POLYNOMIAL BASED LINEAR REGRESSION MODEL TO PREDICT COVID-19 CASES

Nikhil
Student, Faculty of Computer Applications,
Manav Rachna International Institute of Research & Studies,
Faridabad
Nikhil_2019@manavrachna.net

Arushi Saini
Student, Faculty of Computer Applications,
Manav Rachna International Institute of Research & Studies,
Faridabad
Arushisaini61@gmail.com

Santu Panday
Student, Faculty of Computer Applications,
Manav Rachna International Institute of Research & Studies,
Faridabad
Pandeysantu2000@gmail.com

Dr. Neha Gupta
Faculty of Computer Applications,
Manav Rachna International Institute of Research & Studies,
Faridabad
neha.fca@mriu.edu.in

2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) | 978-1-6654-3559-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/RTEICT52294.2021.9574032

***Abstract*— The epidemic COVID-19 has profoundly influenced people's wellness worldwide and the number of fatalities from diseases continues to increase world-wide. Despite technology's remarkable success in our daily lives, notably in ML and DL, AI also helped humanity fight the grueling COVID-19 war. DL is only one approach of ensuring that potential data-driven technologies can help humankind manage COVID-19. Big data and artificial intelligence are used to leverage exceptional efforts to combat the COVID-19 pandemic crisis. In some prior disease outbreaks, various AI offshoots were deployed. AI was applied in the identification of disease clusters, case monitoring, future outbreak predictions, mortality risk, and diagnosis of COVID-19, resource allocation illness management, training facilitation, record maintaining and design identification for the investigation of the trend towards the illness. AI & Machine learning can help to find out the strategies to prevent the Corona virus. This paper presents a polynomial based linear regression model to predict the future cases according to the current situation using data of last few months, showing the output on the graph. The paper also discusses the applications of AI & Machine learning in Corona virus pandemic like forecasting infection rate, diagnose with images comprehensively and will also discuss the role of Machine learning in facilitating the development of vaccine as well.**

Keywords*— *Artificial Intelligence; COVID-19; Machine Learning; SARS-CoV-2, Deep Learning

I. INTRODUCTION

In order to better understand the viral disease patterns and to enhance diagnostic precision, new treatment approaches have been developed to identify the most vulnerable persons based on their specific genetic and physiological property.

Broad data for COVID-19 patients have been merged and analyzed using modern machine-learning algorithms [15]. Following the breakout of COVID-19 and the taxonomic categorization of COVID-19 genomes, the CRISPR-based COVID-19 detecting tests, the prognosis of survival of severe COVID-19 patients and the discovery of possible candidates for COVID-19 drugs were employed inspiringly. Current efforts to create new diagnostic techniques employing machine-learning algorithms have been made. For example, machine learning-based SARS-CoV-2 screening with excellent sensitivity and speed was presented to illustrate a CRISPR-based viral detection system. For large - scale examination of COVID-19 patients based on their individual breathing patterns neural network classifiers were established [14]. Likewise, for the automatic detection and monitoring of COVID-19 patients over time, a data acquisition and data analysis system was designed for chest cell CT images. Machine learning got new success recently in healthcare which made eye-catching headlines. Google has recently made a Machine learning algorithm to accelerate cancerous tumors. To finding out skin cancer Stanford has used a deep learning algorithm (an application of Machine learning). An article by JAMA says Machine Learning algorithms are able to identified diabetic retinopathy in retinal images which has brought revolution in Medical science [9]. Machine learning is not to be trained only to scan the images but also to identify malformations, and point to areas that need attention. Machine learning gives us impartial opinions to improve reliability and accuracy.

II. VARIOUS DEEP LEANING MODELS TO COMPACT CORONA VIRUS

a) Drug Screening for Novel Corona virus using Deep Learning

Deep Learning models can be used to find out the medicines that may help patients having corona virus. Selecting an altered DenseNet (convolution layer) to predict protein-ligand basic interaction can help identify the drugs for treatment [7]. The altered densenet model along with RNAsequence layers of Corona virus simultaneously with chemical compounds can be used to predict the best drug among all the drugs. More research remained to be done but other compounds might help like Adenosine

b) Identifying if an X-ray image or CT scan shows the Corona virus

Detection of a case of Covid-19 related pneumonia from CT scan image can diagnose and can give better treatment. UNet++ based convolutional neural network can be used to get

appropriate features to scan the images and label them. The model has been trained on 40000 scans from 100 patients .

Specifically "The model got a patient to patient sensitivity of 98%, the accuracy of 91.24%, and NPV 98%; image to image responsiveness of 94.34%, specificity of 99.16%, the correctness of 98.85%, and NPV of 99.61% in the retroactive dataset. For 27 potential patients, the model got a high performance [10]. Among the support of the model, radiologists' reading time was significantly decreased by 57%." So, in more scattered words, the model was very successful at increase the workflow of radiologists [13]. This is a great model of the profitable consequence on AI in the actual world. Hence radiologists can look into more CT scan images in the same or even less number of time. It is necessary to note, nevertheless, unusual records were to be on a test set of only 27 inmates for expected testing and 44 sufferers for automotive testing. Moreover, each of the inmates was from the same hospital and it proved that the model learned something, but it is not sure that how powerful this would be on a large population [11].

c) AI-based Cough Sound Analysis for COVID-19

Coughing symptoms include over 30 non-COVID19 medical disorders. This makes COVID-19 coughing alone an important challenge for several difficulties. Doctors also employ sound signals from human bodies. Schuller [20] and other scholars have studied the beneficial purpose of CA (Computer Audition) and AI in the COVID-19 research of cough noises. The ability for CA to automatically identify and cough speech was initially evaluated under various settings, such as breathing, dry cough or wet cough, snowing, fluid speaking, eating habits, tiredness or discomfort. They then advocated the diagnosis and cure of COVID-19 patients utilizing CA technology [12].

III. Plynomial Based Deep Learning Model to Predict the Upcoming Cases According To The Current Situation

In this paper polynomial based linear regression model has been proposed to predict and forecast the upcoming cases for the specified days. Real time dataset has been used which records the Confirmed cases, Deaths, Recover cases of two and half months. To overcome under-fitting, the model has decreased the complexity of features. To produce an essential equation, powers have been added to the original features to create polynomial features.The proposed model is a linear model which can transform original features into polynomial features.

Dataset used: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide

To convert the original features into higher polynomial, the model has used the Polynomial Feature Class & linear regression class implemented in Scikit-Learn. We have used the linear Regression model for its simplicity in forecasting upcoming cases.

datasets which is available on keggle and the model performed well [8].

a. Converting original features into polynomial features

```
polynomial_features= PolynomialFeatures(degree=2)
x_poly = polynomial_features.fit_transform(x)
```

b. Making object of Linear Regression and fitting the data with fit function to check the accuracy of the model:

```
#model for Confirmed Cases and its score
cls_confirmed_cases = LinearRegression(normalize=True)
cls_confirmed_cases.fit(confirmed_ploy_X_train, confirmedcase_y_train.ravel())
cls_confirmed_cases.score(confirmed_ploy_X_train, confirmedcase_y_train.ravel(
```

```
99.4834100350481
```

c. Predicting the upcoming cases for next 25 days:

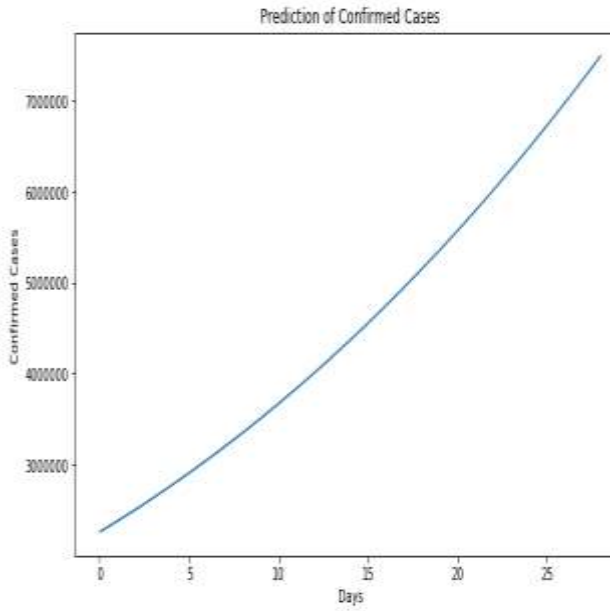

Figure 1: Graph predicting upcoming cases

d. The model has successfully tested the data and have given accuracy of 98.29%

```
cls_confirmed_cases.score(confirmed_ploy_X_train, c
```

```
98.29522354979873
```

e. Predicting the Deaths and Confirmed Cases for upcoming 20 days

```
#prediction of Next 20 days
future_days_deaths = ploy.fit_transform(next_days_prediction)
future_days_Confirmed_cases = ploy.fit_transform(next_days_prediction)
future_days_deaths_prediction = cls.predict(future_days_deaths)
future_days_Confirmed_cases_prediction = cls_confirmed_cases.predict(future_days_Confirm
```

```
#Next 20 day of Deaths and Confirmed Cases DataFrame
deaths = future_days_deaths_prediction[len(datasetdays):].tolist()
confirmed_Cases = future_days_Confirmed_cases_prediction[len(datasetdays):].tolist()
deaths = [round(x) for x in deaths]
confirmed_Cases = [round(x) for x in confirmed_Cases]
date = future_dates[len(datasetdays):]
df = {"Date": date,"Deaths": deaths,"Confirmed Cases": confirmed_Cases}
predicted_dataframe = pd.DataFrame(df)
```

The model has predicted the cases of January 2021 which have been verified with the actual data available on Wordometer.com and the accuracy has been identified as 99.29%

	Date	Deaths	Confirmed Cases
231	2021-01-01	29436294	382859478
232	2021-01-02	29716747	386482156
233	2021-01-03	29998974	390127598
234	2021-01-04	30282979	393795877
235	2021-01-05	30568769	397487063
236	2021-01-06	30856349	401201227
237	2021-01-07	31145724	404938441
238	2021-01-08	31436901	408698777
239	2021-01-09	31729884	412482304
240	2021-01-10	32024680	416289096
241	2021-01-11	32321293	420118221
242	2021-01-12	32619730	423972754
243	2021-01-13	32919996	427849763
244	2021-01-14	33222097	431750321
245	2021-01-15	33526038	435674498
246	2021-01-16	33831825	439622367
247	2021-01-17	34139463	443593996
248	2021-01-18	34448958	447589462
249	2021-01-19	34760316	451608831

The proposed model has predicted the upcoming cases of Covid-19 based on given datasets and has predicted the cases more accurately than other Deep Learning models. Although Deep Learning models can also provide accurate data but they keep tweaking the values of the parameters which may lead to overfit and underfit problem.

The proposed model has first converted the original features into polynomial features and then tested the features using Polynomial based linear regression. To know the necessity for polynomial regression, we generated some random datasets and plotted them on a graph with a best-fit line which will look like

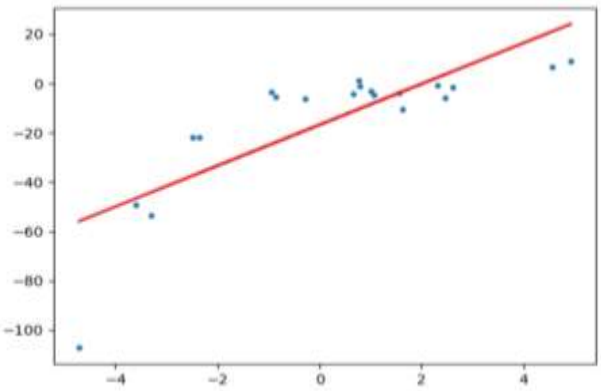

Figure 2: Plotting dataset on a graph to check best fit

If the data points will not fit a linear regression it might cause the under-fitting problem and to overcome this we need to increase the complexity of the model, which is possible using polynomial regression.

We used Polynomial Features from Scikit-learn Class. For degrees 3 to 20, the model captures noise. This is an example of over-fitting and to overcome form overfitting, we can easily add more training datasets so the algorithm does not learn the noise and can become more accurate. Deep learning models take more time to produce good accuracy and even more time to predict the data on a low spec system but our model gives high accuracy and less time to produce accurate result. The model is yet acknowledged as a linear model as the coefficients are still linearly linked with the features. However, the curve is quadratic in nature.

IV. ML AND DL FOR DRUG AND VACCINE DEVELOPMENT

The ability to automatically study the abstract components has a huge influence in effective usage of ML in conjunction with enormous amounts of information [1]. Drug development, vaccination and interactions with ligand proteins influenced two important sectors where ML gave integrated predictions, behavior, and reaction. The focus on proteomics and genomes research has been recommended for mDiverse medication development and vaccination programs for SARS-CoV-2 and COVID-19. [2] The application of ML and DL in developing novel medicinal products and vaccines is one of the important contributions to intelligent medicine, which plays a vital role in the fight against COVID-19 [4]. To analyze how it works, it is necessary to understand the composition of proteins. Researchers can create medicines that act in different protein forms once the scenario is understood. However, it takes quite a long time to examine every protein structure until it finds its unique 3D structure. AI systems based on DL can be used to simplify the process for measuring the structure and genetic sequence of protein. In January Google released AlphaFold, a state-of-the-art algorithm that predict how 3D protein would form when its genetic code is being applied. The technology was tested at COVID-19 in early March [3].

A new technique is presently implemented by several other researchers [5] of MIT's Computer Science and AI Lab (CSAIL), using an improved ML-based algorithm to choose peptides (small amino acid fibre) that are likely to deliver big vaccination quantities. The design programme OptiVax combines new methods in which peptide medications may be

designed, current vaccinations evaluated and existing vaccinations made more compact. Peptides get a mechanical education in this programme for their capacity to speak in response to antibodies and are chosen to extend their coverage to people who benefit from this vaccination [6].

CONCLUSION

The paper reviewed that machine learning can be more auxiliary in the field of healthcare to prevent asunder diseases based on data analysis. With the help of analyzing the vast amounts and the various forms of data, recorded by patient and healthcare institutions, Machine learning can be very supportive for the future needs of medicines. Deep learning which the part of Machine learning is facilitates the development of vaccine. Deep learning has sharp capability to ferret out the Corona virus or any other diseases by just an X-ray, It may remove the need of any other equipment that has been using for just identifying the disease. Many Machine learning communities are still working and recognizing this important trend in modern medicine. They are strengthening their emerging algorithms and technology. The data will increase and making an accurate forecasting model will be more complex for data and machine learning scientists. The paper has presented a polynomial based linear regression model to predict the future and upcoming cases and has shown better outcomes in contrast to other models.

REFERENCES

[1] Banner, N. (2018). Deep Residual Learning for Image Recognition. *Bioethics briefing note*, 1-8.

[2] H, W. K. (2020, July 16). *COVID-19 Outbreak Prediction using Machine Learning Algorithm*. Retrieved August 18, 2020, from Towards data science: https://towardsdatascience.com/covid-19-outbreak-prediction-using-machine-learning-algorithm-ce5641bd55bf

[3] He, K., Zhang, X., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *CVF*, 770-778.

[4] Jiang, F., & Jiang, Y. (2017). Artificial intelligence in healthcare: past, present, and future. *Open Access*, 1-14.

[5] Karpathy, A., & Toderici, G. (2014). Large-scale Video Classification with Convolutional Neural Networks. *CVF*, 1-8.

[6] Schmitt, M. (2020, June 2). *How to fight COVID-19 with machine learning*. Retrieved August 15, 2020, from Datarevenue : https://www.datarevenue.com/en-blog/machine-learning-covid-19

[7] Basu, S.; Mitra, S. Deep Learning for Screening COVID-19 using Chest X-ray Images.arXiv2020, arXiv:2004.10507.16.

[8] Scudellari, M. Hospitals Deploy AI Tools to Detect COVID-19 on Chest Scans. 2020. Available online: https://spectrum.ieee.org/the-human-os/biomedical/imaging/hospitals-deploy-ai-tools-detect-covid19-chest-scans (accessed on 10 September 2020).

[9] BBVA. Artificial Intelligence to detect COVID-19 in Less than a Second Using X-rays. 2019. Available online: https://www.bbva.com/en/artificial-intelligence-to-detect-covid-19-in-less-than-a-second-using-x-rays/

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

[11] Mangal, A.; Kalia, S.; Rajgopal, H.; Rangarajan, K.; Namboodiri, V.; Banerjee, S.; Arora, C. CovidAID: COVID-19 Detection UsingChest X-ray.arXiv2020, arXiv:2004.09803.

[12] Das, D.; Santosh, K.; Pal, U. Truncated inception net: COVID-19 outbreak screening using chest X-rays.Phys. Eng. Sci. Med.2020,43, 915–925.

[13] Minaee, S.; Kafieh, R.; Sonka, M.; Yazdani, S.; Soufi, G.J. Deep-covid: Predicting covid-19 from chest X-ray images using deeptransfer learning.arXiv2020, arXiv:2004.09363.

[14] Yoo, S.H.; Geng, H.; Chiu, T.L.; Yu, S.K.; Cho, D.C.; Heo, J.; Choi, M.S.; Choi, I.H.; Cung Van, C.; Nhung, N.V.; et al. Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging.Front. Med.2020,7, 427.

[15] Zhang, K.; Liu, X.; Shen, J.; Li, Z.; Sang, Y.; Wu, X.; Zha, Y.; Liang, W.; Wang, C.; Wang, K.; et al. Clinically applicable AI systemfor accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography.Cell2020,181, 1423–1433.e11.

[16] VIDA. LungPrint Clinical Solutions. 2020. Available online: https://vidalung.ai/clinical_solutions/

[17] Analytics, H.I. Deep Learning Models Can Detect COVID-19 in Chest CT Scans. 2020. Available online: https://healthitanalytics.com/news/deep-learning-models-can-detect-covid-19-in-chest-ct-scans

[18] Chen, J.; Wu, L.; Zhang, J.; Zhang, L.; Gong, D.; Zhao, Y.; Hu, S.; Wang, Y.; Hu, X.; Zheng, B.; et al. Deep learning-based model fordetecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: A prospective study.medRxiv2020.

[19] Huang, L.; Han, R.; Ai, T.; Yu, P.; Kang, H.; Tao, Q.; Xia, L. Serial quantitative chest ct assessment of covid-19: Deep-learningapproach.Radiol. Cardiothorac. Imaging2020,2, e200075.

[20] Schuller, B.W.; Schuller, D.M.; Qian, K.; Liu, J.; Zheng, H.; Li, X. Covid-19 and computer audition: An overview on what speech & sound analysis could contribute in the SARS-CoV-2 Corona crisis. arXiv 2020, arXiv:2003.11117

[21] Rajendra Prasad P. and Shiva Shankar, "Securing Energy Routing Protocol Against Black Hole Attacks in Mobile Ad-Hoc Network", Recent Advances in Computer Science and Communications, ISSN: 2666-2558(print), ISSN: 2666-2566(online), 14: 2835. https://doi.org/10.2174/2666255813666200814171401, 2021.

[22] Arjun Kumar G B and Dr. Shivashankar, "Optimal power point tracking of solar and wind energy in a hybrid wind solar energy system", International Journal of Energy and Environmental Engineering, Springer Nature Publication, No. 3, pp. 1-27, DOI: https://doi.org/10.1007/s40095-021-00399-9, 2021