do.

References:

[1] FURQAN RUSTAM 1, AIJAZ AHMAD RESHI 2, (Member, IEEE), ARIF MEHMOOD 3, SALEEM ULLAH 1, BYUNGWON ON4, WAQAR ASLAM 3, (Member, IEEE), AND GYU SANG CHOI - COVID-19 Future Forecasting Using Supervised Machine Learning Models – IEEE Access - Accepted May 13, 2020.

[2] Arti M. K., Senior Member, IEEE and Kushagra Bhatnagar - Modeling and Predictions for COVID 19 Spread in India – ResearchGate – April 2, 2020.

[3] Sina F. Ardabili 1, Amir Mosavi 2,3,*, Pedram Ghamisi 4, Filip Ferdinand 2, Annamaria R. Varkonyi-Koczy 2, Uwe Reuter 3, Timon Rabczuk 5, Peter M. Atkinson 6 - COVID-19 Outbreak Prediction with Machine Learning.

[4] R. Sujath1 • Jyotir Moy Chatterjee2 • Aboul Ella Hassanien3 - A machine learning forecasting model for COVID-19 pandemic in India - SpringerVerlag GmbH Germany - 30 May 2020.

[5] Gina Cahill MPH a, Carleigh Kutac MPH b, Nicholas L. Rider DO - Visualizing and assessing US county-level COVID19 vulnerability - American Journal of Infection Control.

[6] Bernhard O. Josephusa, Ardianto H. Nawirb, Evelyn Wijayac, Jurike V. Moniagad *, Margaretha Ohyvere - Predict Mortality in Patients Infected with COVID-19 Virus Based on Observed Characteristics of the Patient using Logistic Regression – Procedia.

[7] Mainak Adhikari a,*, Ambigavathi Munusamy b - iCovidCare: Intelligent health monitoring framework for COVID-19 using ensemble random forest in edge networks - Internet of Things - Accepted 24 February 2021.

[8] Ramesh Kumar Mojjada a,⇑, Arvind Yadav a, A.V. Prabhu b, Yuvaraj Natarajan c - Machine learning models for covid-19 future forecasting - Materials Today: Proceedings - Accepted 24 October 2020.

[9] (https://www.kaggle.com/imdevsk p/covid19-corona-virus-indiadataset?select=complete.csv)

[10] https://www.geeksforgeeks.org/python-pandas-timestamp-toordinal/

[11] Kumar, R., Singh, G., Joshi, K. Emotion Recognition System using Local Binary Pattern. International Journal of Inventive Engineering and Sciences (IJIES)-Volume, 5.

[12] Joshi, K., Joshi, K, N., Diwakar, M., Tripathi, N, A., Gupta, H. Multi-Focus Image Fusion using Non-Local Mean Filtering and Stationary Wavelet Transform. International Journal of Innovative Technology and Exploring Engineering (IJITEE)-Volume,9, 344-350.

[13] Joshi, Kapil and Kirola, Madhu and Chaudhary, Sumit and Diwakar, Manoj and Joshi, N.K., Multi-Focus Image Fusion Using Discrete Wavelet Transform Method (March 14, 2019). International Conference on Advances in Engineering Science Management & Technology (ICAESMT) - 2019, Uttaranchal University, Dehradun, India. Available at SSRN: https://ssrn.com/abstract=3383141 or http://dx.doi.org/10.2139/ssrn.3383141

[14] Gupta, Himanshu and Chauhan, Himanshu and Bijalwan, Anchit and Joshi, Kapil, A Review on Image Denoising (March 14, 2019). International Conference on Advances in Engineering Science Management & Technology (ICAESMT) - 2019, Uttaranchal University, Dehradun, India. Available at SSRN: https://ssrn.com/abstract=3385374 or http://dx.doi.org/10.2139/ssrn.3385374

[15] Gupta, Asheesh and Ahluwalia, Harshit and Adhikari, Preeti and Chaudhary, Sumit and Joshi, Kapil, A New Approach To Transfer Data Using RedTacton Technology (March 15, 2019). International Conference on Advances in Engineering Science Management & Technology (ICAESMT) - 2019, Uttaranchal University, Dehradun, India. Available at SSRN: https://ssrn.com/abstract=3395692 or http://dx.doi.org/10.2139/ssrn.3395692