International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 11 Issue 2, February, 2022, Impact Factor: 7.751

AI based Machine Learning Model for COVID Data Analysis

Basavaraj Chunchure

Vignan's Institute of Management and Technology for Women's Hyderabad, India

ABSTRACT

At present Scenario data science and digital image processing are essential technologies used in many health care applications for quick, accurate detection and analysis of patient's big data. Statistical analysis in data science was useful tool to diagnose quickly and give proper treatment for covid disease effectively and efficiently. Artificial Intelligent based Machine learning techniques efficiently monitoring the cases who take proper treatment and vaccination based on gender and age. Analysis reports are obtained accurately such as diseases spread through community contact and recovered, Cases recovered and not hospitalized, not hospitalized and recovered etc. Big Data analytics assist the early recognition of COVID-19 through the investigate significant characteristics that permit the treatment to classify the factors that facilitate the early detection of the infection.

Keywords: Analysis, Artificial Intelligence, Detection, Diagnosing, machine learning

INTRODUCTION

AI-based algorithm using CT scan images to detect CoVID-19 in such a way to help doctors to diagnose CoVID-19 patients and help them decide what to do next depending on the output of the algorithm, help automate the diagnosis of patients to help doctors to know severe or not, decide how to proceed for patients, free up doctors time as the algorithm will automate a process that can be very time consuming.AI based covid detection using machine learning helpful for radiological diagnostics, prognostics' based on clinical data, pharmaceutical discovery, test kit development, virus function & disease progression, identification of potential drugs and methods [Fig-1]. The versatility of artificial intelligence (AI) has surged up the momentum to implement the technique [1] [2] for medical and societal adversity in the COVID-19 epidemic [3] [4] [5].

Machine learning technique used for diagnosing, calculating, forecasting and examine, evaluation. clinical performance medical AI-based approaches [6–10] can be implemented using machine learning (ML) which can be further subdivided into deep learning, artificial neural net- work (ANN), fuzzy logics, and reinforcement learning. In addition, algorithms like sup- port vector regression for predicting the spread and analysing the growth/transmission rate [11–14], random forest machine learning model for anticipating compound growth rate [15–23] with respect to social distancing stringency and as a discrimination tool for early screening [17–19] have contributed towards gaining an improved understanding of the potential risk factors. Regression models are used for COVID are

- Least Absolute Shrinkage and Selection Operation (LASSO)
- Random forest
- Decision tree regressor
- Linear regression
- Support vector machine
- Polynomial regression

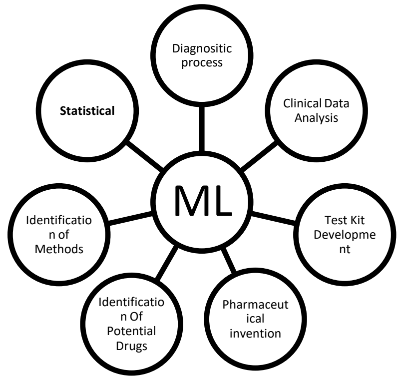

Fig.1: **ML in Covid Care Process**

DETECTION USING ML

Machine Learning algorithms such as Extreme Learning Machine, Support Vector Ma- chine, Decision Tree, Random Forest, K Nearest Neighbor, and Probabilistic Neural Network and deep learning methods Convolution Neural Network (CNN), CNN with transfer learning, Residual CNN (RNN)). The challenging issue in getting optimal per- formance in machine learning algorithms is the design of an appropriate model for clas- sifying or predicting unknown samples into specific groups based on training input. In this case, network parameters of different learning algorithms are playing a significant role in achieving optimal performance in testing the unknown data [Fig.2].

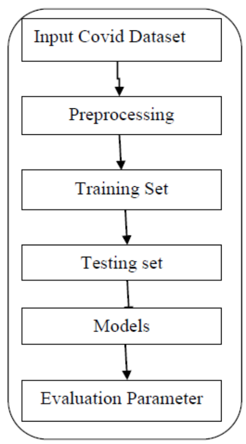

Fig.2 ML Model

STATISTICAL ANALYSIS

The features are validated through statistical approaches such as one analysis of vari- ance (ANOVA) with repeated measures used to identify the significance of each feature (pixel) in distinguishing CoVID-19 and normal. Chi-Square test is used to compare the features of each lung segment between CoVID-19 and healthy groups, and Wilcoxon rank test is used to compare the differences of the left lung, right lung, and total score between CoVID-19 and normal. The above statistical tests are the most powerful tools in data analysis to validate the importance of extracted features in clinical diagnosis. The following parameters are analyzed based on Canadian covid 19 database and re- sults are plotted using graph [Fig.3].

- Gender, Age Analysis.
- Transmitted through community exposures and recovered.
- Cases recovered and not hospitalized.
- Not hospitalized.
- Not hospitalized and recovered

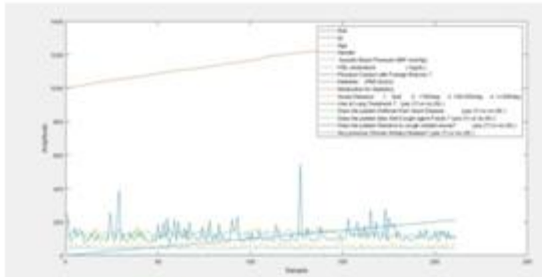

Fig.3: Patient Statistical Analysis

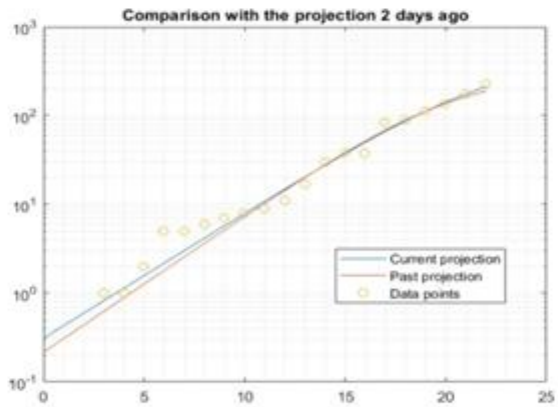

Fig. 4. Active cases analysis

Patient status is taken from the dataset from Covid india dataset. First the data is given as input to preprocessing stage, and then given to training and testing sets. Impor- tant measures are R-squared score, MSE, MAE, and RMSE was used [Table.1].

Table 1: Cases calculation from Data sets

Frequency Pattern	Frequency	
	Absolute	**Relative**
Community Exposure	169,000	23.21%
Recovered	134,000	21.05%
Not Hospitalized	115,800	15.90%
Not Hospitalized, Not Recovered	163,000	22.39%
Community Exposure, Recovered	58,000	7.97%
Recovered ,Not Hospitalized	88,300	12.13%
Covid Dataset	728100	100%

PERFORMANCE MEASURES

The performance of ML are analyzed based on R squared score, Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)

Squared score

Regression model is denoted as

R^2 =variance of model / total variance **….** (1)

Mean Square Error (MSE)

It calculates response time of the error and average square difference of the predicted values and real values.

$$MSE = \frac{1}{m}\sum_{j=1}^{m}(y_j - y_{j}.)^2$$

… (2)

Mean Absolute Error (MAE):

$$MAE = \frac{1}{m}\sum_{i}^{m}|y_i - y_i \wedge|$$

…. (3)

It is the statistics is a calculation of errors that reflect a certain performance

Root Mean Square Error (RMSE)

Root mean square error is always used for prediction, and linear regressions to verify the research results.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p-a)^2}$$

… .(4)

Below [Table 2] results explains, the different methods of evaluation to calculate and estimate the covid condition and find out the statistical analysis accurately and quickly, In ML a sequence of regressions, models used such as linear regression, SVM, RF, polynomial regression, multi-regression, and Lasso regression models.

Table 2: Performance Measure Comparison

Models	R^2	MSE	MAE	RMSE
Lasso	86.29	1839.5	52.2	127.12
Decision tree	83.54	51.24	55.12	39.25
Random forest	89.54	83.24	95.35	389.12
SVM	11.24	2881.23	86.35	151.23
PR	87.52	2143.11	83.27	139.52
LR	85.43	2141.12	84.29	138.15

CONCLUSION

In this paper, Machine learning model of big data was analysed for COVID-19 crisis. An experimental outcome shows that ML model provides rich knowledge about charac- teristics of COVID-19 cases. Artificial Intelligence based machine learning technique helps the doctors for their medical occupation, serving them to get better precision of the analysis in a lesser time and make assessment faster effectively and efficiently.

REFERENCES

[1]. Wong, Z.S.Y., Zhou, J., Zhang, Q.: Artificial Intelligence for Infectious Disease Big Data Analytics". Infect. Dis. Health **24**, 44–48 (2019)

[2]. Chiang, L., Lu, B., Castillo, I.: Big Data Analytics in Chemical Engineering. An- nual Review of Chemical and Biomolecular Engineering **8**(1), 63–85 (2017), 10.1146/annurev-chembioeng-060816-101555;https://dx.doi.org/10.1146/annurev- chembioeng-060816-101555

[3]. Kourti, T.: Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. Journal of Chemometrics **17**(1), 93–109 (2003), 10.1002/cem.778;https://dx.doi.org/10.1002/cem.778
[4]. Qin, S.J.: Survey on Data-Driven Industrial Process Monitoring and Diagnosis". Annu. Rev. Control **36**(2), 220–234 (2012)
[5]. Sliwoski, G., Kothiwale, S., Meiler, J., Lowe, E.W.: Computational Methods in Drug Dis- covery. Pharmacological Reviews **66**(1), 334–395 (2014), 10.1124/pr.112.007336;https://dx. doi.org/10.1124/pr.112.007336
[6]. Klipp, E., Wade, R.C., Kummer, U.: Biochemical network-based drug-target prediction. Cur- rent Opinion in Biotechnology **21**(4), 511–516 (2010), 10.1016/j.copbio.2010.05.004;https:
[7]. //dx.doi.org/10.1016/j.copbio.2010.05.004
[8]. Bragazzi, N.L., Dai, H., Damiani, G., Behzadifar, M., Martini, M., Wu, J.: How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. International Journal of Environmental Research and Public Health **17**(9), 3176–3176 (2020), 10.3390/ ijerph17093176;https://dx.doi.org/10.3390/ijerph17093176
[9]. Swapnarekha, H., Behera, H.S., Nayak, J., Naik, B.: Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. Chaos, Solitons & Fractals **138**, 109947–109947 (2020), 10.1016/j.chaos.2020.109947;https://dx.doi.org/10.1016/j.chaos. 2020.109947
[10]. Castro, R., Luz, P.M., Wakimoto, M.D., Veloso, V.G., Grinsztejn, B., Perazzo, H.: COVID- 19: a meta-analysis of diagnostic test accuracy of commercial assays registered in Brazil. The Brazilian Journal of Infectious Diseases **24**(2), 180–187 (2020), 10.1016/j.bjid.2020.04. 003;https://dx.doi.org/10.1016/j.bjid.2020.04.003
[11]. Banerjee, A., Pasea, L., Harris, S., Gonzalez-Izquierdo, A., Torralbo, A., Shallcross, L., Noursadeghi, M., Pillay, D., Sebire, N., Holmes, C., Pagel, C., Wong, W.K., Langenberg, C., Williams, B., Denaxas, S., Hemingway, H.: Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population- based cohort study. The Lancet **395**(10238), 1715–1725 (2020), 10.1016/s0140-6736(20) 30854-0;https://dx.doi.org/10.1016/s0140-6736(20)30854-0
[12]. Barman, M.P., Rahman, T., Bora, K., Borgohain, C.: COVID-19 pandemic and its recovery time of patients in India: A pilot study. Diabetes & Metabolic Syndrome: Clinical Research & Reviews **14**(5), 1205–1211 (2020), 10.1016/j.dsx.2020.07.004;https://dx.doi.org/10.1016/ j.dsx.2020.07.004
[13]. Menebo, M.: Livadiotis G., "Statistical Analysis of the Impact of Environmental Temperature on the Exponential Growth Rate of Cases Infected by COVID-19". Norway", Sci. Total Environ **737**(5), 233875–233875 (2020)
[14]. Kumar, A., Rani, P., Kumar, R., Sharma, V., Purohit, S.R.: Data-driven modelling and pre- diction of COVID-19 infection in India and correlation analysis of the virus transmission with socio-economic factors. Diabetes & Metabolic Syndrome: Clinical Research & Reviews **14**(5), 1231–1240 (2020), 10.1016/j.dsx.2020.07.008;https://dx.doi.org/10.1016/j.dsx.2020. 07.008
[15]. Chatterjee, A., Gerdes, M.W., Martinez, S.G.: Statistical Explorations and Univariate Time- series Analysis on COVID-19 Datasets to Understand the Trend of Disease Spreading and Death. Sensors **20**(11), 3089–3089 (2020), 10.3390/s20113089;https://dx.doi.org/10.3390/ s20113089
[16]. Kanga, S., Sudhanshu, Meraj, G., Farooq, M., Nathawat, M.S., Singh, S.K.: Report- ing the management of COVID-19 threat in India using remote sensing and GIS based approach (2020), 10.1080/10106049.2020.1778106;https://dx.doi.org/10.1080/10106049. 2020.1778106
[17]. Zarikas, V., Poulopoulos, S.G., Gareiou, Z., Zervas, E.: Clustering analysis of countries using the COVID-19 cases dataset (2020), 10.1016/j.dib.2020.105787;https://dx.doi.org/10.1016/ j.dib.2020.105787
[18]. Wang, P., Lu, J., Jin, Y., Zhu, M., Wang, L., Chen, S.: Statistical and Network Analysis of 1212 COVID-19 Patients in Henan. Int. J. Infect. Dis **95**, 391–398 (2020)
[19]. Neil, M., Fenton, N., Osman, M., Mclachlan, S.: Bayesian Network Analysis of COVID- 19 Data Reveals Higher Infection Prevalence Rates and Lower Fatality Rates Than Widely Reported. J. Risk. Res **0**(0), 1–14 (2020)
[20]. Alsulaiman, M., Alotaibi, Y., Ghulam, M., Bencherif, M.A., Mahmood, A.: Arabic speaker recognition: Babylon Levantine subset case study. Journal of Computer Science **6**, 381–385 (2010)
[21]. Alsulaiman, M., Ghulam, M., Bencherif, M.A., Mahmood, A., Ali, Z.: KSU rich Arabic speech database. Information Journal **16**, 4231–4254 (2013)
[22]. Alt_Ncay, H., Demirekler, M.: Why does output normalization create problems in multiple classifier systems? In: Proceedings of CPR2002, 16th International Conference on Pattern Recognition (2002)
[23]. Anusuya, M.A., Katti, S.K.: Front end analysis of speech recognition: a review. International Journal of Speech Technology **14**(2), 99–145 (2011), 10.1007/s10772-010-9088-7;https:// dx.doi.org/10.1007/s10772-010-9088-7