IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 4, April 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

Covid-19 Future Forecasting Using Supervised Machine Learning Models

Manasi Kanade, Neha Gaikwad, Ankita Akkanavru, Shruti Golait, A. V. Dirgule

Dept. of Computer Engineering, Sinhgad College of Engineering, Pune, India

ABSTRACT- Machine learning (ML)-based forecasting techniques have demonstrated their use in predicting preoperative outcomes and improving decision-making about future activities. Many application domains that required the detection and evaluation of adverse aspects for a threat have long employed machine learning models. To deal with forecasting challenges, a variety of prediction approaches are widely utilised. This research illustrates the capacity of machine learning models to predict the future disease of patients who would be afflicted by COVID-19, which is now regarded as a possible threat to humanity.

KEYWORDS: Machine learning, LSTM

I. INTRODUCTION

MACHINE LEARNING (ML) has established itself as a key study topic during the last decade by addressing a host of exceedingly intricate and sophisticated real-world situations.

Among the application areas examined were healthcare, autonomous vehicles (AV), business applications, natural language processing (NLP), intelligent robots, gaming, climate modeling, voice, and image processing. Unlike traditional algorithms, which follow computer instructions based on decision statements such as if-else, machine learning algorithms rely heavily on trial and error to learn.

Various regression and neural network models have a wide range of application when it comes to forecasting the future health of individuals with a given disease. There have been several research employing machine learning approaches to forecast various diseases, such as coronary artery disease, cardiovascular disease prediction, and breast cancer prediction.

The project is focused on the live forecasting of COVID-19 confirmed cases, as well as the forecasting of a COVID-19 outbreak and early reaction.

COVID-19 is currently posing a significant threat to human life across the planet. The virus was originally discovered in late 2019 in the Chinese city of Wuhan, when a high number of patients suffered pneumonia-like symptoms.

Every day, thousands of new persons are reported to be positive from all over the world. Close personal contact, respiratory droplets, and touching infected surfaces are the most common ways for the virus to spread.

Among all the safeguards, "being informed" on all elements of COVID-19 is seen as crucial. Numerous researchers are examining the various features of the pandemic and producing conclusions to aid mankind to contribute to this component of knowledge.

II. LITERATURE SURVEY

FURQAN RUSTAM et.al [1] Machine learning (ML)-based estimating components have shown their value in predicting preoperative outcomes in order to improve decision-making on future actions. For a long time, ML models have been used in a variety of applications that needed identifiable proof and prioritizing of negative variables for a risk. To cope with gauging concerns, a few expectation strategies are most commonly utilized. This work demonstrates the capacity of machine learning models to predict the number of future patients affected by COVID-19, which is now regarded as a potential threat to humanity.In this study, four common gauging models were used to predict the COVID-19 compromising variables: straight relapse (LR), least outright shrinkage and choice administrator (LASSO), support vector machine (SVM), and outstanding smoothing (ES). Every one of the models makes three types of predictions: the number of freshly contaminated cases, the number of passings, and the number of recoveries in the next 10 days. The review's findings show that incorporating these strategies into the COVID-19 pandemic situation is a promising component.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

Mujeeb Ur Rehman et.al [2] The COVID-19 pandemic has elicited emotional responses all throughout the world, just a year after the WHO issued its official pronouncement. A large number of antibody doses were previously regulated in only a few countries. In any event, the positive effects of these vaccinations are likely to take longer than expected. In these circumstances, detecting COVID-19 quickly remains the most effective strategy to halt the spread of the illness. However, relying solely on obvious side symptoms, it is difficult to predict whether or not a person is infected with COVID-19. In this unusual situation, author propose using AI (ML) algorithms to more thoroughly assess COVID-19-infected individuals.The proposed analytical technique considers a variety of side effects, including flu symptoms, throat pain, susceptibility status, diarrhoea, voice type, internal heat level, joint pain, dry cough, retching, breathing problems, migraine, and chest pain. Given the negative effects that have been documented as a result of ML, our suggested approach can predict the chance of infection with COVID-19. Different exploratory study metrics such as exactness, accuracy, review, and F1-score are used to evaluate this technique. The obtained testing results revealed that the suggested technique can accurately predict the existence of COVID-19 by more than 97 percent.

Yanping Zhang, zhangyp et.al [3] An episode of 2019 novel Covid infections (COVID-19) in Wuhan, Hubei Province, China has spread rapidly cross country. Here, creators report consequences of a spellbinding, exploratory investigation of all cases analyzed as of February 11, 2020. All COVID-19 cases revealed through February 11, 2020 were separated from China's Infectious Disease Information System. Investigations incorporated the accompanying: 1) synopsis of patient attributes; 2) assessment old enough disseminations and sex proportions; 3) computation of case casualty and death rates; 4) geo-fleeting examination of viral spread; 5) epidemiological bend development; and 6) subgroup examination.

Dr. Vakula Rani J#1 and, Aishwarya Jakka#2 [4] COVID-19 pandemic has impacted the economy and changed the human lifestyle, upsetting everybody's psychological, physical, and monetary prosperity. A significant number of the quickest developing economies are stressed inferable from the seriousness and coherence of the pandemic. In view of the rising variety of cases and the subsequent weight on medical care professionals and the public authority, thusly, foreseeing the quantity of tainted COVID-19 cases which could be helpful in arranging the expected emergency clinic assets later on. In creators paper, they focussed on data drove strategies for assessing the quantities of COVID-19 affirmed cases in the nation and their suggestions later on, utilizing different learning models, for example, Sigmoid displaying, ARIMA, SEIR model and LSTM, for defensive measures, for example, social separation or the lockout of COVID-19.

1Saud Shaikh,et.al [5] In creators paper, they are foreseeing and estimating the COVID-19 episode in India in view of the AI approach, where they expect to decide the ideal relapse model for an inside and out examination of the original Covid in India. They are carrying out the two relapse models specifically straight and polynomial and assessing the two utilizing the R squared score and blunder values. The COVID-19 dataset for India is being utilized to serve the examination of this paper. The model is foreseeing the quantity of affirmed, recuperated, and passing cases in view of the information accessible from March 12 to October 31, 2020. For anticipating the future pattern of these cases, we are using the time series guaging approach of scene. Moreover, the time series guaging technique is being utilized to estimate the absolute include of affirmed cases from here on out.

Saksham Gera et.al [6] The COVID - 19 (Novel Corona Virus) Pandemic has strike the world and cause an incredible annihilation throughout everyday life. It is considered as one of the deplorable Pandemic since the beginning of time. This paper expects to gives understanding of how various models of ML are invention experiencing the same thing. Notwithstanding the relapse investigation performed on Indian information, the review analyzes contemporaneous example or pattern in COVID - 19 transmissions in India. Additionally, guaging framework in light of Machine Learning has shown its significance for improvement of the administrative capacity on following game-plan. This examination shows the capacity of various models of Machine Learning to guess the quantity of coming patients impacted by nCov, a final offer to humankind. In this concentrate on 5 models: LR, SVM, Random Forest, KNN and ES have been utilized. Two sorts of forecast are made by each model: 1) number of recently sure affirm cases 2) number of passings. This study demonstrates that among all models ES perform best followed by Random Forest and KNN which perform better compared to SVM that perform inadequately in all forecast regions.

Ovi Sarkar et.al [7] SARS-CoV-2 (n-Covid) is a worldwide pandemic that causes the passings of millions of individuals around the world. It can cause Pneumonia and extreme intense respiratory disorder (SARS) and lead absurdly in serious cases. An asymptomatic illness solidifies our life and work conditions. As there is no successful treatment accessible, numerous researchers and specialists are making an honest effort to battle t he pandemic. This paper centered o n the Covid pandemic circumstance in the worldwide and Bangladesh district and its connected impacts and future status. Creator have used different data portrayal and AI estimations to reproduce the avowed,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

recovered, and passing cases. Creator accept the exploration will help researchers, analysts, and common individuals foresee and investigate this pandemic's effect. At long last, the correlation and examination of various models and calculations effectively showed our representation and expectation achievement.

Shreyansh Chordia and Yogini Pawar et.al [8] The phenomenal flare-up of the COVID-19 infection has tainted in excess of 50 million individuals all around the world in under a year. More than 1 million individuals have lost their lives because of the continuous pandemic. The pandemic struck India on January 30, 2020, when the main positive instance of COVID-19 was distinguished in Kerala. Today, India is one of the most unfavorably impacted nations on the planet. Henceforth, it is of most extreme significance to examine the patterns in India and utilize the embraced information to conjecture the future course of results. Alongside the general pattern investigation in India, this concentrate likewise considers 5 most impacted conditions of the nation: Maharashtra, Andhra Pradesh, Tamil Nadu, Karnataka and Uttar Pradesh as the subjects of the examination.

Narayana Darapaneni et.al [9] In creators paper, they have dissected the COVID-19 movement in India and the three most impacted Indian states (viz. Maharashtra, Tamil Nadu and Andhra Pradesh) starting at 29-Aug-20 and fostered an expectation model to estimate the way of behaving of COVID-19 spread later on months. They involved time series information for India and applied the Susceptible-Infective-Removed (SIR) model and the FbProphet model to foresee the pinnacle infectives and top infective date for India and the three most impacted states. In this paper, they further played out the relative examination of the forecast outcomes from SIR and FbProphet models.

Ashish U Mandayam1 et.al [10] With the movement in the field of AI, prescient examination has turned into a critical part for future expectation. As creator face the COVID-19 pandemic, foreseeing the future number of positive cases for better measures and control would be useful. They utilized two managed learning models to anticipate the future utilizing the time-series dataset of COVID-19. To concentrate on the presentation of expectation, the examination between Linear Regression and Support Vector Regression is done. They have involved these two models as the information were practically direct.

III. PROBLEM STATEMENT

To design machine learning based algorithms to correctly predict the future cases or vaccine status.

In our system LSTM model solve the all problems and get the accurate result not waste of time.

IV. PROPOSED SYSTEM

The study is on COVID-19 predictions, which is a new corona virus. The COVID-19 has shown to be a real and present danger to human life. It kills tens of thousands of people every year, and the death rate is rising every day throughout the world. This study aims to anticipate the mortality rate, the number of daily confirmed infected patients, and the number of recovery cases in the next 10 days in order to help control the pandemic scenario. LSTMalgorithm were used to forecast the data.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

1. Image processing

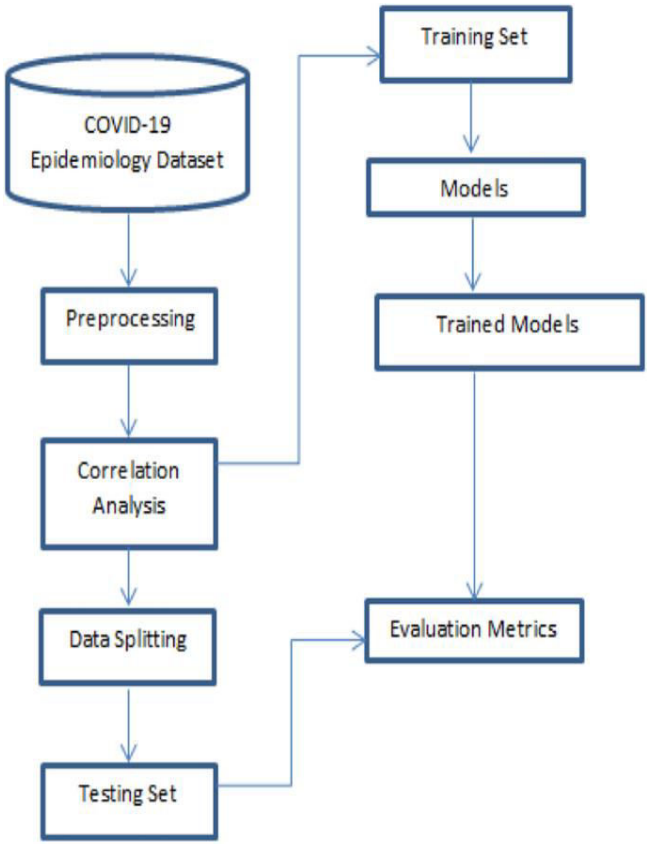


Figure 4.1: Architecture of proposed system

Data preprocessing:

Preprocessing is the process of converting raw data into a format suitable for machine learning. A data scientist can receive more exact findings from a machine learning model when the data is structured and tidy. Data formatting, cleansing, and sampling are all part of the process.

Data formatting:-

When data is collected from a variety of sources by different persons, data formatting becomes increasingly important. A data scientist's initial responsibility is to standardize record formats. A professional examines the variables that represent each characteristic to see if they are all recorded in the same way. Variables include product and service names, pricing, date formats, and addresses. The notion of data consistency also applies to numeric ranges as characteristics.

Data sampling: -

For analysis, large datasets need greater time and computing resources. When a dataset is too huge, data sampling is the best option. This strategy is used by a data scientist to choose a smaller but representative data sample in order to construct and test models faster while still producing reliable results.

Featuraization:-

Featuraization is a technique for converting text, graph, and time-series data into numerical vectors. Featuraization is not to be confused with feature engineering. Feature engineering is simply the process of altering numerical characteristics in such a way that machine learning models can function properly.

Data splitting:-Machine learning datasets should be divided into three subsets: training, test, and validation sets.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

Training set: - A data scientist utilizes a training set to train and define a model's ideal parameters, which it must learn from data.

Test set: - A test set is required for evaluating the trained model's generalization capacity. The latter refers to a model's capacity to spot patterns in previously unknown data once it has been trained on it. To avoid model over fitting, which leads to the inability to generalize discussed earlier, it's critical to employ separate subsets for training and testing.

Modeling:-During this step, a data scientist trains a variety of models in order to determine which one makes the most accurate predictions. Training as a model:- A data scientist can begin model training after preprocessing the acquired data and dividing it into three subsets. The algorithm is "fed" with training data throughout this process. An algorithm will evaluate data and generate a model capable of detecting a target value (attribute) in fresh data — the answer you're looking for using predictive analysis. Model development is the goal of model training.

LSTM

A recurrent neural community is a sort of lengthy brief time period reminiscence. The output of the preceding step is used as enter withinside the contemporary step in RNN. Hochreiter&Schmidhuber created the LSTM. It addressed the problem of RNN lengthy-time period dependency, wherein the RNN is not able to expect phrases saved in lengthy-time period reminiscence however could make extra correct predictions primarily based totally on contemporary facts. RNN does now no longer offer an green overall performance as the distance period rises. By default, the LSTM can also additionally store the statistics for a protracted time. It is used for time-collection facts processing, prediction, and classification.

Structure of LSTM:

LSTM has a chain structure that contains four neural networks and different memory blocks called cells.

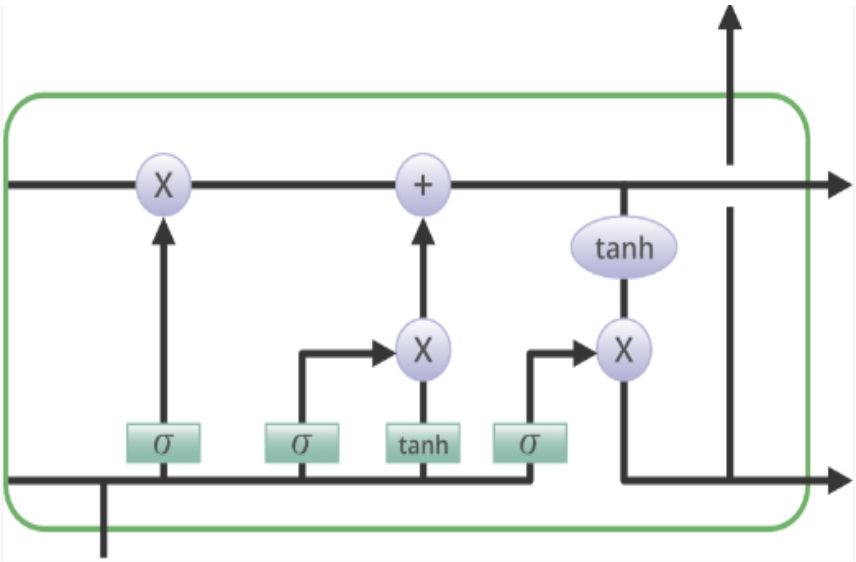


Fig 4.2: LSTM Structure

Information is retained by the cells and the memory manipulations are done by the gates. There are three gates –

1. Forget Gate: The forget gate removes information from the cell state that is no longer helpful. The gate receives two inputs, x t (input at a certain time) and h t-1 (prior cell output), which are multiplied with weight matrices before bias is added. The result is sent into an activation function, which outputs a binary value. If the output for a given cell state is 0, the piece of information is lost, however if the output is 1, the information is saved for future use.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

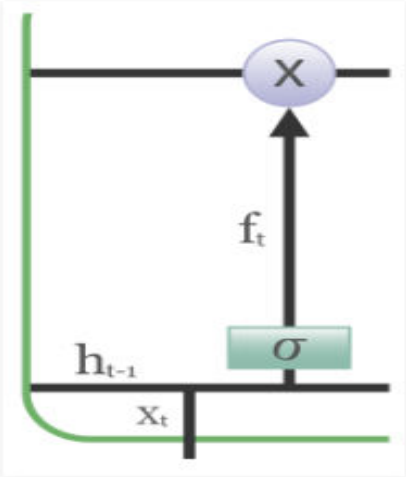


Fig 4.3: Forget Gate of LSTM

2. Input gate: The input gate is responsible for adding important information to the cell state. The information is first controlled using the sigmoid function, which filters the values to be remembered using the inputs h t-1 and x t, similar to the forget gate. Then, using the tanh function, which returns a value between -1 and +1, a vector is generated that contains all of the possible values from h t-1 and x t. Finally, the vector's values are multiplied by the controlled values to produce the usable information.

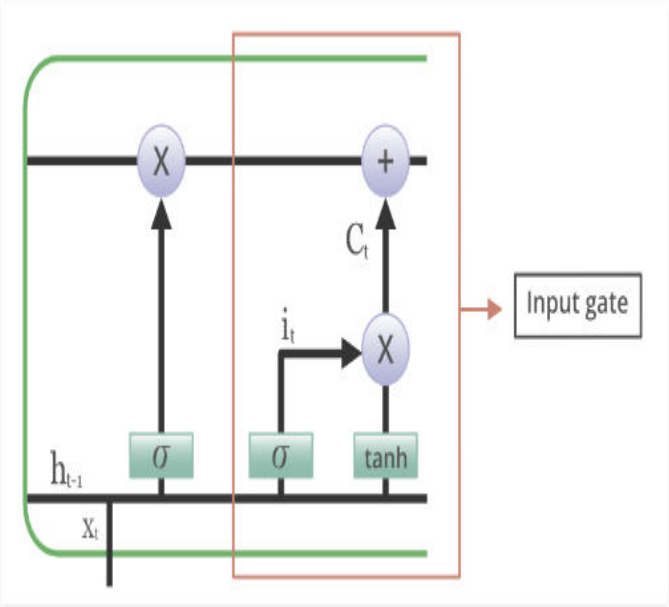


Fig 4.4: Input Gate of LSTM

3. Output gate: The output gate's job is to extract meaningful information from the current cell state and display it as output. The cell is first used to construct a vector using the tanh function. The information is then filtered by the values to be remembered using the sigmoid function and inputs h t-1 and x t. Finally, the vector's values are multiplied by the regulated values and supplied as an output and input to the next cell.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

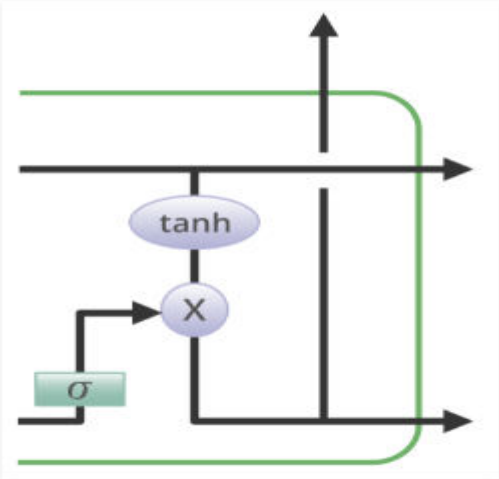


Fig 4.5: Output Gate of LSTM

V. RESULT AND DISCUSSION

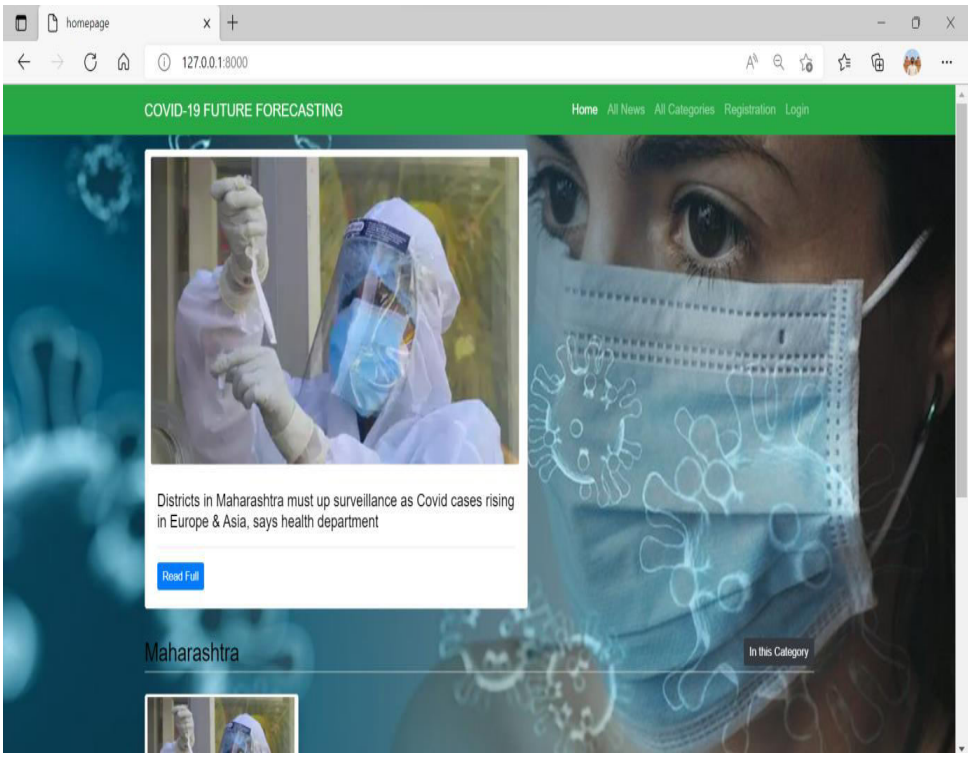


Fig 5.1: Home Page

In the above fig we can see the home page of our system Covid-19 Future Forecasting by using LSTM.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

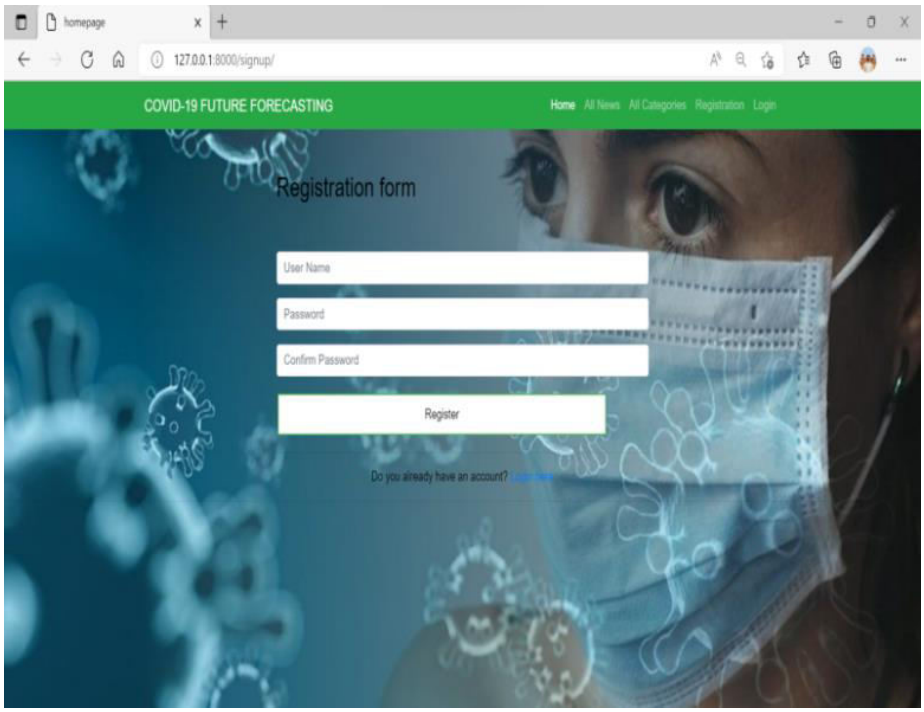


Fig 5.2: Registration Page

In the above fig we can see the registration page that means if we can register in that system then first we register in that page. We can enter the User name, Password, Confirm the Password and Register in that system.

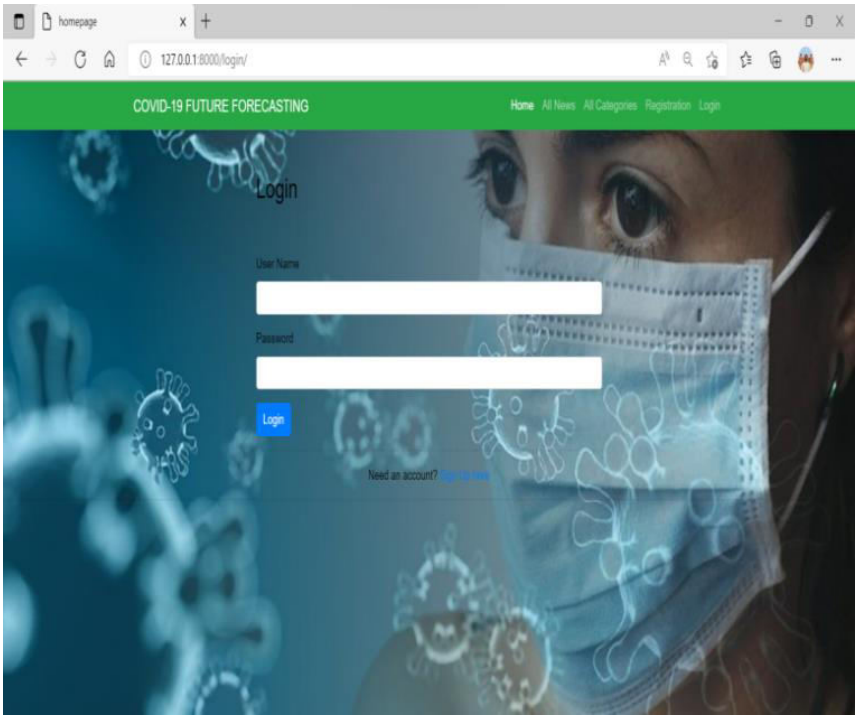


Fig 5.3: Login Page

In the above fig we can see the Login page of our system. This indicates that if we can login in the system we can enter the User name and Password.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

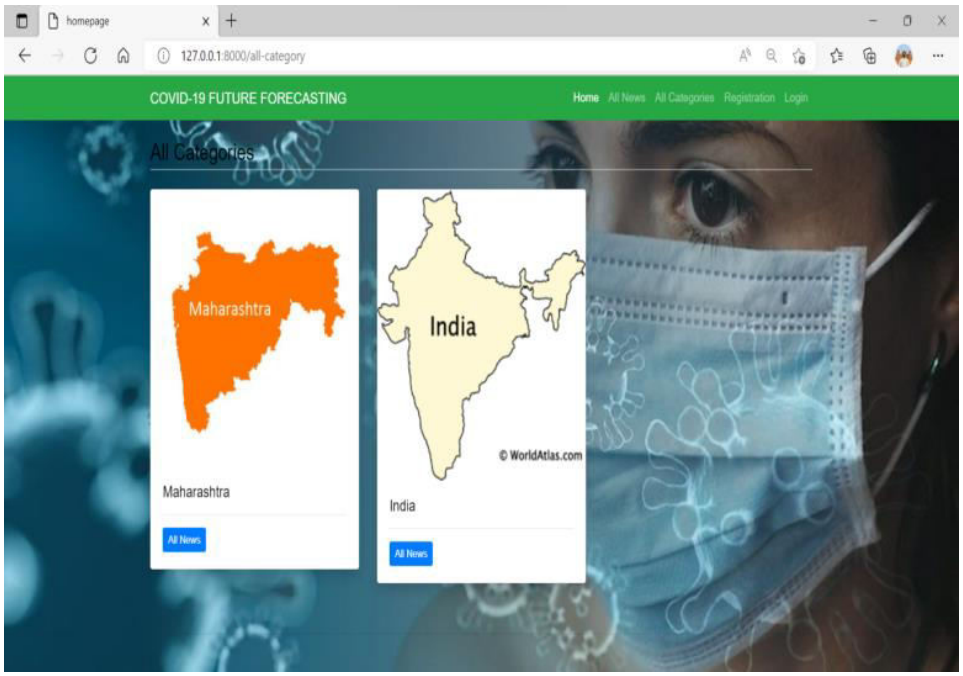


Fig 5.4: News Category

In this above fig we can see there is News Category seen in that fig. There we can see the graph of Maharashtra and the graph of the India for Covid-19 Future Forecasting.

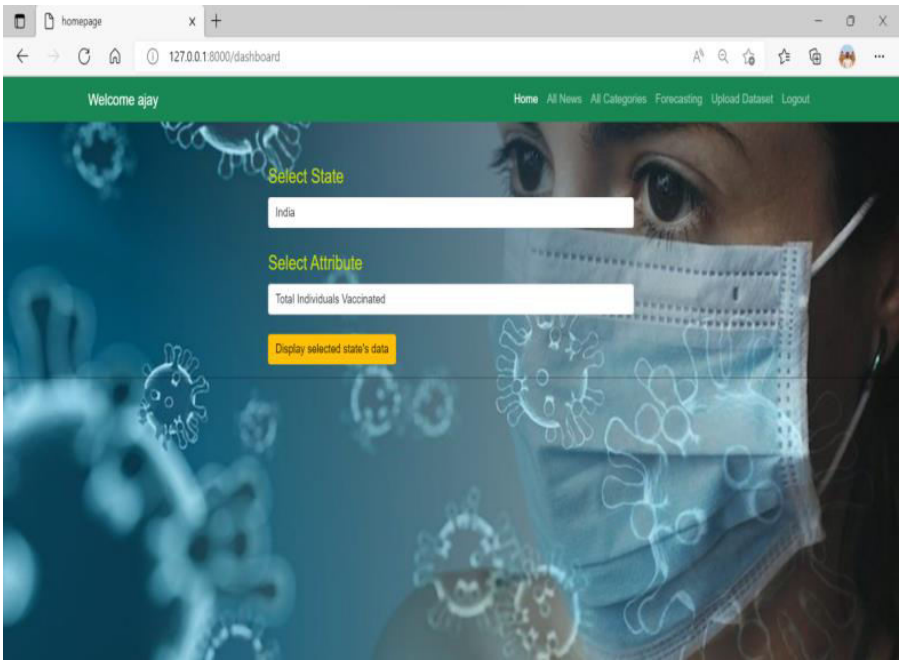


Fig 5.5: Input to State

We can choose the state that we have to forecast the future of Covid-19. In that fig we can see the select state and select attribute. After that we can display the selected state's data in the next fig.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

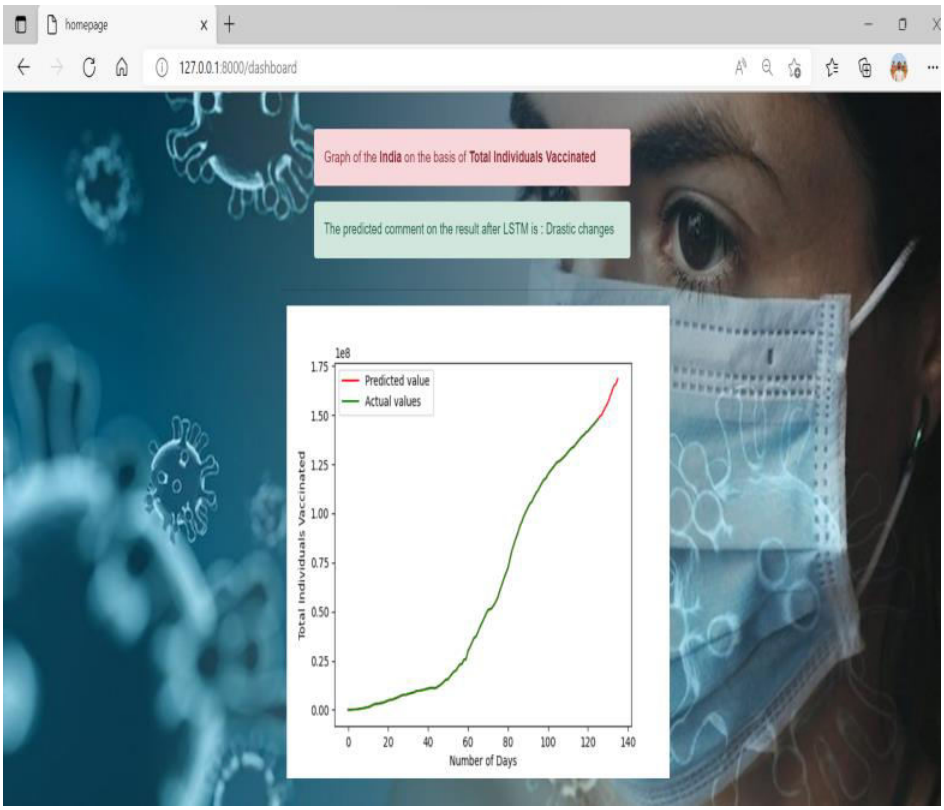
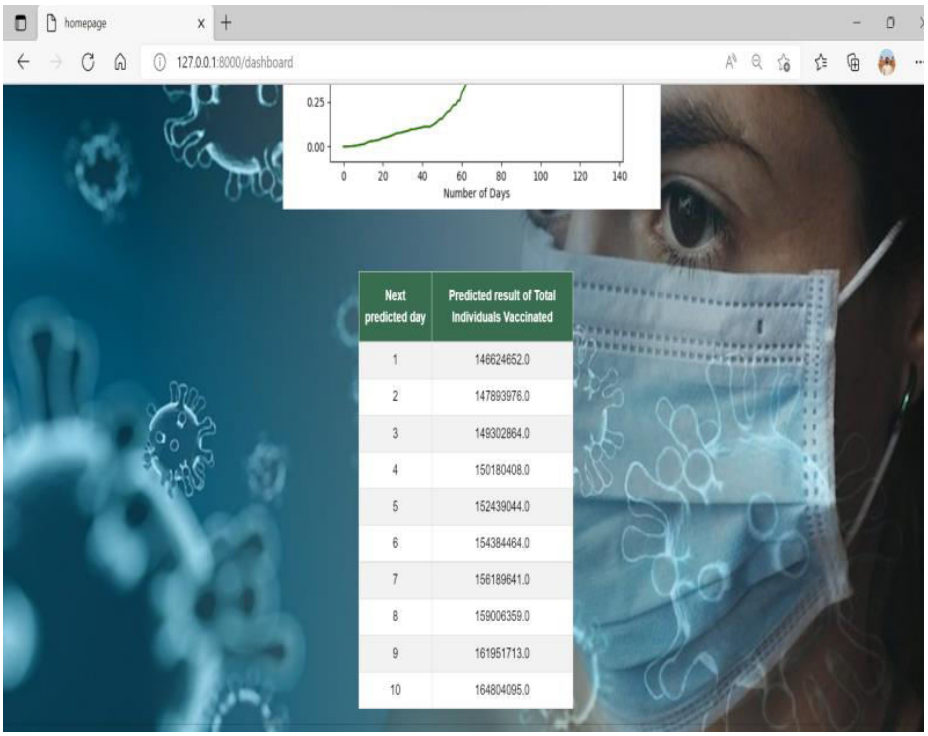


Fig 5.6: Graph for result of selected state

If we select the state is India. We can display the graph of Number of days and total number of vaccinated. The predicted comment on the result after LSTM is the drastic changes.



Next predicted day	Predicted result of Total Individuals Vaccinated
1	146624652.0
2	147893976.0
3	149302864.0
4	150180408.0
5	152439044.0
6	154384464.0
7	156189641.0
8	159006359.0
9	161951713.0
10	164804095.0

Fig 5.7: Predicted values

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004115|

In this above fig we can see the table that can predict the next day. That is it can predict the result of total individual vaccinated. By using LSTM we can gives better results for forecasting the future of Covid-19.

VI.CONCLUSION

Using machine learning approaches, this work seeks to establish a system for predicting the number of cases impacted by COVID-19 in the future. The study's dataset includes daily records on the number of newly infected patients, recoveries, and fatalities caused by COVID-19 across the world. As the fatality rate and verified cases rise day by day, the globe faces an unsettling scenario. The number of persons who might be impacted by the COVID-19 pandemic in various parts of the world is unknown.

REFERENCES

[1] FURQAN RUSTAM 1, AIJAZ AHMAD RESHI 2, (Member, IEEE), ARIF MEHMOOD 3, SALEEM ULLAH 1, BYUNG-WON ON4, WAQAR ASLAM 3, (Member, IEEE), AND GYU SANG CHOI 5 "COVID-19 Future Forecasting Using Supervised Machine Learning Models "Received May 4, 2020, accepted May 13, 2020, date of publication May 25, 2020, date of current version June 10, 2020.

[2] Mujeeb Ur Rehman 1,* ,ArslanShafique 1, Sohail Khalid 1, MahaDriss 2,3 and Saeed Rubaiee 4 "Future Forecasting of COVID-19: A Supervised Learning Approach "Sensors 2021, 21, 3322. https:// doi.org/10.3390/s21103322 Received: 5 April 2021 Accepted: 6 May 2021 Published: 11 May 2021

[3] Yanping Zhang, zhangyp "The Epidemiological Characteristics of an Outbreak of 2019 Novel Corona virus Diseases (COVID-19) — China, 2020 "Submitted: February 14, 2020; Accepted: February 14, 2020

[4] Dr. Vakula Rani J#1 &Aishwarya Jakka#2 "Forecasting COVID-19 cases in India Using Machine Learning Models "Authorized licensed use limited to: IEEE Xplore. Downloaded on January 24, 2022 at 11:34:26 UTC from IEEE Xplore.

[5] 1Saud Shaikh, 2Jaini Gala, 1Aishita Jain, 1Sunny Advani, 1Sagar Jaidhara, 1Dr. Mani Roja Edinburgh "Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting "2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.9377137

[6] Saksham Gera, MrMridul , Mr. Kireet Joshi "Regression Analysis And Future Forecasting Of COVID-19 Using Machine Learnings Algorithm. "2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.9377065

[7] Ovi Sarkar , MdFaysalAhamed , Pallab Chowdhury "Forecasting & Severity Analysis of COVID-19 Using Machine Learning Approach with Advanced Data Visualization "2020 23rd International Conference on Computer and Information Technology (ICCIT), 19-21 December, 2020

[8] Shreyansh Chordia & Yogini Pawar "Analyzing and Forecasting COVID-19 Outbreak in India "2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.937711

[9] Narayana Darapaneni , Praphul Jain , RohitKhattar , Manish Chawla, RijyVaish "Analysis and Prediction of COVID-19 Pandemic in India "2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) |978-1-7281-8337-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICACCCN51052.2020.9362817

[10] Ashish U Mandayam1 , Rakshith.A.C2, Siddesha S3, S K Niranjan4, "Prediction of Covid-19 pandemic based on Regression "Authorized licensed use limited to: IEEE Xplore. Downloaded on January 24, 2022 at 11:39:32 UTC from IEEE Xplore.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details