Covid-19 Trend Analysis and Forecasting

Major Project II

Submitted By

Gaurav Parihar(9917103015)

Shubham Tak(9917103031)

Under the Supervision of

Dr. Arti Jain

Department of CSE/IT

Jaypee Institute Of Information Technology University, Noida

May 2021

TABLE OF CONTENTS

DECLARATION

I/We hereby declare that this submission is my/our own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place: Morena
Date: 08/05/2021

Signature:
Name: Gaurav Parihar
Enrollment No: 9917103015

Place: Jodhpur
Date: 08/05/2021

Signature:
Name: Shubham Tak
Enrollment No: 9917103031

CERTIFICATE

This is to certify that the work titled "**Covid-19 Trend Analysis and Forecasting**" submitted by **Gaurav Parihar** And **Shubham Tak** in partial fulfillment for the award of degree of B. Tech of Jaypee Institute of Information Technology, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Signature of Supervisor ……

Name of Supervisor Dr. Arti Jain

Designation Assistant Professor

Date 08 May 2021

ACKNOWLEDGEMENT

We would highly like to place on record our deep sense of gratitude to Dr. Arti Jain, Assistant Professor at, Jaypee Institute of Information Technology, Noida for his generous guidance and constant supervision as well as for providing necessary information regarding the project and also his support in completing this work. I express my sincere gratitude to faculty, Dept. of Computer Science, and project evaluators for their stimulating guidance, continuous encouragement and supervision throughout the course of present work. I also extend my thanks to my college for providing adequate resources for the project, my group members and seniors for their insightful comments and constructive suggestions to improve the quality of this project work.

Signature(s) of Students

Gaurav Parihar (9917103015)

Shubham Tak (9917103031)

Signature(s) of Supervisor:

Dr. Arti Jain
Assistant Professor
(Department of CSE & IT,JIIT)

ABSTRACT

The novel coronavirus (COVID-19), known as SARS-CoV-2 previously known as 2019-nCoV causes an illness which affects the respiratory system mainly. It has caused a huge impact on the lives of human lives, health and global economy. The critical problem the world is facing apart from the invention of a vaccine is to detect the presence of coronavirus in the human body in the early stage and prevent the situation from getting worse. To detect the presence of coronavirus the need of tools for diagnosis increased at the fast pace. Recent studies and radiologists specify that the physical examination of the body shows the chest radiographs were found abnormal. Studies also state that chest X-Rays are insensitive in the early stages of the disease. Further if quarantined the patient's X-Rays often reveals changes in the lungs. So, the above findings could be a green signal to make advancement in the field of computer vision, machine learning and deep learning as X-Ray images could help to detect the presence of Coronavirus. Computer Vision is a scientific field which completely focuses on how computers gain a high level of understanding from digitally created images and videos. Tabular data also plays an important role in making people aware of how big is the number that really matters when a slight increase in the number can cause a huge effect on social life.

LIST OF FIGURES

LIST OF TABLES

Chapter-1: Introduction

1.1 General Introduction

In this epidemic of COVID-19, people will continue to suffer until people receive the vaccine. A lot of scientific research is being done continuously. So, until people receive vaccines they should be tested at a correct time so that they can be cured and prevent the spread of coronavirus from that person. Test analysis is often performed in automated, high-throughput, medical laboratories by medical laboratory scientists. Test samples can be obtained by various methods, including a nasopharyngeal swab, sputum (coughed up material), throat swabs, deep airway material collected via suction catheter or saliva.

Reverse transcription polymerase chain reaction (RT-PCR) first uses reverse transcription to obtain DNA, followed by PCR to amplify that DNA, creating enough to be analyzed. RT-PCR can thereby detect SARS-CoV-2, which contains only RNA. The RT-PCR process generally requires a few hours. This complete process from collection, testing and making of reports takes about a day. The situation is even worse at places that are not equipped with testing laboratories and the biggest problem is contamination of the equipment with which samples are taken.

After going through many articles the world demands a fast, reliable and secure way of testing. Here comes the role of Computer Vision, Machine Learning and especially Deep Learning as it is described as the state-of-the-art for all analysis of medical imaging and plays an important role. Recent studies and radiologists specify that the physical examination of the body shows the chest radiographs were found abnormal as it infects the respiratory system mainly. Studies also state that chest X-Rays are insensitive in the early stages of the disease. Further if quarantined the patient's X-Rays often reveals changes in the lungs. So, the above findings could be a green signal to make advancement in the field of computer vision, machine learning and deep learning as X-Ray images could help to detect the presence of Coronavirus.

Recent advancement in Machine Learning and Deep Learning give us the leverage to analyze the trend the data follows. So, we gathered the data from the Indian Covid19 Website and merged the different period data files.

We have built a complete interface for the user to detect and see the trend the data follows collected around India. We are fetching live data and doing the analysis like Number of Cases/States, Number

of Cases/ Gender, Number of Cases/ Age,etc. We have predicted the Case Load Rate for next 2 months so as to make people aware and they can take preventive measures.

1.2 Problem Statement

The analysis and prediction of COVID-19 may be vital for formulating communicable disease control and prevention strategies. The prediction of infectious diseases is a crucial task in health work, which can detect the event trend of the disease early and increase the predictability of epidemic prevention work. The same is applicable with COVID-19 it is completely based on people to people and some say it's in the air too. So, to help people to bring insight from the data and graphical data we have built a web dashboard to take preventive measures.

1.3 Significance/Novelty of the problem

The standard forecasting techniques are forecasting the number of cases will increase in future. Which in turn not only increases fear in the people as the recovery rate and deceased that is responsible for forecasting data. The forecasting methods are the requirement of the current situation we are in , so that we are ready with the health infrastructure and other other requirements.

By bringing up solutions, which can automate the process is to forecast values much accurately and can rely upon upto some extent.

1.4 Empirical Study

The analysis and prediction of COVID-19 may be vital for formulating communicable disease control and prevention strategies. The prediction of infectious diseases is a crucial task in health work, which can detect the event trend of the disease early and increase the predictability of epidemic prevention work. Most algorithms use an integrated decision-making system using mathematical and mechanical calculation methods to predict future growth based on past event data. Classical methods like Arima Model, Facebook Prophet forecasting model and polynomial regression outperform machine learning methods for one step forecasting on the dataset.

1.5 Brief Description of the Solution Approach

In our approach, we use a Long Short Term Memory algorithm to forecast Case Load Rate and feed it with augmented data. We are using multiple layers to provide accurate results. So, it takes the best from all rather than be biased towards one approach.

Moreover, we have built a simple UI interface for end users to quickly interact with our model at no cost.

1.6 Comparison of existing approaches to the problem framed

LSTM variants like deep LSTM, convolutional LSTM and bi-directional LSTM models are tested on 32 states/union territories and support absolute error, the model with maximum accuracy is chosen based on prediction errors, bi-directional LSTM gives the simplest results, and convolutional LSTM gives the worst. this paper proposes a system for detecting the COVID-19 based on speech and sound analysis of the various extracted acoustic features using RNN, specifically the LSTM architecture. The existing solutions were accurate till the 1st wave. But, this 2nd wave have changed the trend completely and forecasting result were found inaccurate.

CHAPTER 3: LITERATURE SURVEY

Title: An Improved Method for the Fitting and Prediction of the Number of COVID-19 Confirmed Cases Based on LSTM[1]

Author: Bingjie Yan, Xiangyan Tang, Boyi Liu, Jun Wang, Yize Zhou, Guopeng Zheng

Identification :

The analysis and prediction of COVID-19 may be a vital prerequisite for formulating communicable disease prevention and control strategies. The prediction of infectious diseases is a crucial task in health work, which can detect the event trend of the disease early and increase the predictability of epidemic prevention work. Long Short Term Memory networks are a special quite Recurrent Neural Networks. it's a far better effect on statistical prediction. There's a time period for the new coronavirus. Using LSTM for statistical prediction may find the influencing factors of potential cases.

We use the data on the amount of daily diagnoses as different countries and therefore the time when the country takes similar "the curfew" measures because of the input of the model data. In view of the large data within the training set, to scale back the fitting deviation, we improved the LSTM model, used the unified variance because the judgment standard, reduced the data using the tactic, supported the general cardinality, and adjusted the data. They proposed improvement methods are mainly in the following aspects:

- We have fully considered the spatiotemporal characteristics of the info , instead of single standardized data.
- The improved parameter settings and evaluation indicators are more accurate for fitting and forecasting.
- We consider the impact of the epidemic stage and conduct reasonable data processing for different stages.
- Training description-
- Time Series Length=21 Days(Considered it as incubation period).
- Forecast Days=7 Days
- The outcome shows that LSTM has better results than digital prediction models(HIll Equation and Logistic)

Title: Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India[2]

Autor: Parul Arora, Himanshu Kumar Bijaya, Ketan Panigrahi

Identification :

Recurrent neural networks contain hidden states distributed across time, and this permits them to store plenty of data about the past. they're most commonly utilized in forecasting applications thanks to their ability to process variable length sequential data.Recurrent neural network (RNN) based long short term memory (LSTM) cells are used as prediction models.For prediction tasks,LSTMs are considered to be among the foremost feasible solutions, which they anticipate the long run forecasts enthusiastic to varied highlighted features present within the dataset.The data-set utilized during this paper is taken from the Ministry of Health and Family Welfare (Government of India) . This data is extremely stochastic in nature as increase/decrease in number of cases depend on other environmental/physical variables.

LSTM variants like deep LSTM, convolutional LSTM and bi-directional LSTM models are tested on 32 states/union territories and support absolute error, the model with maximum accuracy is chosen supported prediction errors, bi-directional LSTM gives the only results, and convolutional LSTM gives the worst.The error of these models is evaluated on testing data-set and supported the errors, the only model with maximum accuracy is chosen because the prediction model for COVID-19 data.

The output of our proposed model can help planners and authorities to form a choice on lockdown measures.

Outcomes-

- Comparison of Mean Absolute Percentage Error- Stacked LSTM-4.81%,Bi-LSTM-3.22% and Conv-LSTM-5.05%.
- Bi-directional LSTM performs better.
- Bi-directional LSTM has limited error range and is preferred for prediction purposes.
- Conv-LSTM gives the worst result.

Title: COVID-19 growth prediction using multivariate long short term memory[3]

Autor: Novanto Yudistira

Identification :

The spread of Coronavirus (covid-19) is an important function of tracking the growth of the epidemic. The LSTM layer structure is statistically searched until excellent points are obtained. First, we trained training data containing validated cases from around the world. We achieve better performance compared to the RNN method with a relatively low error verification error. Tests are based on graph visibility and RMSE.We found that it is difficult to achieve an equal number of validated cases over time, however, LSTM is able to provide the same pattern between real and predictable.

We used Adam Adjuster with a reading rate of 0.001 and a multiplication number of 10000. We train and test 5 times to reduce bias due to random initiation. In this way, we can calculate size and interval from multiple curve predictions. We conducted a study of how a number of hidden countries affect RMSE.

This study confirms the ability of LSTM to predict the covid-19 growth curve given sufficient training data.

The results will be convincingly better if we add a lot of data at great prices.Our method is much better than the traditional mathematical method or the quality model because the model is trained to represent global data completely.

Outcome:-

- LSTM outperformed RNN. The results will be better if more variety of data with large quantities is added.
- LSTM captures the pattern of the dynamic growth of graphs with minimum RMSE compared to RNN.
- Future work is expected to increase training data by adding data or data augmentation strategy.

Title: Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks[4]

Autor: Hafiz Tayyab Rauf, M. Ikram Ullah Lali, Muhammad Attique Khan,

Identification :

It is well known that inputs and outputs do not depend on each other in standard neural networks, but in cases where a sentence name should be predicted, the previous key words and the previous words should be preserved. This approach was developed by RNN, which solved this problem through a hidden layer. The RNN is a "memory" that stores all the information about the compound.The recurrent neural network (RNN) is a well-known neural network in which the outcome of a previous step is taken as a step forward.

The proposed framework is strong enough in the environment as it reviews the forecast results for each day in each target country. Data used for the training and evaluation of in-depth study models is taken from the standard WHO website and is available and publicly available

To validate the predictions obtained using the proposed framework, actual cases of the past 10 days were taken with a verified number of COVID19 patients.

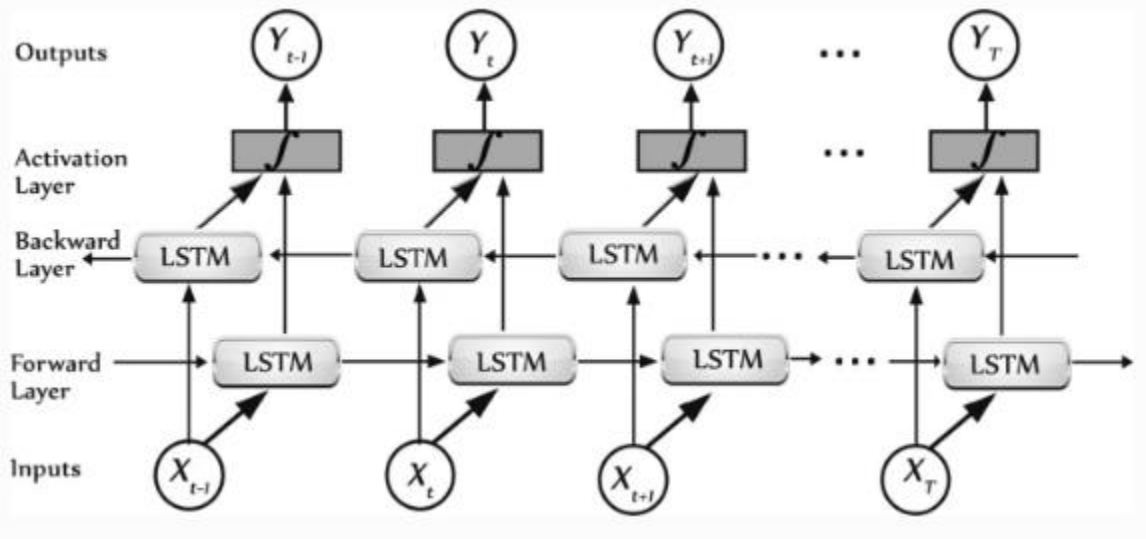


Fig 2.1 Graphical Representation of LSTM Architecture.

Title: Prediction of the Number of COVID-19 Confirmed Cases Based on K-Means-LSTM[5]

Author: Shashank Reddy Vadyala, Sai Nethra Betgeri, Eric A. Sherer, Amod Amritphale

Identification :

This paper presents a composite algorithm that combines the selection of similar dates based on the region using Xgboost, K-Means, and interim network (LSTM) neural networks to create a predictive predictor model for COVID-19 interim cases. SEIR-based models have been widely used to model COVID-19 outbreaks, containing some uncertainty. As part of the SEIR model, this study proposed an in-depth learning model (K-Means-LSTM) as a new practice to advance the outbreak model. The K-Means-LSTM strategy does not predict an epidemic of infection and spread. Instead, it predicts a time series of infected cases.K-MeansLSTM has three main advantages. First, the model is readable not only in the historical data of COVID19 but also in the data used for COVID-19, climate and risk for each region. The second advantage of the model is its ability to be used on various scales. At the moment it can predict the spread of the parish and the province. Finally, the model can predict short-term and long-term predictions that can be a reliable decision-making tool.

We compared our predictive K-means-LSTM results with the SEIR model results. To accurately assess the effectiveness of the SEIR, K-means-LSTM and LSTM models, Root means that the square error (RMSE) is calculated. Climate change is strongly associated with changes in mortality rates due to pneumonia.

Title: Neural Network Based Country Wise Risk Prediction of COVID-19[7]

Autor: Ratnabali Pal, Arif Ahmed Shekh, Samarjit Kar

Identification :-

Most algorithms use an integrated decision-making system using mathematical and mechanical calculation methods to predict future growth based on past event data.

In this paper, we have proposed the AI-guided approach to predicting the world-class long-term risk of coronavirus. The main challenges of this problem are: Small Database: Multidisciplinary ML (ML) required skills require a large volume of training data. Significantly, the COVID-19 database is less than a year old and it is difficult to design accurate AI training methods for such a small amount of data.

The country-specific network is used to predict the number of cases (κ), the number of deaths (δ), and the number of adoptions (K).) .The method can be useful to predict the long-term risk of an epidemic such as COVID-19. We found an average accuracy of 77.6% over all country-specific databases.

Title: Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM[8]

Autor: Farah Shahid, Aneela Zameer, Muhammad Muneeb

Identification :-

This paper aims to compare predictive models from statistics, machine learning and in-depth study of the COVID-19 database of ten countries including Brazil, Germany, Italy, Spain, UK, China, India, Israel, Russia, and the USA everywhere in the world. in the first phase, the development of ARIMA and SVR as a simple ML algorithm is discussed, and within the next phase, a description of the various models of DL Database has three components of verified cases, deaths and acquired cases. Unlocked data delays the merging process. MinMaxScaler removes a very small number of features and is previously separated by a range of features.

The performance of all models is compared with the concept of standard mathematical measurement measures in terms of absolute mean error (MAE), square root error (RMSE), and r2_score. ARIMA and SVR models cannot follow the trend of those features with high predictive error and bad r2_score values. Without measurement, LSTM produces low MAE rates of certified cases and deaths such as 2.0463 and 0.0095, while RMSE values of 2.2428 and 0.0103, respectively, in China with a good number of cases obtained by the union government. While among all countries, Bi-LSTM predicts

three of the simplest cases in China with high accuracy among all methods of all countries with low MAE and RMSE values of 0.0070 and 0.0077, respectively, deaths in China.

Title: Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model[9]

Autor: Smita Rath, Alakananda Tripathy, Alok Ranjan Tripathy

Identification :

The virus spread among the people more often through small droplets released by coughing, sneezing and talking in close contact. rather than moving long distance through air the droplet falls onto the bottom or surface. the essential symptoms of the COVID-19 are fever, cough, shortness of breath, loss of sense, and fatigue.

The whole data set was collected from the WHO site https://covid19india.org for the daily new positive cases, active cases, deceased and recovered cases during a CSV file. Initially, the info cleaning process is performed on the 2 data set to get rid of any missing values. Then a correlation analysis is performed on the info sets using Python programming through Spyder of Anaconda Navigator App.

The two data sets are first to separate into eighty percent training set and one-fifth testing set then rectilinear regression is performed to train the primary 80% set. Here the amount of daily active cases is predicted supported daily positive cases. rectilinear regression comprising various variables is known as linear multiple correlation . The steps for multiple rectilinear regression are nearly almost like those for easy rectilinear regression . the excellence lies with estimation. From the above training and testing of the prediction models, it had been found to be an efficient thanks to forecast subsequent number of daily active cases during the second week of August.

Title: Covid-19 Pandemic Data Analysis and Forecasting using Machine Learning Algorithms[10]

Autor: Sohini Sengupta, Sareeta Mugde, Dr Garima Sharma

Identification :

These machine learning and deep learning models don't deliver their promise of accurate statistic forecasting. The Machine Learning and deep learning models used above are traditionally used for time-series forecasting and provides brilliant results elsewhere for example sales forecasting. . Prophet is another forecasting model given by Facebook, which is supposed to manage business problems. it's two techniques of growth forecasting- here we are using the saturated growth model. Model is extremely flexible to handle periodic effects.Linear regression model trains itself on some historic data of independent and variable and predicts the long term considering a linear relationship between both. Polynomial Regression uses a similar approach; only difference is that the variable is modeled as nth degree of the polynomial in x.Support vector regression could also be a supervised machine learning model which pulls a hyperplane between the information points and creates a boundary of possible data points (high and low) in future. Support Vector Regression traditionally features an enormous forecasting ability.

We are currently at a growth phase when it involves the number of daily active cases; however considering the graphs of several other countries, in India also we expect some extent from which the curve will fall. this kind of scenario are often best analyzed and supported by a Sigmoid model. The Sigmoid Model has very cheap Root mean square value, so it is the simplest fit.

Also there are many other predictive models using machine learning and deep learning that are accustomed forecast. However the Sigmoid model only gives us an accurate representation of date once we will expect the peak .

Title: An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19[6]

Autor: Denis A. Pustokhin,Irina V. Pustokhina,Phuoc Nguyen Dinh,Son Van Phan

Identification :

The proposed approach is made up of two phases, consisting of pre-processing and an in-depth LSTM model. In the first phase, pre-processing techniques were used in CX images to increase partition performance.In the second step, MCWS was used to classify areas in gradient images. In the final step, these processed images were adapted to include an input layer of a deep LSTM model, which contained sequential data for creating a block and an LSTM network.

The proposed approach supported the in-depth LSTM model provided a high degree of success for the three-class classification comprising COVID-19, pneumonia and general.

Good performance was achieved with 80% training and a 20% testing rate. The proposed model could significantly improve existing radiology methods and could be of great help to radiologists and specialists to assist them in detecting, determining, and tracking COVID-19 cases throughout the epidemic.

Title: Time series forecasting of COVID-19 transmission in Canada using LSTM networks[16]

Author: Vinay Kumar Reddy Chimmula, Lei Zhang

Identification :

Every infectious disease outbreak reflects certain patterns and such patterns need to be identified based on the transmission capacity of such an outbreak. Interventions to eradicate these infectious diseases relied on methods used to detect outbreaks when they occurred. Any outbreak in a country or province usually occurs at different levels of magnitude over time, which means seasonal changes, viral adaptation over time. Often the patterns shown in such cases are not in harmony with the environment and this motivates us to create a system that can capture such a dynamic change. Database is available in the form of a time series and date, month and year so that temporary items may not be available.

The patterns from the data show that the faster and more effective measures taken by Canadian public health authorities to reduce exposure show a positive effect compared to other countries such as the USA and Italy.

Title: Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study[17]

Autor: Abdelhafid Zeroual, Fouzi Harrou, Abdelkader Dairi,

Identification :

Today, COVID-19 pandemic is one in every one of the first difficult issues catching our occasions because of its very adversely influences general wellbeing. enduring neural organizations Regular feedforward neural organizations territory unit wide utilized with achievement in different fields. In such organizations, information stream changes territory unit passed by means of covered up layers a single way any place the yield is affected exclusively with this case.The essential substance in RNNs is to accept the impact of past data for creating the yield. to the current completion, cells portrayed by entryways impacting the yield exploitation recorded perceptions territory unit encased to encourage the yield. Precisely articulation the quantity of affirmed and recuperated cases gives relevant information to governments and leaders with respect to the normal situation and in this way the necessary measures to force.

Title: COVID-19 Detection System Using Recurrent Neural Networks[18]

Autor: Abdelfatah Hassan, Ismail Shahin, Mohamed Bader Alsabek

Identification :-

This paper proposes a system for detecting the COVID-19 based on speech and sound analysis of the various extracted acoustic features using RNN, specifically the LSTM architecture. The RNN is employed mainly to predict the longer-term data sequence through the utilization of previous data samples.RNN suffers from the vanishing gradient problem, which rises with the length of the training sequences. Therefore, LSTM is employed to beat this problem. LSTM stores data information for an extended period, and it's easier to recollect the past data within the memory. During this work, the RNN architecture that we are using is LSTM. We used the extracted features from our collected database for network training and testing. Besides, 70% of the info was used for
training and 30% for testing.To evaluate our system, we use the performance metrics of precision, recall, F1-score, Area Under Curve (AUC), and accuracy.

This paper has provided a groundbreaking and modern approach for the first diagnosis of COVID-19. It also illustrates the mechanism of the proposed COVID-19 detection system. This analysis is completed by evaluating the various acoustic features of cough sound, breathing sound, and voice.

based on the obtained results, it are often stated that Patients' voice has shown such inconvenient accuracy compared to cough and breathing sounds.

Title: The Prediction of COVID-19 Using LSTM Algorithms[19]

Autor: Myung Hwa Kim, Ju Hyung Kim, Kyoungjin Lee

Identification :-

This study aims to create a prediction model as a more accurate decision support tool by using deep learning to predict economic indicators consistent with epidemic trends in statistic . so as to reduce anxiety and uncertainty about economic damage caused by the pandemic of latest infectious diseases, it's necessary to reasonably predict the economic impact by methodology and establish a response strategy the primary verification uses 20% of the collected influenza trend data to verify the economic indicator prediction model for influenza trends, and therefore the second verification is verified using COVID-19 actual data from January 1, 2020 to August 10, 2020 to verify the predictive performance of other communicable disease cases. This model was developed using LSTM algorithms that show high performance in statistic forecasts, and performance results showed good performance within the rate of inflation describing the worth index. n future studies, it's thought that a more suitable model for COVID-19 are often developed if the epidemic data is assessed and predicted by the eight influenza types classified by the CDC.

Title: Deep learning via LSTM models for COVID-19 infection forecasting in India[11]

Autor: Rohitash Chandra, Ayush Jain, Divyanshu Singh Chauhan

Identification :-

Coronavirus virus 2019 (COVID-19) is an infectious disease caused by acute acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has become a worldwide epidemic. In practice, the COVID19 Daily Model Prediction Model STM Network (LSTM) Neural Network Model, Bidirectional LSTM Network (BD-LSTM) and Encoder-Decoder JSTM Network (ED-LSTM) are provided with structural details in India. Utilizing the best LSTM model provides a two-month approach to the newest day-to-day affairs by returning expectations to the trained model. Adam

Optimizer was used to build the corresponding LSTM model with the corresponding squared error (MSE) loss function.

Demonstration of relevant methods in terms of scalability and visibility indicating the power to take care of continuous assessment performance as the assessment horizon increases. RMSE is used because it is the main performance measure for assessment accuracy. This model provides a visual representation of total COVID-19 infections in various states and Union Territories of India. The results show the challenges of estimating limited data, which take into account cases across India.

Title: Transfer Learning for COVID-19 cases and deaths forecast using LSTM network[12]

Autor: Yogesh Gautam

Identification :-

An effective modeling method to forecast the spread of the virus among the population are often extremely useful to organize and formulate health and economic policies for any government or administrators. When new cases rise at a rate of thousands per day then even the foremost developed nation's healthcare system has been overwhelmed to handle an outsized number of patients. The info has been obtained from an open dataset of Our World in Data which maintains and updates global data daily acquired from the european Center for Disease Prevention and Control (ECDC). the info was available from the 31st of December to the 10th of June. However, the info from February 28th onwards has been used for training purpose during which cases were more active. Data for the amount of latest cases per 100 thousand of the population per day is one among the forecasting variables. The paper implements Transfer Learning for the LSTM network to find out from one region and forecast for a totally different region. For training purpose, Keras API with TensorFlow backend is employed . Various python libraries are used for other evaluations. The networks were trained in GPU available within the Google Colab environment.This approach is more useful for countries that are in its early phase of virus spread, to forecast based on other country's models. Such forecasts are often an excellent help to governments and policymakers.

Title: Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method[13]

Autor: Peipei Wang, Xinqi Zheng, Gang Ai, et al

Identification :-

Breathing and close contact are the most channel of COVID-19 transmission, while controlling the source of infection, isolating the route of transmission and protection of vulnerable groups are the key to stop the spreading of COVID-19. As a result, many countries have adopted unprecedented nationwide interventions, like keep social distance, a curfew, closure of faculties and businesses, bans on travel et al. to stop the further spread of the epidemic. Moreover, traditional forecasting methods mostly use current data like historical confirmed cases to create models, while the real-time update of knowledge isn't considered. it's worth noting that the data were studied within the LSTM network through deep learning, and no change was caused to the infected database during the study. The above results provide that the epidemic of Peru will peak on early December, leading to 398,991 total infections. The model also predicted the epidemic size of Russia will reach 1,343,751 after 150 days, and therefore the accumulative confirmed cases in Iran are going to be greater than 470 thousand by early December consistent with the results of our model.It is a challenging task to forecast the spread of outbreaks supported small data sets, limited by the tiny data available, from which much data and law can't be deciphered if the initial data is nearly a flat curve.

Title: Deep Ensemble Learning Method to Forecast COVID-19 Outbreak[14]

Autor: Farah Shahid, Aneela Zameer, Muhammad Muneeb

Identification :-

This investigation centers around medical services examination, information investigation AI arrangements, to manage anticipating of COVID-19 flare-up. The objective of this paper is to propose a generic,data-free, COVID-19 episode prescient strategy upheld by profound learning procedures which will be utilized as a medical care choice help device for various nations whose preventive and control measures are similar. This technique embraces a stacking procedure to accomplish better exactness by melding the anticipated qualities and construct three learners (DNN, LSTM, and CNN) that will independently gauge COVID-19 plague patterns. To assess the expectation paces of the

proposed leaners and meta-students, the model utilized the root mean squared mistake (RMSE) and the air conditioner curacy. RMSE might be a typical estimation for the contrast among anticipated and genuine qualities. It is generally utilized in different fields just as in the forecast of irresistible sicknesses spreading and the precision recognizes the general viability of the models.

The model effectively built a nonexclusive COVID-19 flare-up pre-mandate technique that might be utilized as a choice help instrument for improving its observation, controlling, and overseeing pestilences. Notwithstanding, a constraint of this investigation is that we didn't consider during preparing different elements which are related to the spread and flare-up of COVID-19 and that can give more significant examinations and ideally more sensible expectations. These components incorporate, for example, legislative issues, culture, schooling, limiting outside exercises, wellbeing offices, implementation of wearing covers, geological position, and so forth.

Title: COVID-19 Patient Count Prediction Using LSTM[15]

Autor: Muhammad Iqbal, Feras Al-Obeidat , Fahad Maqbool , Saad Razzaq

Identification :-

COVID-19 is a zoonosis that begins in animals and can be passed to humans following recombination, mutation, and adaption. Although Covid-19 seems similar to influenza virus , the transmission speed is a critical contrast between the two infections. Influenza incorporates a shorter middle incubation period and a shorter serial interim than the Covid19 virus The LSTM model is trained on the data extracted from Covid-19 data source. The model was trained with varying LSTM hidden nodes, the number of epochs, and the batch size to find the best performing results.Rather than training the LSTM model using daily patient count (PC) and then predicting PC in upcoming days, we have used the Percentage of Positive Patients (PPP).

The proposed model can also predict Covid-19 cases, areawise, that help smart lockdown decisions and area wise sampling. In the future, we have plans to train the model on death cases and will try to forecast the mortality ratio and its correlations with the critical cases.

Title: COVID-19 prediction using LSTM algorithm: GCC case study[20]

Autor: Kareem Kamal A. Ghany, Hossam M. Zawbaa, Heba M. Sabri

Identification :-

The current outbreak of disease due to COVID is accounted for in late December 2019. This new infection is exceptionally infectious and has rapidly spread around the world. LSTM is a novel recurrent network architecture in combination with a proper gradient-based learning algorithm. LSTM is designed to overcome these error back-flow problems. It can learn to bridge time intervals over 1000 steps, even in the noisy case, compact input sequences, without the loss of short time delay capabilities.

They worked over Vanilla LSTM which consists of a cell, an input gate , a forget gate and an output gate. Periods and the 3 gates manage the progression of data related to the cell. After the implementation of Vanilla LSTM they proposed a LSTM model with an input layer, and a single layer, and an output layer.

They are able to predict which country will recover fast and which will take more time after looking at the increase in the number of cases. The obtained results may vary given the stochastic nature of the LSTM model; therefore, we have run it several times. Finally, we enter the last sequence with output to forecast the next value in the series.

Outcome:-

- The learning rate is set to 0.001 and it decays every five epochs.
- Number of epochs used=1000
- Optimizer used Adam, mean square error, as the loss function and Relu activation function.

2.2 Integrated Summary of the literature studied

- After going through the literature survey many models have been built to forecast the future trend of the data. But, most of the models are based on previous data which was basically dependent on the trend in the 1st wave.
- Some of the mostly used models are Vanilla LSTM, Stacked LSTM, Bi-directional LSTM and Conv-LSTM.
- Most of the studies are based on linear knowledge but, current data is based on non-linear insights.
- Models with least Mean Absolute Percentage Error are preferred.
- Worked on hyper-parameter tuning- Lag size, No.hidden layers, dropout rate, learning rate,batch size, and epoch size.
- Preferred incubation period 21 days.
- Prediction for less interval of time is preferred over long intervals of time.

CHAPTER 3: REQUIREMENT ANALYSIS AND SOLUTION APPROACH

3.1 Overall description of the project

The complete project is built keeping in mind the over need of the time and how we can help the higher authorities and people to take proper precautions and other preventive measures to help oneself or the population of the country. We build a complete dashboard using React which showcases the trend of the data collected from the Indian Covid-19 website.

We have built the forecasting model to predict the Case Load Rate of Covid patients in the country. The complete interface is user friendly and scalable and the trend described over the dashboard can be an alarming stage when the situation becomes worse and actions need to be taken over that.

3.2 Requirement Analysis

The requirement analysis was done in four phases:

- Eliciting requirements: We've gathered multiple research papers and found out that usually the existing solutions were accurate for 1st wave and the predictions were inaccurate for current situations.
- Analyzing requirements: We analysed every requirement and came up with an idea to provide a free and open source solution where users can watch out the trend and look for future forecasts.
- Requirements modeling: We've modelled multiple user cases and explained for better coverage.
- Review and retrospective: We've kept iterating above phases to make sure we keep our workflow flexible over time.

3.3 Solution Approach

Our approach is to build a simple UI where the end user can look out the trend graphs and look for future trends of Case Load Rate. The forecasting model for Case Load Rate is built over LSTM(Long Short Term Memory). In order to build models, we've utilised multiple models as layers to provide a generalised model for more accurate results.

CHAPTER 4: MODELING AND IMPLEMENTATION

4.1 Design Diagram

4.1.1 Use Case Diagrams

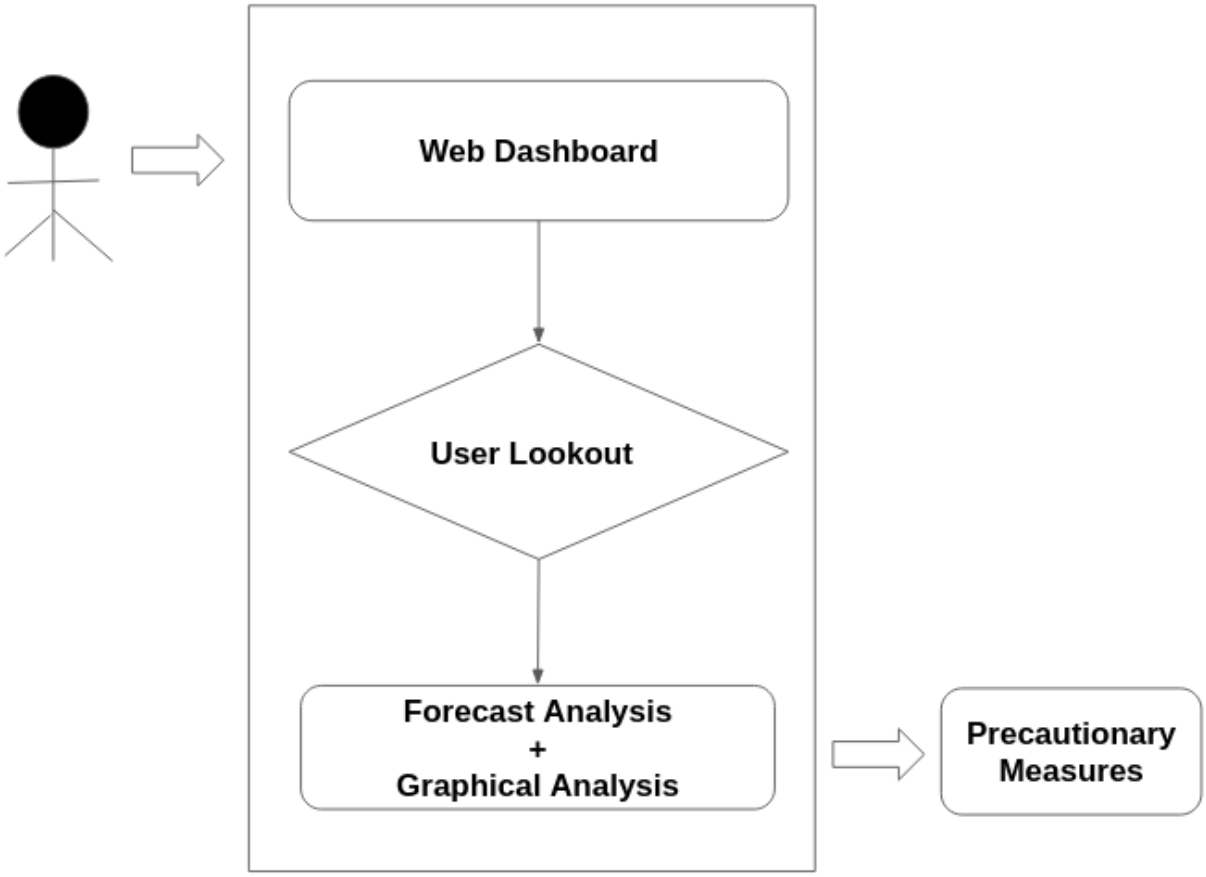


Fig 4.1 Use Case Diagram

4.1.2 Control Flow Diagram

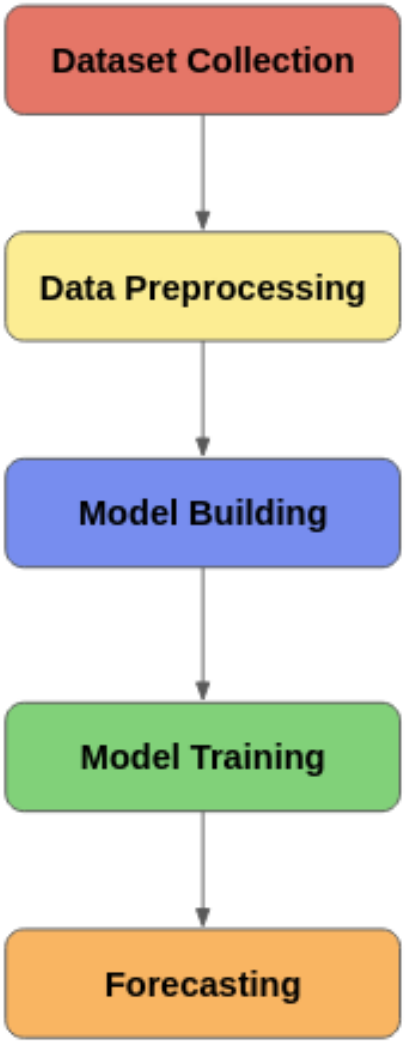
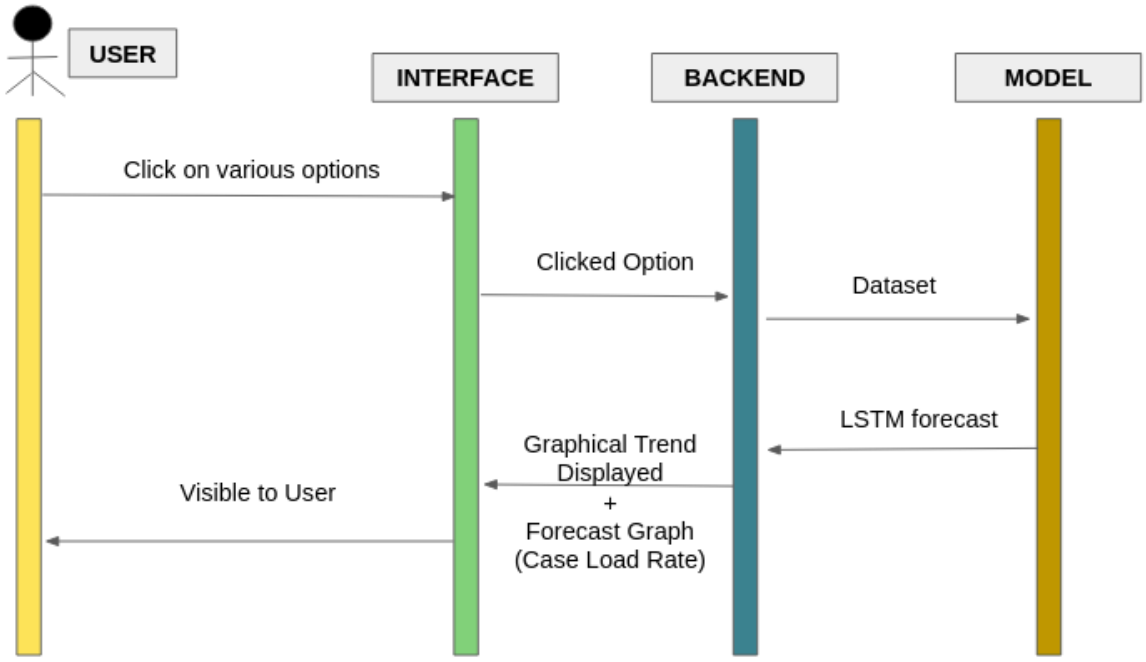


Fig 4.2 Control Flow Diagram

4.1.3 Sequence Diagram



4.3 Sequence Diagram

4.2 Implementation Details and Issues

We have utilized the COVID-19 Indian dataset for the LSTM model preparation and forecasting. The dataset consists of tabular details with following columns:

Column Name
Date
Date_YMD
Daily Confirmed
Total Confirmed

Daily Recovered
Total Recovered
Daily Deceased
Total Deceased

Table 4.1: Dataset Description

	Date	Date_YMD	Daily Confirmed	Total Confirmed	Daily Recovered	Total Recovered	Daily Deceased	Total Deceased
0	30 January 2020	2020-01-30	1	1	0	0	0	0
1	31 January 2020	2020-01-31	0	1	0	0	0	0
2	1 February 2020	2020-02-01	0	1	0	0	0	0
3	2 February 2020	2020-02-02	1	2	0	0	0	0
4	3 February 2020	2020-02-03	1	3	0	0	0	0

Fig 4.4: Dataset Example

4.2.1 Implementation Steps:

Step 1:

Fetch the dataset file from Covid19.org website through API calls. And followed the data preparation steps.

Step 2:

Displayed the data insights on the graph present React based dashboard.

Step 3:

Created the concept of Case Load Rate that is basically dependent over Daily Confirmed, Daily Recovered Cases, Daily Deceased.

Case Load=Daily Confirmed-Daily Recovered Cases-Daily Deceased

Case Load Rate= (Case Load/Daily Confirmed)*100

Step 4:

In the next step we calculated Case Load Rate and added to the dataframe as a column named as Case Load Rate.

Step 5:

Represented graphically the Recovery Rate, Case Load Rate, Delta.

Recovery Rate=(Daily Recovered/Confirmed Case)*100

Delta=Case Load Rate-Recovered Rate

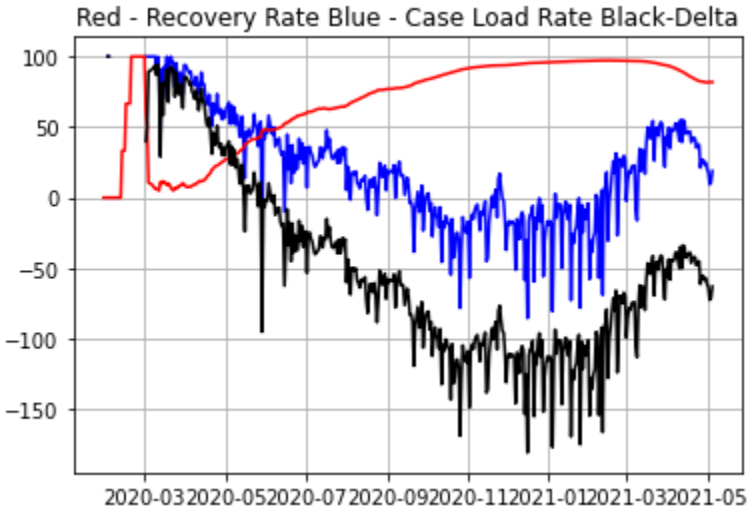


Fig 4.5: Graphical Representation

Step 6:

Worked over creation of LSTM model. We created a univariate LSTM model with one trainable parameter as Case Load Rate. The Sequential model has various different layers. The model is a sliding window as it is learning from last 14 days values. The reason behind taking 14 days is that the recovery period is 14 days and the patient starts to recover.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 14, 64)            16896
_________________________________________________________________
lstm_1 (LSTM)                (None, 32)                12416
_________________________________________________________________
dense (Dense)                (None, 1)                 33
=================================================================
Total params: 29,345
Trainable params: 29,345
Non-trainable params: 0
```

Fig 4.6: Model Summary

Step 7:

The next step involves the training of LSTM model. The data is then fit into the model with

Number of Epochs=90

Number of Steps per Epochs=1

Fig 4.7: Loss and Mean Error Graphs

Step 8:

To make the whole process of trend analysis and forecasting of Covid-19, we have designed a react based dashboard. The dashboard is user friendly and easy to use. The user just needs to choose some of the options and look out various different graphical insights and take precautionary measures. The images below give the description about the web dashboard.

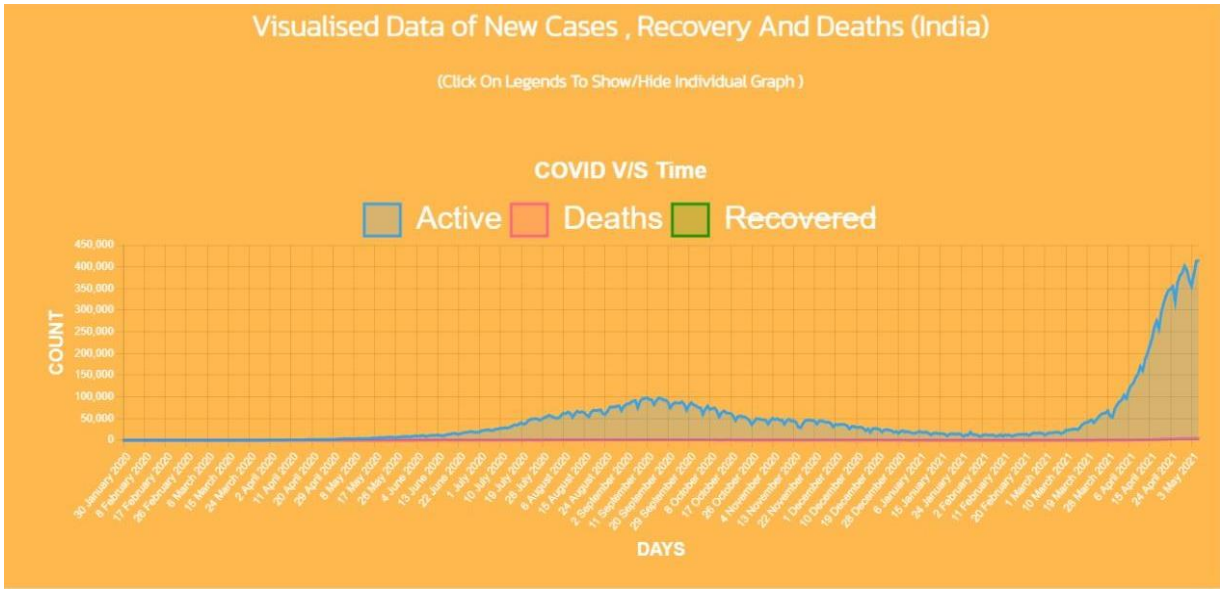


Fig 4.8: UI Showing The Graphical View(New Cases, Recovery, and Deaths)

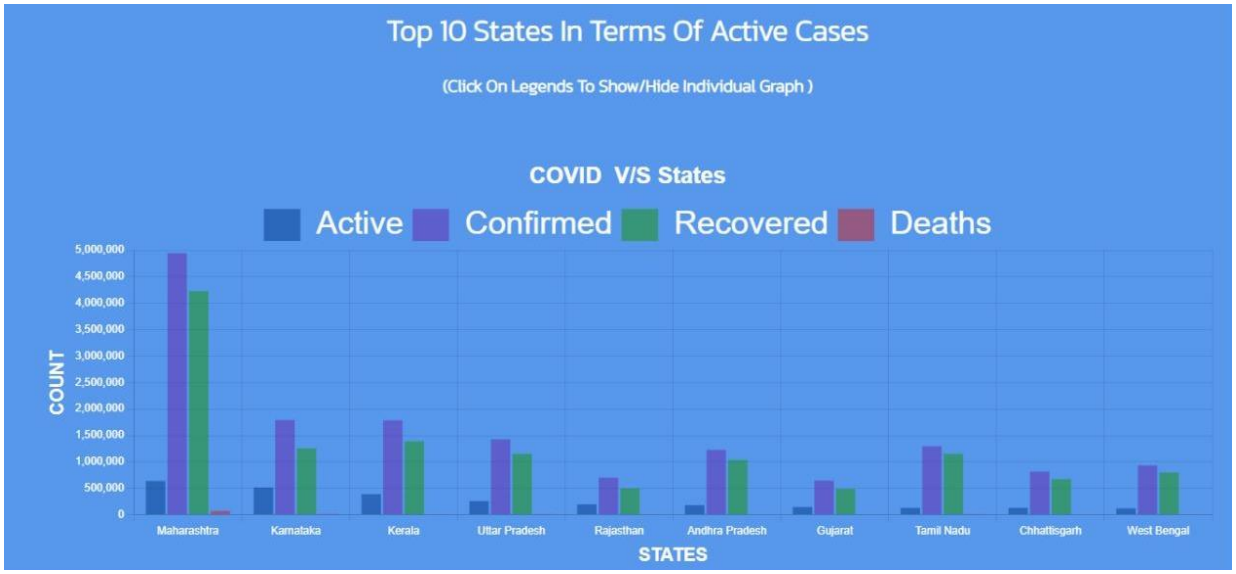


Fig 4.9: UI Showing The Graphical View(Top 10 States with Active Cases)

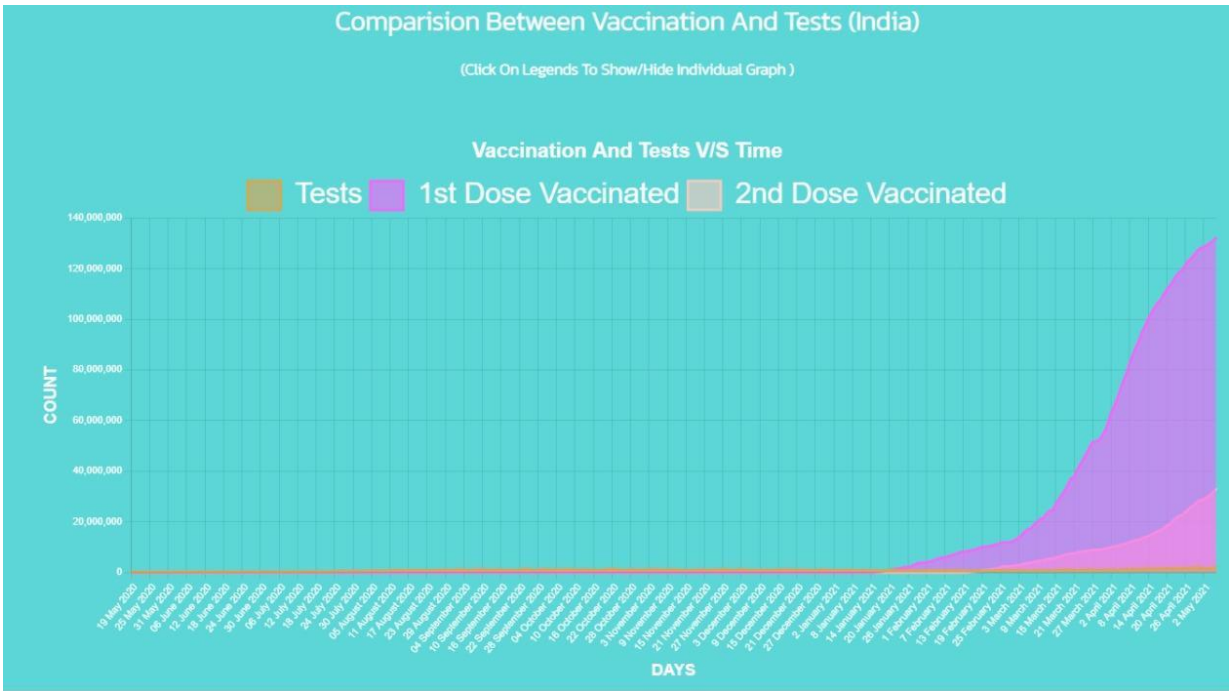


Fig 4.10: UI Showing The Graphical View(Tests+Vaccination)

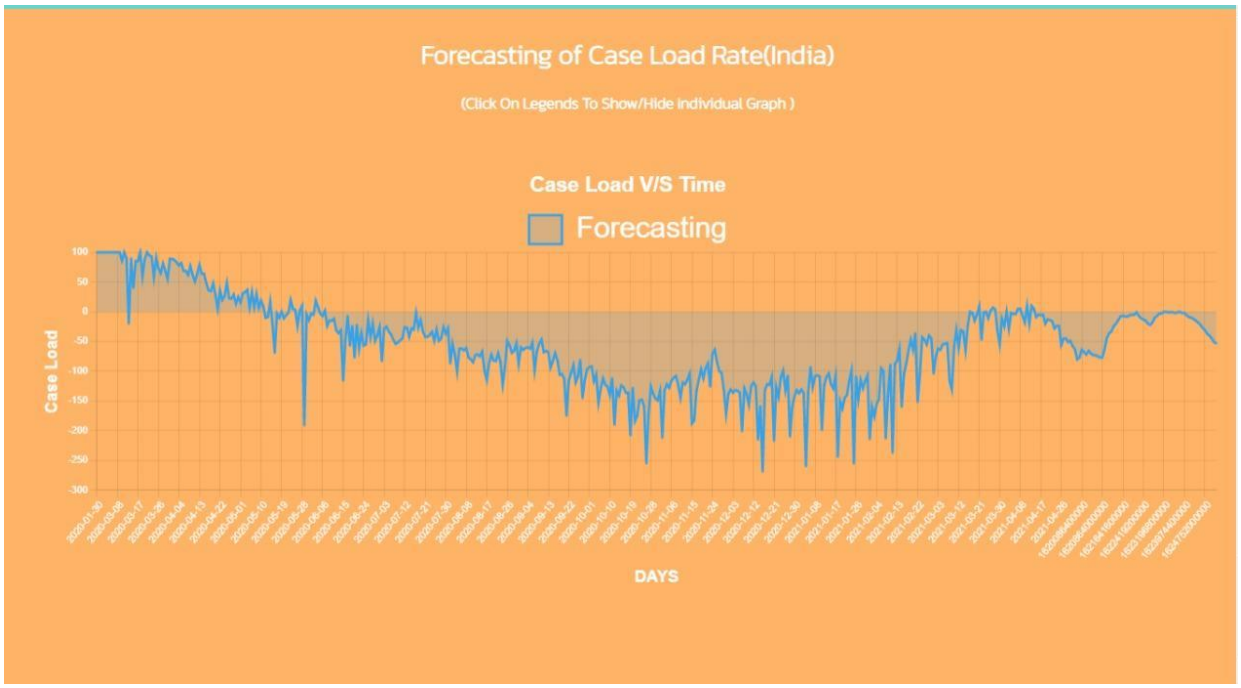


Fig 4.11 : UI Showing The Graphical View(Forecasting of Case Load Rate)

4.2.1 Working of Applied Model:

We have implemented the LSTM model keeping in mind that the forecasting can be accurate and does not give suggestions to completely rely upon. But, take precautionary measures.

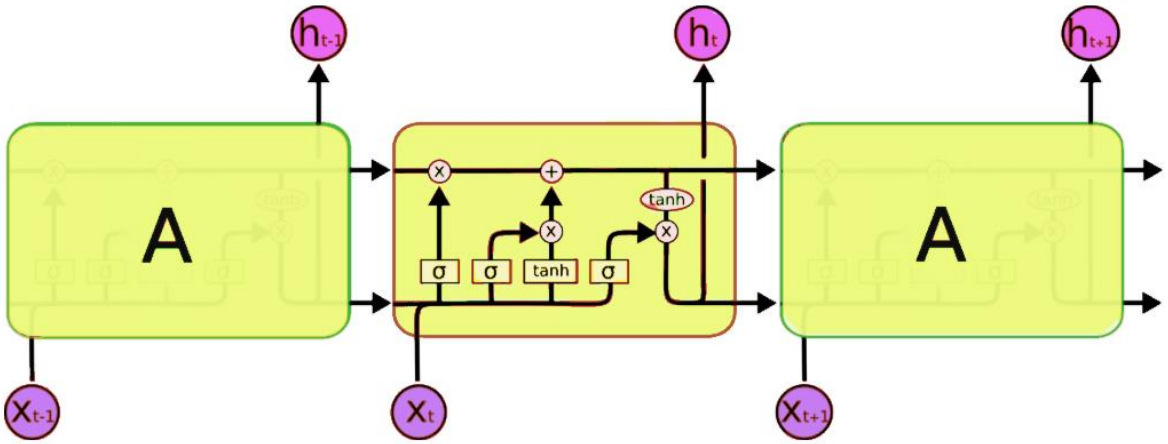


Fig 4.12: LSTM Architecture

Recurrent neural networks like the Long Short-Term Memory network or LSTM add the explicit handling of order between observations when learning a mapping function from inputs to outputs, not offered by MLPs or CNNs. They are a type of neural network that adds native support for input data composed of sequences of observations.

4.2.2 Implementation Issues:

- The complete implementation is done to help users who will use it. Users can rely upon it completely till there are other factors like lockdown or some relief measures are considered because are non statistical factors which can't be taken out of the data.

CHAPTER 5: Testing

5.1 Testing Plan

Type of Testing	Performed(Yes/No)	Explanation	Software Components
Requirement	Yes	Requirement specification must contain all the requirements that are needed by our system.	Manual work needs to plan out all the software requirements, the time needed to develop, technology to be used, etc.
Unit	Yes	Testing technology using which individual modules are tested to determine if there are any issues, by the developer itself.	Unit tests have been written to ensure automated testing of the application
Integration	Yes	Testing where individual components are combined and tested as a group.	Compiling full code and testing
Performance	Yes	Testing to evaluate the input where the best and most optimal output is yielded by the system	Testing results ensure this.

Table 5.1. Testing Plan

5.2 Component Decomposition and Types of Testing

Sr.No.	List of Modules Required for Testing	Type of Tests Required	Techniques used for Testing
1	Fetching Data	Requirement, Unit, Performance, and Integration	White Box
2	Data Augmentation		White Box
3	LSTM Model		White Box
4	Web Dashboard	Integration	Black Box

Table 5.2. Components and Types of Testing

5.3 Test Cases

The next step involves the training of LSTM model. The data is then fit into the model with

Number of Epochs=90

Number of Steps per Epochs=1

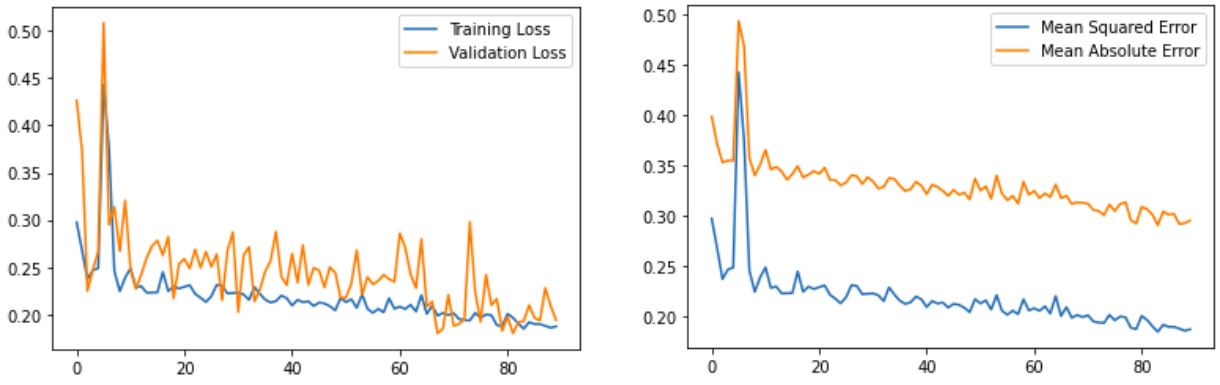


Fig 5.1: Loss and Mean Error Graph

CHAPTER 6: FINDINGS, CONCLUSION AND FUTURE WORK

6.1 Findings

- The proposed model design is found to be efficient in forecasting the Case Load Rate. The use of LSTM for forecasting made the prediction more reliable and efficient.
- The data is found with insight that has various ups and downs. The data is non statistical.
- It shows that Deep Learning gives better results than contemporary Machine Learning techniques as the feature extraction from the image is more and leads to better model accuracy.

6.2 Conclusion

The aim of the project was to forecast the Case Load Rate so as to find out how much daily active cases has increased to the daily confirmed cases. Due to the non statistical trend the data has followed over the past year makes it hard to learn the pattern the data follows. The data pattern is dependent on natural factors such as lockdown and preventive measures. We have built this model keeping in mind the above stated problems. But since we have managed to obtain a decent model with decent accuracy, it can be concluded that the model will definitely perform. The support with the Web Dashboard built on Flask is user friendly and scalable.

6.3 Future Work

On considering the available dataset this model will definitely perform better. But, to increase its integrity and scalability of the model we will work in adding following features.

- We will work on finding the better insight out of the data.
- We will integrate this dashboard with the detection model of Covid-19 through X-Rays.

CHAPTER 7: REFERENCES

[1] Bingjie Yan, Xiangyan Tang, Boyi Liu, Jun Wang, Yize Zhou, Guopeng Zheng,"An Improved Method for the Fitting and Prediction of the Number of COVID-19 Confirmed Cases Based on LSTM.",University of Chinese Academy of Science, Beijing, China.,National University of Defense Technology, Changsha, China,2020

[2] Parul Arora, Himanshu Kumar, Bijaya Ketan Panigrahi "Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India", Department of Electrical Engineering, Indian Institute of Technology, Delhi, New Delhi, India,2020

[3] Novanto Yudistira"COVID-19 growth prediction using multivariate long short term memory",IAENG International Journal of Computer Science,2020

[4]Hafiz Tayyab Rauf, M. Ikram Ullah Lali, Muhammad Attique Khan, "Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks"Department of Computer Science, University of Gujrat, Gujrat, Pakistan, Department of Computer Science, University of Education,Lahore 54770, Pakistan,2020

[5]Shashank Reddy Vadyala, Sai Nethra Betgeri, Eric A. Sherer, Amod Amritphale, "Prediction of the Number of COVID-19 Confirmed Cases Based on K-Means-LSTM",Department of Computational Analysis and Modeling, Louisiana Tech University, Ruston, LA United States, Department of Chemical Engineering, Louisiana Tech University, Ruston, LA United States

[6] Denis A. Pustokhin,Irina V. Pustokhina,Phuoc Nguyen Dinh,Son Van Phan,"An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19",Firat University, Technology Faculty, Electrical-Electronics Engineering Department, Elazig, Turkey,2021

[7] Ratnabali Pal, Arif Ahmed Shekh, Samarjit Kar, ”Neural network based country wise risk prediction of COVID-19”,Department of Computer Science, UiT The Arctic University of Norway, Tromsø 9019, Norway, September 2020

[8]Farah Shahid, Aneela Zameer, Muhammad Muneeb, “Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM”,Department of Computer & Information Sciences, Pakistan Institute of Engineering & Applied Sciences (PIEAS), Nilore, Islamabad 45650, Pakistan

[9]Smita Rath, Alakananda Tripathy, Alok Ranjan Tripathy, “Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model”, Department of Computer Science and Engineering, Siksha ‘O’ Anusandhan Deemed to be University, Odisha, India , Department of Computer Science,Ravenshaw University, Cuttack, India

[10]Sohini Sengupta, Sareeta Mugde, Dr Garima Sharma, “Covid-19 Pandemic Data Analysis and Forecasting using Machine Learning Algorithms”,Welingkar Institute of Management Development and Research,2020

[11] Rohitash Chandra, Ayush Jain, Divyanshu Singh Chauhan, “Deep learning via LSTM models for COVID-19 infection forecasting in India”, Department of Electronics and Electrical, Indian Institute of Technology Guwahati, Assam, India 2021

[12] Yogesh Gautam, “Transfer Learning for COVID-19 cases and deaths forecast using LSTM network”, Department of Mechanical and Aerospace Engineering, Pulchowk Campus, Institute of Engineering, Kathmandu,2020

[13]Peipei Wang, Xinqi Zheng, Gang Ai, et al, “Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method”,Technology Innovation Center for Territory Spatial Big-data, MNR of China, Beijing, China,2020

[14]Farah Shahid, Aneela Zameer, Muhammad Muneeb, “Deep Ensemble Learning Method to Forecast COVID-19 Outbreak”, University of Manouba,2020

[15]Muhammad Iqbal, Feras Al-Obeidat , Fahad Maqbool , Saad Razzaq, “COVID-19 Patient Count Prediction Using LSTM”, IEEE TRANSACTIONS ON COMPUTATIONAL,2020

[16]Vinay Kumar Reddy Chimmula, Lei Zhang, “Time series forecasting of COVID-19 transmission in Canada using LSTM networks”, Faculty of Engineering and Applied Science, University of Regina, Regina, Saskatchewan, S4S0A2 Canada,2020

[17]Abdelhafid Zeroual, Fouzi Harrou, Abdelkader Dairi, “Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study”,Faculty of technology, Department of electrical engineering, University , Algeria , LAIG Laboratory, University Guelma, Algeria

[18]Abdelfatah Hassan, Ismail Shahin, Mohamed Bader Alsabek ,”COVID-19 Detection System Using Recurrent Neural Networks.”,Electrical Engineering Department, University of Sharjah,UAE

[19]Myung Hwa Kim, Ju Hyung Kim, Kyoungjin Lee,”The Prediction of COVID-19 Using LSTM Algorithms”,Department of IT Policy and Management, Graduate School, Soongsil University, Seoul, Korea,2021

[20]Kareem Kamal A. Ghany, Hossam M. Zawbaa, Heba M. Sabri,”COVID-19 prediction using LSTM algorithm: : GCC case study, College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia, Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni Suef, Egypt, Sadat Academy for Management Sciences, Information Systems Dept., Cairo, Egypt