International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

Covid-19 Disease Prediction Using Machine Learning – An Overview

[1]Dr Arunkumar PM, [2]Dr Kannimuthu Subramanian

[1]Associate Professor, Department of Computer Science and Engineering ,Karpagam College of Engineering, Coimbatore, Tamilnadu, India.
[2]Professor, Department of Computer Science and Engineering ,Karpagam College of Engineering, Coimbatore, Tamilnadu, India.

Email: [1]arunkumarpm@gmail.com

ABSTRACT

Scientific society and medical fraternity are initiating wide range of technological solutions in controlling the spread of novel coronavirus disease (COVID-19). Artificial Intelligence can be a potential game changer in combating corona virus. In the past, the less pronounced epidemic disease spread was predicted accurately by machine learning models. But, in the current situation, COVID-19 pandemic is more severe and is doing irrevocable harm to people's daily life and country's economic development. Countries all around the world are affected adversely by this viral infection. In this paper, an overview of COVID-19, origin of the disease and symptoms are discussed. The paper also explores the broad spectrum of data analysis related to this pandemic and concepts of modern machine learning algorithms for analyzing and predicting COVID-19.

KEYWORDS
COVID-19, Machine learning, Corona virus, Disease prediction, Artificial Intelligence

1. INTRODUCTION

The novel coronavirus is one of the most infectious diseases to have hit the countries worldwide. The deadly virus has spread over 200 countries and it is in a rampaging mood.The virus was originated from Wuhan, Hubei Province, China during December 2019.The virus has more than 171 million people and claimed at least 3.5 million lives across the world(as on 1st June 2021). World Health Organization (WHO) announced that the official name of the 2019 novel coronavirus is coronavirus disease (COVID-19). Initially, the disease evolved in china with the symptoms of lower respiratory tract infection of patients with pneumonia[1].The researchers across the world started to explore the cause of this disease. The early study showed the relation of this virus with Severe Acute Respiratory Syndrome (SARS) and Middle East respiratory syndrome (MERS) corona viruses.The results of initial research demonstrated the fact that, the primary infection of virus is related to the contact history of local seafood market in Wuhan City,China.The subsequent human-to-human transmission through close contacts is reported to be the major reason for the unbelievable amount of virus spread across the world. Research studies indicate reported COVID-19 has been found to have higher levels of pandemic risk than the SARSCoV. Health department of China have released the communication route of COVID-19 in three different ways[2]. The droplets transmission is the foremost reason and this is due to the

coughing and sneezing of infected person.The normal person may inhale this droplet and get the virus absorbed. The next level of transmission is contact transmission. When the person touches the surface or any products in household that has virus contamination, and if he or she then touches eye, nose or mouth, then there is a clear possibility of virus infection. The final way is through aerosol transmission. The formation of aerosols due to the mixing of respiratory droplets in air can be lethal in certain cases. Such aerosols if inhaled can be a possible reason for the virus spread[3].

The epidemiological indicators of COVID-19 are analyzed by various researchers. The overview of the disease related to epidemiology is tabulated with respect to data available during January 2020 [4].The age of the COVID-19 detected patients range from 2 to 72 years. The male community is highly affected by this virus and the exposure mainly points to Wuhan residential area in China. In the study, the average incubation time is reported as of 6.4 days (5.6–.7.7 days interval).The factors such as age, incubation period, Basic Reproduction number (R0) and Mortality rate are reported with small variations among the research conducted by various scientists during the same period [5].

COVID -19 Symptoms And Treatment Options

The term Coronavirus disease was coined in 1931, with the first coronavirus (HCoV-229E) isolated from humans in 1965. HCoV-229E and HCoV-OC43 are the two human corona viruses (HCoV) prevailed till 2002. SARS coronavirus (SARS-CoV) got much attention after that period. Severe Acute Respiratory Syndrome (SARS) is a species of coronavirus that infects humans, bats and certain other mammals. Structure of corona virus is depicted in Figure.1 :

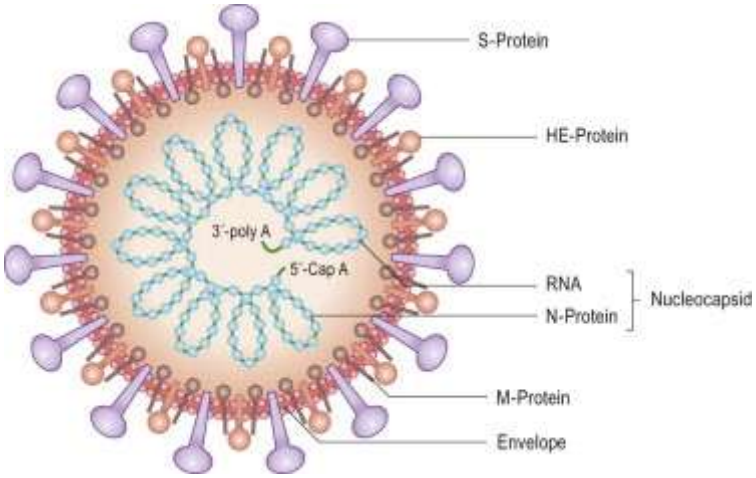

Figure 1. Coronavirus [6]

Replication cycle of Corona virus is shown below in Figure:2

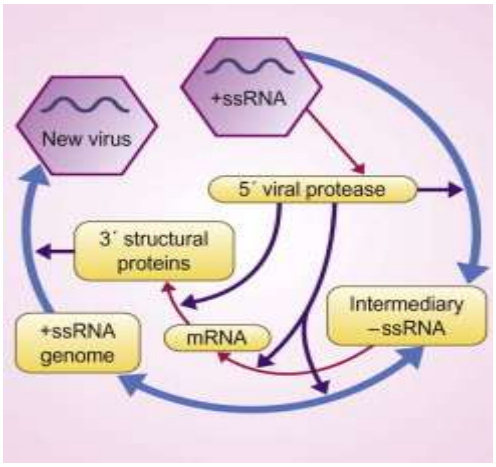

Figure 2 .Corona virus –replication cycle [6]

The common symptoms of COVID-19 include fever, cough, sore throat, headache, fatigue, myalgia and breathlessness. Abnormal features such as RNAaemia, acute respiratory distress syndrome and acute cardiac injury are also reported. In a particular patient history, chest CT showed multiple peripheral ground-glass opacities in both lungs. Interferon inhalation is of very little use in such cases [7].The diagrammatic model that portrays the symptoms of corona virus is elucidated in Figure : 3.

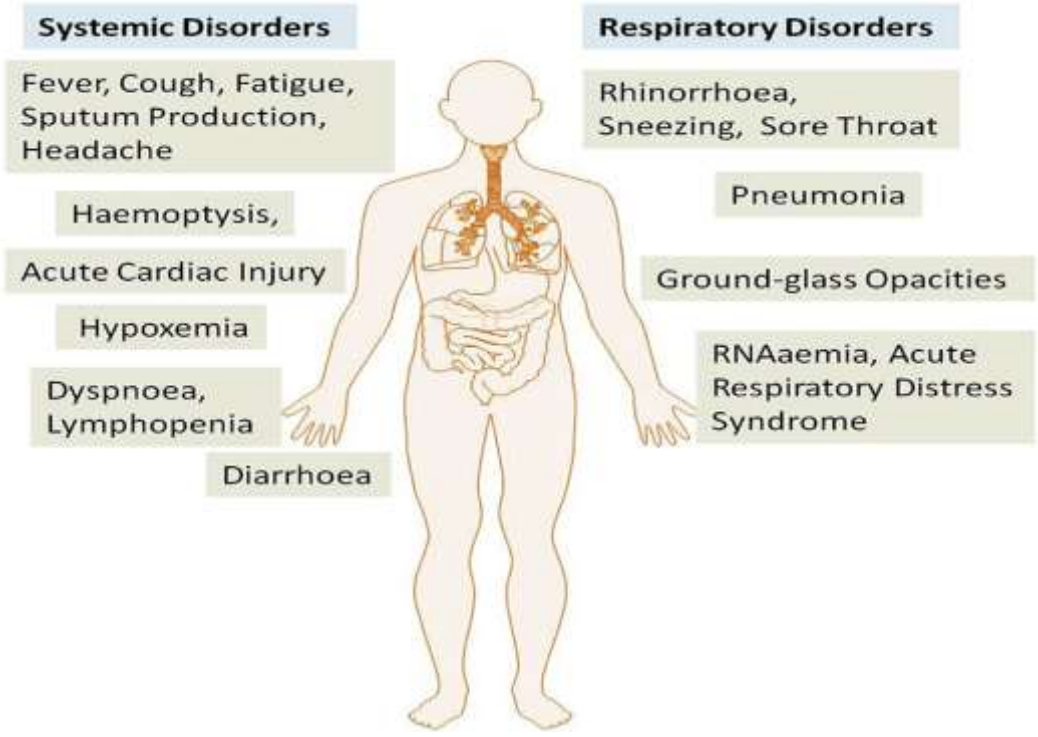

Figure 3 . Clinical features of COVID-19 [8]

The WHO and Centers for Disease Control and Prevention (CDC) have issued direction on important clinical and epidemiological findings suggestive of a COVID-19 infection. Avoiding travel to high-risk countries, using face masks, social distancing and

maintaining basic hand hygiene are the important precautionary measures recommended. Polymerase chain reaction (PCR) testing is used as diagnostic method to confirm COVID-19 in the United States. Reverse transcription PCR (RT-PCR) technique is performed to explore the virus sample in a better way and to conclude the disease development. General methods of testing can be classified as Molecular tests and Serological tests. Polymerase chain reaction test comes under molecular testing and this method involves the testing of sample taken from throat to detect active infection. During the test, if two genes are identified within the SARS-CoV-2 genome, then patient may have acquired the virus (positive case). Serological tests involve the discovery of antibodies that fight such viruses [8].

As the virus still has no active medicine, it continues to affect large group of population in a very severe fashion. Antipyretic therapy and oxygen therapy are recommended as treatment options to control the severity of patient condition. Drug named Remdesivir is used initially in US to treat COVID-19 case. Indian Council of Medical Research (ICMR) approved the use of hydroxychloroquine as prophylaxis and can be used by medical practioners who are treating the patients affected by COVID-19.As on date, researchers around the world are inconclusive in identifying one specific line of treatment for controlling this virus spread. The outcome of the control measures and treatment options vary from one country to other. The general pitfalls of the process carried out till date is summarized [9].The transparency about the COVID-19 virus is not sufficient for the global health community to take immediate measures. So, stringent policy measures needed to improve transparency. The delay in imposing travel restriction is another big issue. Though the virus got attention during December 2019, the airport restriction came in to force only after two or three months in many countries. The screening process is not foolproof at the port of entry. Quarantine delay and lack of research on pandemic diseases are also noted. Moreover, the rumors and fake news about the viral infection results in mental stress among the general public. These issues need to be tackled in future with clear rules and regulations formulated by the highest health authorities [10]. Evolution of COVID-19 based on complete genome analysis is carried out. Phylogenetic analysis was done using the MEGA 7.0 version applying the Maximum likelihood model [11].

USING MACHINE LEARNING FOR COVID-19 PREDICTION

Artificial intelligence is progressing rapidly and the intelligence demonstrated by computers is used effectively in medical domain too. Disease prediction using advanced Machine learning techniques have reached pinnacle in the field of computer science and Information Technology. In recent years, many diseases such as diabetes, lung infection and heart diseases are predicted using supervised machine learning algorithms. The advancement in machine learning is applied to disease prediction of air borne and water borne diseases such as tuberculosis, influenza, small pox, cholera and typhoid. Diseases transmitted by mosquitoes such as malaria and dengue are also explored by researchers in the machine learning community. Many technologies are used in controlling the spread of epidemic diseases. Armed Forces Medical Services (AFMS) and other authorities have the responsibility of initiating emergency operations for controlling disease spread and following WHO guidelines. Real-time map of the outbreak can be visualized using Bigdata tools. Daily reports depicting the disease spread ratio and statistics can be viewed for further analysis. In the same way, Artificial intelligence can be utilized to control COVID-19 disease. Machine learning can be applied in many dimensions with respect to COVID-19 disease control. The application of machine learning can provide warnings and alerts, prediction of disease,

diagnosis and prognosis. The advanced deep learning algorithm can also be used to offer better treatments by giving enhanced and automated medical decision to the doctors.

Lot of research is carried out to devise accurate prediction of the epidemic trend of COVID-19.The disease trend using SEIR (Susceptible, Exposed, Infectious, and Removed) model is initiated. Combination of SEIR model and machine learning approach is utilized by few researchers to track the epidemic trend.[12]. The spread analysis of this virus based on travel volume data (December 2019-January 2020) is conducted [13] .A novel machine learning prediction is carried out by using Eureqa algorithm. Automatic prediction of COVID-19 spread across china and other parts of the World is performed [14]. Public data related to confirmed cases (CC), new cases (NC), and death cases (DC) of COVID-19 is used for the study. National Health Committee of China and Worldometer are the sources of data fetched for cases in China and rest of the world respectively. The machine learning model has predicted that the virus spread may get stopped during May 2020 in China.Also, the model predicts a total positive cases of COVID-19 in china to be 89, 000 in count. Using a particular time span, the same model is used to predict the disease spread analysis for remaining countries excluding China.Here, the analysis is done by keeping Feb 27, 2020 as start date for 14 day period. The results conclude the possibility of 403,216 count for the COVID-19 cases [14] .

A novel machine learning-based prognostic model to predict the survival of corona virus affected patients is proposed [15].The data with a time stamp of January 10th to February 18th , 2020 is collected from Tongji hospital,Wuhan.The data is collected from confirmed and suspected COVID-19 patients. After initial screening stage, the data from 3,129 patients is scaled down to 375 with complete details. Lactic dehydrogenase (LDH), lymphocyte and High-sensitivity C-reactive protein (hsCRP) are considered as the critical features and the machine learning model is developed. Supervised XGBoost classifier is used to evaluate the performance of the model. The flow diagram of the model is elucidated as shown in Figure: 4.

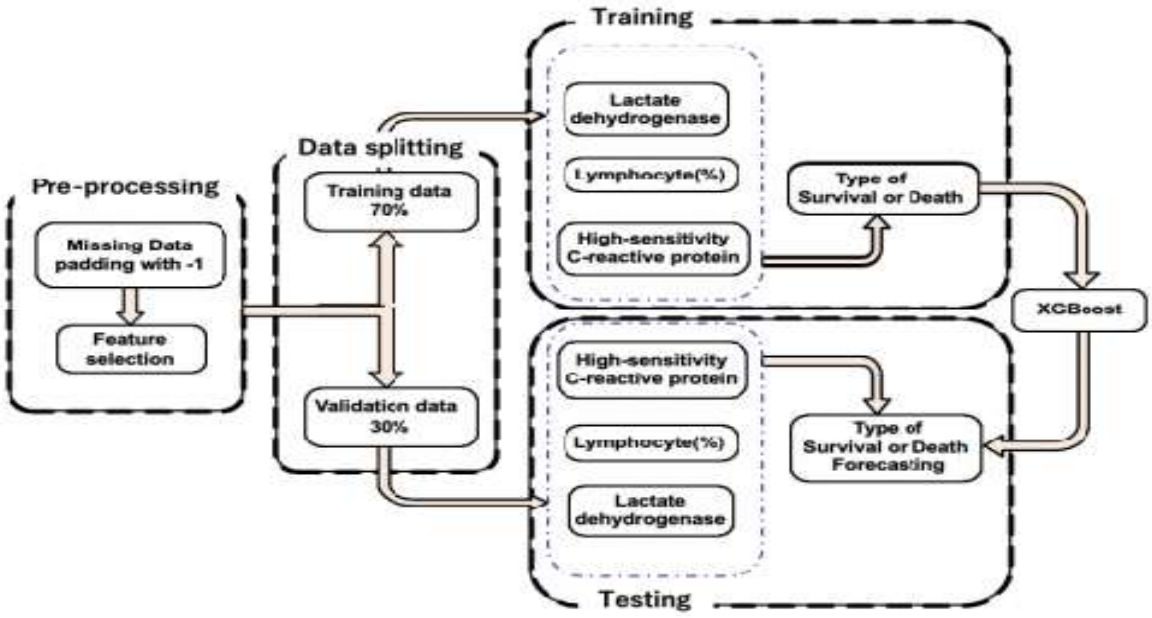

Figure 4 . XGBoost machine learning algorithm [15]

The preprocessing is done with padding method to address the missing data. Key features were ranked by Multi-tree XGBoost and 3 important features are selected. The data split is done and the training and test data are formulated. The model is evaluated by using the XGBoost classifier. 100% death prediction accuracy and 90% survival prediction accuracy are observed as results.

A new deep learning model is proposed as a classification network for distinguishing the COVID-19 from Influenza-A viral pneumonia [16]. Real time reverse transcription-polymerase chain reaction (RT-PCR) detection has low positive rates in detecting COVID-19. Computed tomography (CT) based Deep learning model is best suited for early detection. 618 CT samples are collected from selected hospitals of Zhejiang Province, China. Location-attention classification is carried out. Two CNN three-dimensional classification models were evaluated. ResNet-18 network structure was used for feature extraction. The output of the convolution layer was modified accordingly and final confidential score was evaluated. This model was compared with the network model with and without the added location-attention mechanism. The proposed deep learning model with location-attention mechanism offers 86.7% accuracy for classifying COVID-19 from other viral infections. Another deep learning model for COVID-19 analysis using CT images is carried out using 453 CT images. Decision tree and Adaboost machine learning models are combined as ensemble model. The performance indicators used in this work are Accuracy, Sensitivity, Specificity, Area Under Curve (AUC), Positive predictive value (PPV), Negative predictive value (NPV), F1 score and Youden Index. The internal validation of the model gave an accuracy of 82.9%, whereas external testing produced 73.1% accuracy [17] .

CRISPR-based virus detection system for COVID-19 is proposed [18]. Designing of nucleic acid detection assays using machine learning approach is performed in this work. The complete results are yet to be accumulated by the authors. Respiratory and breathing characteristics are taken in to consideration to evaluate COVID-19 patients [19]. Respiratory Simulation Model (RSM) is adopted in this work and a novel neural network model is evaluated. The result shows an accuracy of 94.5%.Based on data from Israeli Ministry of Health, a novel Machine learning approach for predicting COVID-19 disease is developed [20]. The training set includes 51,831 medical samples with 8 binary features. Gradient-boosting machine and decision tree algorithms are utilized and the performance is observed using auROC (area under the receiver operating characteristic curve).The results are most encouraging and the outcome of the result will help the machine learning experts to explore much deeper to innovate novel ideas in future. As on date, lot of researchers are exploring both supervised and unsupervised machine learning models for predicting COVID-19 disease. Logistic regression and Artificial neural networks are used frequently in many research articles.Moreover, certain research investigation reveals that supervised learning is more accurate to detect COVID-19 with more than 90% accuracy compared to unsupervised machine learning [21].

ANALYSIS OF COVID-19 DATA

The machine learning model can be applied to country-wise data of COVID-19 patients. A snapshot of first 10 countries affected severely (in terms of number of positive cases) is shown in Table 1 .This data is recorded as on 2nd April 2020 by Worldometer data source [22].

Table 1 : 2020 COVID-19 data[22]

Country	Total cases until 2nd April 2020	New cases on single day (2nd April 2020)	Total deaths until 2nd April 2020	New deaths on single day (2nd April 2020)
USA	244,877	+29,874	6,070	+968
Italy	115,242	+4,668	13,915	+760
Spain	112,065	+7,947	10,348	+961
Germany	84,794	+6,813	1,107	+176
China	81,589	+35	3,318	+6
France	59,105	+2,116	5,387	+1,355
Iran	50,468	+2,875	3,160	+124
UK	33,718	+4,244	2,921	+569
Switzerland	18,827	+1,059	536	+48
Turkey	18,135	+2,456	356	+79

The second largest populated country in the world, India is also suffering from the effects of COVID-19.A total count of 2543 active COVID-19 cases are detected on 2nd April 2020, with a death count of 72.This real-time data is updated on a daily basis and hence regression based models can be used to predict the disease count for the coming days. The data observed during April 2020 shows that USA is leading the world with maximum number of COVID-19 cases. The staggering count has crossed 2 lakh entries. Italy is leading in terms of total death with a death count of 13,915 as on 2nd April 2020.

The spread of COVID disease vary time to time. The trend in 2021 is different when compared to previous year. Few countries reported less number of disease count showing the possible decline in the severity.But,few geographical locations are now clustered with large disease spread.US and India are the worst affected countries with largest number of COVID infections as on date. As like the data shown in Table1 (2020 data) , the top 10 countries leading in the COVID-infected count (as on 31st May 2021) is listed in Table 2.

Table 2 : 2021 COVID-19 data[23]

Country	Total cases until 31stMay 2021	New cases on single day (31stMay 2021)	Total deaths until 31stMay 2021	New deaths on single day (31stMay 2021)
USA	34,113,146	+5,235	609,767	+115
India	28,173,655	+126,698	331,909	+2,782
Brazil	16,547,674	+32,554	462,966	+874
France	5,667,324	+1,211	109,528	+126
Turkey	5,249,404	+6,493	47,527	+122
Russia	5,071,917	+8,475	121,501	+339
UK	4,487,339	+3,383	127,782	+1
Italy	4,217,821	+1,820	126,128	+82
Argentina	3,781,784	+28,175	78,093	+637
Germany	3,689,918	+2,203	89,148	+97

As on 31st May 2021,USA leads the world countries with more than 34 million COVID cases. India is worst affected in the year 2021 with a total cases above 28 million and it is the only country reporting more than 0.1 million new cases per day [23]. Such data can be used by machine learning community to explore and evaluate more results. The proper data processing with foolproof modeling of machine learning algorithms is mandatory to acquire better results.

By Observing Table 1 and Table 2, it is imperative to note that the countries namely, USA, Italy, Germany, France, UK and Turkey are consistently placed in the top 10 category while exploring data during 2020 and 2021. The death count per one million population is analyzed among the above said six countries. The same data instances (April 2020 and May 2021) are used and the outcome is elucidated in Figure 5.

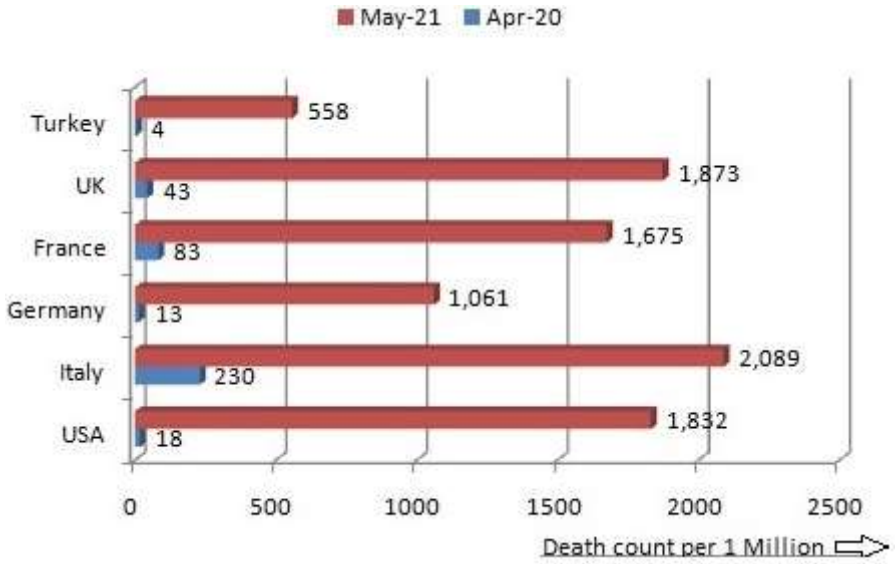

Figure 5 : COVID deaths per 1 Million comparative chart

The figure portrays the death count per 1 million population of countries with severe infection reported during the last 24 months. The mortality due to COVID disease with respect to population cluster is important to identify the most vulnerable countries. Among the countries with highest COVID infections, Italy shows very large death count per 1 million population.UK and USA are also showing staggering death rate. Hence, such countries need immediate revival techniques to combat this deadly disease. Modern technologies such as Machine learning will help to leverage better solutions.

2. CONCLUSION :

Artificial Intelligence is a potentially powerful tool in the fight against the pandemic. The advancement in Machine learning algorithms and improvisation in medical innovations can be combined together to address the disease prediction related to COVID-19.The initial research by scientists have given enormous scope in applying machine learning models to predict the disease progression. The time saved by the automated system could help to save many valuable lives. But on the downside, Machine learning algorithms extract decisions from large amount of data. Such large and relevant data for this disease is scarcely available now and this could hinder the progress for a while.But, due to the advancing in technology and data rich sources emulating quickly from medical fraternity, this issue can be sorted out

easily. The proliferation of data analytics field can be used along with the conventional medical aids to combat the threats posed by the deadly COVID-19 in near future.

3. REFERENCES

[1]. WHO. Novel Coronavirus–China. 2020. https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/. [Online], Available: 1 Feb 2020.

[2]. National Health Commission of People's Republic of China. Prevent guideline of 2019-nCoV. 2020. http://www.nhc.gov.cn/xcs/yqfkdt/202001/bc661e49b5bc487dba182f5c49ac445b.shtml . . [Online], Available: 1 Feb 2020.

[3]. National Health Commission of People's Republic of China. Pneumoniadiagnosis and treatment of 2019-nCoV infection from Chinese NHC andCDC 2020. 2020. [Online], Available: http://www.nhc.gov.cn/xcs/zhengcwj/202001/4294563ed35b43209b31739bd0785e67/files/7a9309111267475a99d4306962c8bf78.pdf.Accessed 1 Feb 2020.

[4]. Adhikari, S. P., Meng, S., Wu, Y. J., Mao, Y. P., Ye, R. X., Wang, Q. Z., ... & Zhou, H. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. Infectious Diseases of Poverty, 9(1), 1-12,2020.

[5]. Backer JA, Klinkenberg D, Wallinga J. The incubation period of 2019-nCoV infections among travellers from Wuhan. China Euro Surveill. ; https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062.,2020.

[6]. Korsman, S. N., Van Zyl, G., Preiser, W., Nutt, L., & Andersson, M. I. Virology E-Book: An Illustrated Colour Text. Elsevier Health Sciences, 2012.

[7]. Lei, J., Li, J., Li, X., & Qi, X. CT imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia. Radiology, 200236,2020.

[8]. Rothan, H. A., & Byrareddy, S. N.. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of Autoimmunity, 102433,2020.

[9]. https://www.medicalnewstoday.com/articles/coronavirus-testing [Online], Available: 03 April 2020.

[10]. Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., ... & Agha, R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery,2020.

[11]. Malik, Y. S., Sircar, S., Bhat, S., Sharun, K., Dhama, K., Dadar, M., ... & Chaicumpa, W. Emerging novel coronavirus (2019-nCoV)—current scenario, evolutionary perspective based on genome analysis and recent developments. Veterinary Quarterly, 40(1), 68-76, 2020.

[12]. Yang, Z., Zeng, Z., Wang, K., Wong, S. S., Liang, W., Zanin, M., ... & Liang, J. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease, 12(2) ,2020.

[13]. Wu, J. T., Leung, K., & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet, 395(10225), 689-697,2020.

[14]. Li, M., Zhang, Z., Jiang, S., Liu, Q., Chen, C., Zhang, Y., & Wang, X. Predicting the epidemic trend of COVID-19 in China and across the world using the machine learning approach. medRxiv, ,2020.

[15]. Yan, L., Zhang, H. T., Xiao, Y., Wang, M., Sun, C., Liang, J., ... & Tang, X. Prediction of survival for severe Covid-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in Wuhan. medRxiv, ,2020.

[16]. Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., ... & Li, Y. Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia. arXiv preprint arXiv:2002.09334,2020.

[17]. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., ... & Xu, B. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). medRxiv, ,2020.

[18]. Metsky, H. C., Freije, C. A., Kosoko-Thoroddsen, T. S. F., Sabeti, P. C., & Myhrvold, C. CRISPR-based COVID-19 surveillance using a genomically-comprehensive machine learning approach. bioRxiv,2020.

[19]. Wang, Y., Hu, M., Li, Q., Zhang, X. P., Zhai, G., & Yao, N. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. arXiv preprint arXiv:2002.05534,2020.

[20]. Zoabi, Y., Deri-Rozov, S., & Shomron, N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. npj digital medicine, 4(1), 1-5,2021.

[21]. Kwekha-Rashid, A. S., Abduljabbar, H. N., & Alhayani, B. Coronavirus disease (COVID-19) cases analysis using machine-learning applications. Applied Nanoscience, 1-13,2021.

[22]. 22.Worldometer online resource : https://www.worldometers.info/coronavirus/ [Online], Available: 3rd April 2020.

[23]. Worldometer online resource : https://www.worldometers.info/coronavirus/ [Online], Available: 1st June 2021.