Covid-19 Future Forecasting using Machine Learning and Deep Learning Models

Megha M Veerkar Dept.of ECE Vishwanathrao Deshpande Institute Of Technology Haliyal, India <u>meghaverker143@gmail.com</u>

Abstract— Forecasting mechanism used by Machine learning (ML) and deep learning model have proved their significance to predicting the in operative outcomes to enhance the choice making on the longer term course of actions. The ML models and deep learning models have long been utilized in many application domains which needed the identification and prioritization of adverse factors for a threat. Several prediction methods are being popularly wont to handle forecasting problems. This study demonstrates the potential of ML and Deep Learning models to forecast the amount of upcoming patients suffering from COVID-19 which is presently considered as a possible threat to mankind. In particular, three standard forecasting models polynomial regression Autoregressive (PR), Integrated moving average model (ARIMA), and **CNN** based Long Short-Term Memory networks(CNN -LSTM) have been used in this study to forecast the threatening factors of COVID-19. Two types of predictions are made by each of the models, such as the number of newly infected cases, the number of deaths in the next 10 days. The result gained by this study proves that these methods work best for the current scenario of the covid-19 pandemic.

IndexTerms—COVID-19, polynomial regression, ARIMA, CNN-LSTM.

I. INTRODUCTION

The COVID-19 virus was first detected in Wuhan, China, in December 2019. It is characterized by symptoms such as fever, fatigue, and dry cough. The virus, initially linked to a seafood and live animal market in Wuhan, is believed to have originated from zoonotic transmission. However, it rapidly spread Prof.Meenaxi T Dept.of ECE Vishwanathrao Deshpande Institute Of Technology Haliyal, India <u>torase.meenaxi@gmail.com</u>

through human-to-human transmission, leading to widespread outbreaks globally. On January 30, 2020, the World Health Organization (WHO) declared the outbreak a Public Health Emergency of International Concern (PHEIC), and later, on March 11, 2020, the WHO declared it a global pandemic due to its widespread transmission and severe impact on public health[1]. COVID-19 is spreading rapidly across the globe, with the number of cases doubling in increasingly shorter timeframes, posing a significant global threat, particularly to older adults and those with underlying health conditions. The virus primarily affects the respiratory system, leading to severe respiratory issues in some cases, while others may remain asymptomatic, unknowingly contributing to the spread of the virus. As of now, a vaccine for COVID-19 is not yet available due to the complexity of its pathological mechanisms. In the meantime, public health measures such as maintaining social distance and enhancing immune system health are crucial in the fight against the virus.[3] The death toll from COVID-19 is likely to increase as many patients may not survive until a vaccine is developed. However, early identification of infected individuals can help control the number of deaths[4].

Machine learning model (ML) and deep learning algorithms are extensively used in the study for solving many very complex and sophisticated real-world problems. ML has the capacity of data visualization, data analysis and gives more accurate results for the given information. In most of the areas like healthcare, business applications, intelligent robots, gaming, climate modelling ML models are used.[5].

ML algorithms are works unproved and error methods because the conventional algorithms follow the programming instructions supported decision statements like if-else or nested-if, and different loops like for, while. Deep learning is a branch of machine learning that holds different sections to generate an "artificial neural network",[6]. It is able to take brilliant

Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed
1	1/30/2020	6:00 PM	Kerala	1	0	0	0	1
2	1/31/2020	6:00 PM	Kerala	1	0	0	0	1
3	2/1/2020	6:00 PM	Kerala	2	0	0	0	2
4	2/2/2020	6:00 PM	Kerala	3	0	0	0	3
5	2/3/2020	6:00 PM	Kerala	3	0	0	0	3
6	2/4/2020	6:00 PM	Kerala	3	0	0	0	3
7	2/5/2020	6:00 PM	Kerala	3	0	0	0	3
8	2/6/2020	6:00 PM	Kerala	3	0	0	0	3
9	2/7/2020	6:00 PM	Kerala	3	0	0	0	3
10	2/8/2020	6:00 PM	Kerala	3	0	0	0	3

resolution on its own. The ML and deep learning are

foremost used for foreseeing the various standards of algorithms, abundant standard ML algorithms are utilized to guide the longer term course of actions required in many application areas like disease foretelling, weather forecasting etc [7]. There are bunch of researches performed for the prediction of different diseases using machine learning techniques like breast cancer prediction, cardiovascular disease prediction and coronary artery disease prediction[8]. In present condition ML and deep learning techniques are used in analysis of the covid-19 affected patients and foreseeing the condition of the patients in upcoming days. These models have the capacity to foretell the state of COVID19 affected cases and COVID19 out break and early response.

The key features of this study are listed below.

- CNN-LSTM performs best for time series forecast.
- Different ML algorithms are also used for different class of predictions.

The rest of the paper consists of six sections. Section I presents the Introduction, section II contains the description of the data set and methods used in this study. Section III Block diagram. Section IV presents the methodology, section V presents the results, and section VI summarizes the paper and presents the conclusion.

MATERIALS AND METHOD

A. DATASET

The aim of this paper is to predict the new infected cases and death cases of covid-19 for next 10days. The dataset

for this study has been obtained from Kaggle.The dataset file consisting the information regarding dates,

infected cases and lost cases on daily base of different states in India. The table 1 shows the sample dataset.

Table 1: Covid-19 patient details

B. MACHINE LEARNING AND DEEP LEARNING MODEL

The machine learning models are used for prognosticating the given dataset. Basically the Machine learning models works on trial and error methods and deep learning models have capacity to automatically learn the temporal dependencies present in time series. Different regression model are applied to trained model which generates the output for given dataset.Details of the models used for this study are explained as following.

- a. Polynomial Regression
- b. Autoregressive Integrated moving average
- c. CNN-LSTM

a. POLYNOMIAL REGRESSION

Polynomial regression is a statistical technique commonly used in demography to analyze the relationship between an investigational variable (x) and a progeny variable (y). This method models the relationship as an nth degree polynomial, allowing for a more flexible and accurate representation of the data. When traditional statistical regression models fail to capture the complexity of the relationship between the variables, polynomial regression can be a useful tool.By fitting a polynomial function to the data points, researchers can better describe and predict the relationship between the variables with minimal error.

Overall, polynomial regression provides a more nuanced and detailed analysis of the relationship between dependent and independent variables, offering valuable insights into demographic trends and patterns. The equation for polynomial regression can be written as

 $y = b_0 + b_1 x + b_2 x^2 + \epsilon$ (1) where y is dependent variable and x is independent and $\boldsymbol{\epsilon}$ is the error term associated the regression model.

b. AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

An alternative method to predict time series data and offer a insightful perspective on the Covid-19 situation is through the use of the ARIMA model. The primary objective of this model is to elucidate the relationships between different data points.

By analyzing the discrepancies between values in the series rather than the actual values themselves, the model is able to forecast future trends. The model foretell the future stages by examining the differences between values in the series instead of through actual values.

The frame work of the ARIMA are

AR (p): The amount of fall behind observations.

 $I \ (d)$: The number of times primary $\ data \ are subtracted$.

MA (q): The sequence of operating window.

ARIMA model can be written as

 $y't = c + \phi_1 y't - 1 + \dots + \phi_p y't - p + \theta_1 \epsilon_{t-1} + \dots + \theta_q$ $\epsilon_{t-q} + \epsilon_r$ (2)

where y't is the differenced series and right hand side include lagged value of yt and lagged error. To determine the order of ARIMA model, a useful criterion is the Akaike information criterion (AIC). It is written as

AIC = $-2 \log (L) + 2(p+q+k)$ (3)

Where L is the likelihood of the data, p is the order of the autoregressive part and q is the order of the moving average part. The k represents the intercept of the ARIMA model.

For AIC, if k = 1 then there is an intercept in th ARIMA model ($c \neq 0$) and if k = 0 then there is no intercept in the ARIMA model (c = 0). The corrected AIC for ARIMA models can be written as

$$AICc = AIC + 2 (p+q+k+1) (p+q+k+2)$$
(T-p-q-k-2) (4)

Smaller the value of AIC, better the performance of the

model.

c. CNN-LSTM

The feature of CNN is listening the foremost sequence for given time step. The CNN will expound every sub-samples given data by two time step and samples with the four stage as input and one stage as output. Every specimen then divided into two sub – sequence for given time step. The CNN will expound every sub-samples given data by two time step and gives the clarity of sub-sequence to LSTM as input.

The architecture of CNN LSTM, the CNN is used for quality exacting on the given data. LSTM is used to reinforce the series of foretell. This architecture is also called as LRCN, but now it's called as "CNN-LSTM". The keras is used to trained the CNN- LSTM model.

A CNN LSTM model is usually examine by appending layer of CNN one side, accompany by the LSTM layer and dense layer as output. LSTM Model for expound the options across time steps.CNN model and a sequence of LSTM models, one for every time step. The CNN model is used to consider the input file and forward the file to output for every one time step.

In the beginning separate the series of input into sub- sequence and they are operated by CNN model along with its degree. Consider one specimen ,we have tendency to tend to able to initial split our univariate data points information as input and output samples with the four stage as input and one stage as output. Every specimen then divided into two sub –sequence for given time step. The CNN will expound every subsamples given data by two time step and gives the clarity of sub-sequence to LSTM as input.

From this conclusion is that the outline variety is the amount of sub-sequence for n number sequence and collecting the figure of it's slow moves per subsequence for n steps.

Then the pc file will be remodeled to process the specific shape. The dense layer might even be natural network layer that 's connected deeply, that implies the earlier data from the all the cells of neurons as collected and given as input to the dense layer. The dense layer perform multiplication of the matrix vector. The merits used inside the matrix are literally variables which are instructed and renovated for back propagation. M

Т

dimensional vector are nursed by result of the dense layer.

C) EVALUATION PARAMETER : R-SQUARED SCORE

The R-squared (R2) score is a valuable statistical tool used to evaluate the effectiveness of regression models. This metric provides insight into the percentage of variability in the dependent variable that can be explained by the independent variable. By assessing the R2 score, we can gain a better understanding of the relationship between the dependent variable and the regression models on a scale ranging from 0 to 100%. Once a regression model has been trained, it is crucial to examine the quality of the model by analyzing the R2 score. This score signifies how closely the data points cluster around the regression line, serving as a measure of the coefficient of determination. The R2 score always falls between 0 and 1, with 0 indicating a weak relationship between variables and 1 indicating a strong relationship.

A high R2 score indicates that the trained model is accurate and effectively explains the variability in the independent variable. Essentially, R2 serves as a linear model that quantifies the percentage of variation in the independent variable that can be accounted for by the regression model.

It can be found as $R^2 = \frac{Variance explained by model}{Total variance}$ (5)

III BLOCK DIAGRAM

This section explain the block diagram of purposed work. Dataset is obtained from the Kaggle. preprocessing the raw dataset. Splitting the dataset as trained and test dataset. And applying the ML and deep learning regresses on trained dataset to obtained the best fit curve. The accuracy of three models are compared by R^2 method.



Figure 1 Purposed Block diagram

IV. METHODOLOGY

The study is about novel corona virus also called COVID19 estimation.. The year 2019 was considered as curse on human life. Due to this Covid-19 lots of peoples from all over the country lost their life. Due to this covid-19, the economic condition in India become very poor. Many researches were going on to help the human beings to come out of this tuff situation. This study focusing on foretelling of covid-19 freshly infected cases and figure of mortality. In order to implement the proposed framework two ML and one deep learning algorithms are used.

To achieve the above object the dataset is collected which include the details of covid-19 affected cases and figure of mortality. At the beginning ,given data set has to be pre-processed to remove the unwanted data from the dataset.

Then the data is broken into test dataset and train dataset... The educational algorithms like PR, ARIMA and CNN-LSTM are employed in this proposed framework. These educational models efficiency is tested by MSE reported within the results.

V. RESULTS AND DISCUSSION

This section examines the outcome accomplished for proposed framework. This study attempts to develop a system for the future forecasting



of the number of cases affected by COVID-19 using machine learning and deep Learning methods. This study is an effort to forecast the amount of individuals which will be affected in terms of latest infected cases and deaths for the upcoming 10 days. Machine learning models PR, ARIMA and deep learning model CNN-LSTM have been used to predict the number of newly infected cases and the number of deaths. The study perform the prediction of newly predicted cases and death cases of Covid-19, CNN-LSTM model perform the best among the three models. Table 2 shows the R² values.



shows CNN-LSTM model applied on given dataset and figure 10,11 shows future prediction of data using CNN-LSTM all represented in the form of graph.



Graph 2 .Infected cases on trained data(PR)

Graph 3 Deaths on train data(PR)

Table 3 shows the number of confirm cases and death cases for next 10 days.

PR	Next 10days						
Confir	16397	164868	16575		-	 _	17206
М	91	3	99				65
Death	6597	6633	6670				6929

Sl.No	Model	R ² value
	Polynomial	
1	Regression	0.49
2	ARIMA	0.99
3	CNN-LSTM	1

Graph 4. Confirmed cases of covid-19 for ARIMA model

Graph 5 .Death cases of covid-19 for ARIMA model

Table 2. R² values.

Figure 2,3 shows the PR model result. Figure 4,6 shows the ARIMA model for trained data and figure 5,7 shows future prediction of data using ARIMA. figure 8,9

I





Graph 6 .Future forecast of confirmed cases of covid-19 for ARIMA model



Graph 7 .Future forecast of death cases of covid-19 for ARIMA modeL







© 2024, IJSREM | <u>www.ijsrem.com</u>

Graph 9 Death on train data.(CNN-LSTM)







Graph 11 Prediction of confirmed cases(CNN-LSTM)

VI. CONCLUSION

- The accuracy of the polynomial regression model used to predict the confirm and death cases of covid-19 is very low.
- ARIMA Model which gives the accuracy around 0.9494 for predicting the confirm and death cases of covid-19.
- Deep learning model (CNN-LSTM) gives the accuracy of 1 for predicting the confirm and death cases of covid-19.

By comparing the accuracy provided by all 3 models the CNN- LSTM model gives the best fit curve for the given dataset.

REFERENCES

Review Paper for Detection of COVID-19 from Medical Images and/ or Symptoms of Patient using Machine Learning approachesakshay Kumar Siddhu1, Dr. Ashok Kumar2 and Dr. Shakti Kundu3 1 Research Scholar, CCSIT, TMU 2,3Associate Professor, CCSIT, TMU

Eduardo Soares et al., "SARS-cov-2 CT-scan dataset: A

[19]

large dataset of real patients CT scans for SARS-cov-2 identification", *medrxiv*, 2020.

[3]

Maciej M. Kowalik et al., "COVID-19—toward a comprehensive understanding of the [13] disease", *Cardiology Journal*, 2020.

[4]

lex Zhavoronkov et al., "Potential non-covalent SARScov-2 3C-like protease inhibitors designed using generative deep learning approaches and reviewed by [14] human medicinal chemist in virtual reality", 2020.

[5]

. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: [15] Concerns and ways forward," PloS one, vol. 13, no. 3, 2018.

[6] G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne, "Machine learning strategies for time series [16] forecasting," in European business intelligence summer school. Springer, 2012, pp. 62–77.

[7] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Zha, W. Liang, C. Wang, K. Wang et al., "Clinically applicable AI system for accurate diagnosis [17] quantitative measurements and prognosis of covid-19 pneumonia using computed tomography", *Cell*, vol. 181, pp. 1423-1433, 2020.

[8] M. Scudellari, *Hospitals Deploy AI Tools to* [18] *Detect COVID-19 on Chest Scans*, [online] Available: https://spectrum.ieee.org/the-human-

os/biomedical/imaging/hospitals-deploy-ai-toolsdetect-covid19-chest-scans.

[9]H. W. K., COVID-19 Outbreak Predictionusing Machine Learning Algorithm, July 2020, [online]Available:https://towardsdatascience.com/covid-19-outbreak-prediction-using-machine-Iearning-[20]algorithm-ce5641bd55bf.[20]

[10] K, P. J, M. Kasoar et al., "Predicting global patterns of long-term climate change from short-term[21] simulations using machine learning", *npj Clim Atmos Sci*, vol. 3, no. 44, 2020.

[11] S. B. Lajevardi and B. Minaei-Bidgoli, "Combination of Time Series Decision Tree and Clustering : A Case Study in Aerology Event Prediction", *International Conference on Computer and Electrical Engineering*, pp. 111-115, 2008.

[12] S. Mehrmolaei and M. R. Keyvanpour, "A

Brief Survey on Event Prediction Methods in Time Series" in Artificial Intelligence Perspectives and Applications, Springer International Publishing, pp. 235-246, 2015.

K. Chen, K. Chen, Q. Wang, Z. He, J. Hu and J. He, "Short-Term Load Forecasting With Deep Residual Networks", *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943-3952, 2019.

G. Černe, D. Dovžan and I. Škrjanc, "Short-Term Load Forecasting by Separating Daily Profiles and Using a Single Fuzzy Model Across the Entire Domain", *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7406-7415, 2018.

V. Ashish, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is all you need", *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, no. 1, pp. 105-109, 2017.

F. E. Harrell Jr, K. L. Lee, D. B. Matchar, and T. A. Reichert, "Regression models for prognostic prediction: advantages, problems, and suggested solutions." Cancer treatment reports, vol. 69, no. 10, pp. 1071–1077, 1985.

P. Lapuerta, S. P. Azen, and L. LaBree, "Use of neural networks in predicting the risk of ronary artery disease," Computers and Biomedical

Research, vol. 28, no. 1, pp. 38-52, 1995.

K. M. Anderson, P. M. Odell, P. W. Wilson, and W. B. Kannel, "Cardiovascular disease risk

profiles," American heart journal, vol. 121, no. 1, pp. 293–298, 1991.

H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," Procedia Computer Science, vol. 83, pp. 1064–1069, 2016.

F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus covid-19," Plos one, vol. 15, no. 3, p. e0231236, 2020.

Y. Grushka-Cockayne and V. R. R. Jose, Combining prediction intervals in the m4 competition," *Int. J.*

Forecasting, vol. 36, no. 1, pp. 178_185,Jan. 2020W. Bleesser and P. Liicoff, "Predicting stock returns with bayesian vector autoregressive," *Data Analysis*, *Machine Learning and Applications*, vol. 1, pp. 499– 506, 2005.

[23] D. Das, K. Santosh and U. Pal, "Truncated inception net: COVID-19 outbreak screening using

I

chest X-rays", Phys. Eng. Sci. Med, vol. 43, pp. 915-925, 2020.

[24] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim and U.R. Acharya, "Automated detection of COVID-19 cases using deepneural networks with X-ray images", *Comput. Biol. Med*, vol. 121, pp. 103792, 2020.

[25] M. Schmitt, *How to fight COVID-19 with machine*

learning,[online]Available:https://www.datarevenue.co m/en-blog/machine-Iearning-covid-19.

[26] R. Medar, V. S. Rajpurohit and S. Shweta, "Crop Yield Prediction using Machine Learning Techniques", 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019.

[27] J Xin Ge and Yuan Xue, "Machine Learning Technique and Applications – An Classification Analysis", *Journal of Machines and Computing*, vol. 1, no. 4, pp. 185-190, October 2021.

[28] A. Adebiyi, A. Adewumi, and C. Ayo, "Stock price prediction using the ARIMA model," in *Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, IEEE, Cambridge, UK, March

L