



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VIII Month of publication: August 2021

DOI: <https://doi.org/10.22214/ijraset.2021.37740>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Review Paper on Machine learning-based prediction of Corona Virus

Miss. Aakansha P. Tiwari¹, Prof. S. P. Akarte², Prof. S. V. Deshmukh³

¹PG Scholar, ²Assistant Professor, ³Assistant Professor, Computer Science & Engineering, Prof.Ram Meghe Institute of Technology & Research, Maharashtra, INDIA

Abstract: *Effective contact tracing of SARS-CoV-2 enables quick and efficient diagnosis of COVID-19 and might mitigate the burden on healthcare system. Prediction models that combine several features to approximate the danger of infection are developed. These aim to help medical examiners worldwide in treatment of patients, especially within the context of limited healthcare resources. They established a machine learning approach that trained on records from 51,831 tested individuals (of whom 4769 were confirmed to own COVID-19 coronavirus). Test set contained data from the upcoming week (47,401 tested individuals of whom 3624 were confirmed to own COVID-19 disease). Their model predicted COVID-19 test results with highest accuracy using only eight binary features: sex, age ≥ 60 years, known contact with infected patients, and also the appearance of 5 initial clinical symptoms appeared. Generally, supported the nationwide data publicly reported by the Israeli Ministry of Health, they developed a model that detects COVID-19 cases by simple features accessed by asking basic inquiries to the affected patient. Their framework may be used, among other considerations, to prioritize testing for COVID-19 when testing resources are limited and important.*

Keywords: *Machine Learning, SARS-COV-2, COVID-19, Coronavirus.*

I. INTRODUCTION

The novel coronavirus disease- 2019 (COVID-19) pandemic situation caused by the SARS-CoV-2 virus continues to pose a critical and urgent threat to global health. The outbreak in early December 2019 within the Hubei province of the People's Republic of China has spread worldwide. As of October 2020, the final number of patients confirmed to possess the disease has exceeded 39,500,000, in >180 countries, though the number of people infected due to covid-19 is probably much higher. Over 4.2 million people have died from COVID-19. Since the disease can cause multi-organ sequelae and death, COVID-19 patients who have a poor immunity power and are likely to deteriorate to critical status need to be identified promptly. Additionally, it's difficult for medical systems to need within the high prevalence of critically ill patients with COVID-19. This disease has been detrimental to medical resource availability. [1]

Effective screening enables quick and efficient diagnosis of COVID-19 patient and should manipulate the burden on our healthcare systems. Prediction models that combine several features to estimate the danger of infection are developed rapidly, within the hope of assisting medical staff worldwide in triaging patients, especially within the context of limited healthcare resources available. These models use features like computer tomography (CT) scan, clinical symptoms, laboratory tests, and so the combination of these features. However, most previous models were supported data of hospitalized infected patients, thus don't seem to be effective in touch tracing for SARS-CoV-2 within the nice amount population. [2]

During this paper, they proposed that a machine-learning model that predicts a positive SARS-CoV-2 infection in a very RT-PCR test by asking eight basic questions. This model was trained on data of all individuals tested for SARS-CoV-2 during the first months of the COVID-19 pandemic situation. Thus, our model could also be implemented globally for effective screening and prioritization of testing for the virus within the overall population. within the study, a machine learning (ML) pipeline was developed to predict COVID-19 patient severity and so the danger of progression to critical illness within specific time intervals using chest CT and clinical data of patient.

II. MATERIALS AND METHODS

A. Clinical Information

Patient data on demographics and co-morbidities were retrospectively collected. The patients' conditions were also determined to be critical or severe if they reached any of the subsequent endpoint: ICU admission, mechanical ventilation, or death of patient. If not, their conditions were noncritical or non-severe, non-symptomatic.[4] For critical or severe patients, the duration for his or her progression to critical events was calculated from the time of CT to the earliest time of developing one in the entire aforementioned critical event.

The patient data, including age, sex, symptoms (presence or absence of fever), white corpuscle count, lymphocyte count, comorbidity status (cardiovascular disease, hypertension, chronic obstructive pulmonary disease, diabetes, chronic disease, chronic renal disorder, cancer, and human immunodeficiency virus), and history of exposure to the COVID-19 or another patient with COVID-19, were collected simultaneously. These were used because the 15 clinical variables for model training. The employment of mechanical ventilation, ICU care, and progression to death was also recorded frequently.

B. Machine Learning Pipeline

The lung tissues and abnormalities caused by COVID-19 were automatically segmented on CT images employing a deep-learning model supported a deep convolutional neural network. Second, radiomics features were extracted from CT, and an ML pipeline utilizing both image-based and clinical variables were accustomed predict a patient’s COVID-19 severity and progression to critical illness at the time of the CT scan. [5]

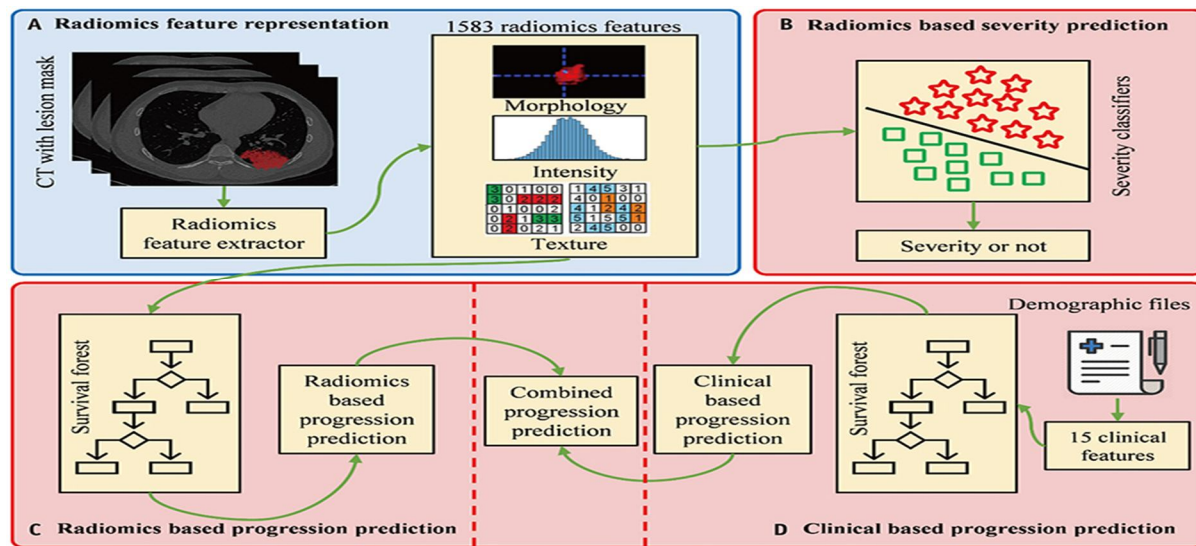


Fig.1. Illustration of our analysis pipeline.

C. Progression Prediction

Specifically, these progression prediction models were supported survival forests that were optimized to assign risk scores to infected patients with different progression outcomes in line with their input features (radiomics features or clinical variables). Missing values of some clinical variables were imputed by a widely used imputation method. Survival forests use a set of decision trees for predictions and also the ranking of radiomics features or clinical variables by their importance for time-to-event risk prediction. Both survival forests were trained and validated on the identical training, validation, and test sets in an exceedingly 7:1:2 split of patient data that were used for the radiomics models.

III. METHODS

A. Setting and Study Data

The Ministry of Health publicly released data of people who were tested for SARS-CoV-2 via RT-PCR assay of a nasopharyngeal swab. The dataset contains initial records, on a everyday, of all the residents who were tested for COVID-19 nationwide. Additionally to the test date and result, various - information is accessible, including clinical symptoms, sex and a binary indication on whether the tested individual is aged 60 years or above. Supported these data, we developed a model that predicts COVID-19 test results using eight binary features: sex, age 60 years or above, known contact with an infected individual, and five initial clinical symptoms. [1] The training-validation set consisted of records from 51,831 tested individuals (of whom 4769 were confirmed to possess COVID-19), from the period March 22th, 2020 through March 31st, 2020. The test set contained data from the following week, April 1st through April 7th (47,401 tested individuals, of whom 3624 were confirmed to possess COVID-19). The training-validation set was further divided to training and validation sets at A ratio of 4:1 (Table 1). The following list describes each of the dataset’s features employed by the model:

- 1) *Basic Information*
 - a) Sex (male/female).
 - b) Age ≥ 60 years (true/false)
- 2) *Symptoms*
 - a) Cough (true/false).
 - b) Fever (true/false).
 - c) Raw throat (true/false).
 - d) Shortness of breath (true/false).
 - e) Headache (true/false).
- 3) *Other Information.*

(#) Feature	Total n = 99,232		COVID-19 n = 90,839		negative COVID-19 positive n = 8393	
	n	%	n	%	n	%
(1) Sex						
Male	50,350	50.74	45,545	50.1	4805	57.2
Female	48,882	49.26	45,294	49.8	3588	42.7
(2) Age 60+						
True	15,279	15.4	13,619	14.9	1660	19.7
False	83,953	84.6	77,220	85	6733	80.2
(3) Cough						
True	14,768	14.88	10,715	11.8	4053	48.2
False	84,223	84.87	79,909	87.9	4314	51.4
(4) Fever						
True	8122	8.18	4387	4.83	3735	44.5
False	90,868	91.5	86,237	94.9	4631	55.1
(5) Sore throat						
True	1273	1.28	96	0.11	1177	14
False	95,062	95.8	88,059	96.9	7003	83.4
(6) Shortness of breath						
True	930	0.94	71	0.08	859	10.2
False	95,405	96.14	88,084	96.9	7321	87.2
(7) Headache						
True	1799	1.81	68	0.07	1731	20.6

Fig. 2: Comparison of Patient Characteristics across the Training, Validation, and Test Sets Training Set

IV. LITERATURE REVIEW

A. *Paper Title: Machine learning-based prediction of COVID-19 diagnosis based on symptoms*

Author Name: Yazeed Zoabi, Shira Deri-Rozov and Noam Shomron

In this paper, they proposed a machine-learning model that predicts a positive SARS-CoV-2 infection during a RT-PCR test by asking eight basic questions. This model was trained on data of all individuals in Israel tested for SARS-CoV-2 during the primary months of the COVID-19 pandemic situation. Thus, their models are often implemented globally for effective screening and prioritization of testing for the virus within the global population.

This research isn't without shortcomings. They relied on the info reported by the Israeli Ministry of Health, which has limitations, biases and missing information regarding a number of the feature. For instance, for patients labelled as having contact with an individual confirmed to own COVID-19 disease, additional information like the duration and placement (indoors/outdoors) of the contact wasn't available to infected patient. A number of the symptoms like lack of smell and taste were identified as being very predictive of a COVID-19 infection by previous studies, but weren't recorded by the Israeli Ministry of Health. They showed that training and testing a model while filtering out symptoms of high bias prior to still achieved very high accuracy. They also note that everyone the symptoms were reported and a negative value for a signal might mean that the symptom wasn't reported. It's therefore important to assess the model's performance within the circumstance that more values are unreported or missing instead of with negative values too. To simulate a less biased condition, in our prospective test set, they randomly selected negative reports of all five symptoms at a time, and removed the negative values also. [1]

B. Paper Title: Machine Learning-Based Prediction of COVID-19 Severity and Progression to Critical Illness Using CT Imaging and Clinical Data

Author Name: Subhanik Purkayastha, Yanhe Xiao², Zhicheng Jiao, Rujapa Thepumnoeysuk, Kasey Halsey, Jing Wu, Thi My Linh Tran, Ben Hsieh, Ji Whae Choi, Dongcui Wang, Martin Vallières, Robin Wang, Scott Collins, Xue Feng, Michael Feldman, Paul J. Zhang, Michael Atalay, Ronnie Sebros, Li Yang, Yong Fan, Wei-hua Liao, Harrison X. Bai.

In this paper, a machine learning (ML) pipeline was developed to predict COVID-19 disease severity and therefore the risk of progression to critical illness within specific time intervals using chest CT and clinical data. This study has several limitations. First, this was a retrospective study with patient selection bias. Data heterogeneity may have also affected the model performance. Further, the present study defined critical outcomes as mechanical ventilation, admission to the ICU, and death, whereas other studies may have different definitions which will account for the various overall mortality rates for his or her cohorts. Considering the critical outcomes of mechanical ventilation, ICU admission, and death as separate events, rather than a composite category, may be beneficial but it requires a bigger sample with sufficient statistical power.[6] ML model supported radiomics features obtained from chest CT and clinical variables predicted COVID-19 severity and progression to critical events with good accuracy. The model supported the mixture of chest CT data and clinical variables also showed higher performance than the model supported only clinical variables, and similar performance to the model supported the mixture of the visual CT severity scores and clinical variables. Further research and development are needed to see the sensible role ML can play in COVID-19 severity predictions within the clinical setting. [2]

V. ADVANTAGE OF MACHINE LEARNING

A. Easily Identifies Trends and Patterns

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviours and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

B. No Human Intervention Needed

With ML, you don't need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus software's; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

C. Continuous Improvement

As ML algorithms gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

VI. CONCLUSION

In this paper, we have summarized different machine learning-based approaches and the current situation of spreading Coronavirus. According to the results and studies reviewed, it is indicated that using convolutional neural networks, which is the deep learning-based technique, might have significant effects on automatic tools in terms of detection, distinguishing, and extraction of essential features automatically from the CT scan images that are related to Coronavirus diagnosis.

VII. ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my guide Prof.S.P.Akarte and Co-guide Prof. S. V. Deshmukh who has in the literal sense, guided and supervised me. I am indebted with a deep sense of gratitude for the constant inspiration and valuable guidance throughout the work.

REFERENCES

- [1] Yazeed Zoabi, Shira Deri-Rozov and Noam Shomron. "Machine learning-based prediction of COVID-19 diagnosis based on symptoms, Tel Aviv University, 6997801 Tel Aviv, Israel.
- [2] Subhanik Purkayastha, Yanhe Xiao2, Zhicheng Jiao, Rujapa Thepumnoesuk, Kasey "Machine Learning-Based Prediction of COVID-19 Severity and Progression to Critical Illness Using CT Imaging and Clinical Data". eISSN 2005-8330, Korean J Radiol 2021;22(7):1213-1224
- [3] Song, Y. et al. Deep learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT images. medRxiv <https://doi.org/10.1101/2020.02.23.20026930> (2020).
- [4] Wang, S. et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). medRxiv, <https://doi.org/10.1101/2020.02.14.20023028> (2020).
- [5] Jin, C. et al. Development and evaluation of an AI system for COVID-19 diagnosis. medRxiv, <https://doi.org/10.1101/2020.03.20.20039834> (2020).
- [6] Punni, N. S. & Agarwal, S. Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. arXiv:2004.11676 [cs, eess] (2020).
- [7] Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 2020;8:475-481
- [8] Long B, Brady WJ, Koyfman A, Gottlieb M. Cardiovascular complications in COVID-19. *Am J Emerg Med* 2020;38:1504- 1507 3. Sheraton M, Deo N, Kashyap R, Surani S. A review of neurological complications of COVID-19. *Cureus* 2020;12:e8192
- [9] Lai CC, Ko WC, Lee PI, Jean SS, Hsueh PR. Extra-respiratory manifestations of COVID-19. *Int J Antimicrob Agents* 2020;56:106024
- [10] Aziz S, Arabi YM, Alhazzani W, Evans L, Citerio G, Fischkoff K, et al. Managing ICU surge during the COVID-19 crisis: rapid guidelines. *Intensive Care Med* 2020;46:1303-1325
- [11] Sun Q, Qiu H, Huang M, Yang Y. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. *Ann Intensive Care* 2020;10:33
- [12] Zhai P, Ding Y, Wu X, Long J, Zhong Y, Li Y. The epidemiology, diagnosis and treatment of COVID-19. *Int J Antimicrob Agents* 2020;55:105955
- [13] Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA* 2020;324:782-793
- [14] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA* 2020;323:1061-1069
- [15] Menni, C. et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.* 26, 1037–1040 (2020)



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)