

# Empirical Learning on Machine Learning Approaches for Discovery and Forecast of Coronavirus

**A. Vaideghy**

Assistant Professor, Department of Computer Technology, PSG College of Arts & Science, Coimbatore, India

**Dr.C. Thiyagarajan**

Associate Professor, Department of Computer Technology, PSG College of Arts & Science, Coimbatore, India

**Dr.G. Meena Suguanthi**

Assistant Professor, Department of Management, Karpagam Academy of Higher Education, Coimbatore, India

*Abstract* - The new, fatal Coronavirus disease originated in the Wuhan city of China in December 2019 by spreading like a fire in different stages across the entire world. It has affected millions of people and has caused deaths in thousands worldwide. It is a requisite to find the infected people, classify them and take necessary precaution in earlier stages. Due to the increased spread of Coronavirus, limited polymerase chain reaction kits are available in the hospitals for undergoing testing for Coronavirus patients. It is more essential to develop artificial intelligence-based tools for automatic diagnostic and classification of Coronavirus epidemic. This paper aims to acquire knowledge about the disease epidemiology, prevention from spreading of Coronavirus Severe Acute Respiratory Syndrome, and suggests various algorithms of machine learning. The work classifies the images of Coronavirus obtained from X-Ray and computed tomography. Machine learning technologies have potential clinical applications for medical images.

*Keywords* - Coronavirus, Epidemic disease, Machine learning algorithm, Treatments.

## INTRODUCTION

Viruses get it's originate air, water, and soil. They can cause infectious diseases such as common flu, cold, and warts and severe illnesses such as Acquired Immunodeficiency Syndrome (AIDS), Ebola, Smallpox [1], etc. A virus goes into the normal cells, replicates themselves, destroys or damages the cells making us sick, and this formation is called "Pandemic virus". Viruses injure certain cells of the human body such as our respiratory system, liver, or blood. For epidemiologic infections, treatment of symptoms is the only way to save human life. There are some antiviral drugs and medicines that cure or heal the viral infections instead of antibiotics. Moreover vaccines save mankind from viral diseases.

Pandemic situations arise due to an outbreak of any infectious disease in a wide-scale causing a rapid increase in mortality and morbidity rate over a large scale. This situation disturbs a large population of the world leading to significant social, economic, and political disturbances. Also a previous study has been done to find the chances of pandemics raised in the past centuries [2]. When a novel virus spreads rapidly globally, pandemic outbreak happens and epidemic situations happen in a specific region or city. Evidence shows that the death rate of the pandemic outbreak is generally higher than the epidemic. A history shows that, Spanish flu is an unfavorable pandemic which has affected more than a hundred million people leading to death [3].

There is a rapid spread of Coronavirus around the world till now. Coronavirus is a giant family of viruses initiating illness from a common cold to acute or severe disease. SARS-CoV-2 identifies 80% of SARS-CoV-1 and 50% of the Middle East Respiratory Syndrome (MARS-CoV) from genetic sequences having their origin from bats. As per genetic sequence and phylogenetic reports, Coronavirus and SARS- CoV are similar passably. Spread of this disease is most likely from bats. Also, from a variety of animal species, this virus has a high mark of homology of ACE2 (Angiotensin Converting Enzyme 2) receptors [4]. High consistent fever, Persistent dry cough, respiratory syndromes (breathing difficulties, breath shortness) are the common signs of the Coronavirus infections [5]. This infection has transitioned pandemic globally like a fire. There is no therapeutic vaccine or drug found by any medical researchers. Identifying the diseases at the early stage is substantial. Also, infected people should be isolated from the non-affected population immediately.

The development of abundant digital technologies in the domains of Big-data analytics, IoT (Internet of things), Block chain, Artificial Intelligence (AI) has paved ways to tackle different diseases and major clinical problems by earlier and easy diagnosis. AI uses machine learning (ML) and deep learning (DL) technologies in many aspects of healthcare sector to predict the outcomes and understand the healthcare trends [6]. Diagnosis of pneumonia is done from CT and X-ray scans of the chest region. To detect, screen and measure Coronavirus, AI-based automated CT or X-ray images analysis techniques can be availed [7].

## 1. Outburst of Coronavirus

As of 14 April 2020, WHO reports 111,828 deaths from 177,687 confirmed cases from 213 countries for this Coronavirus infection. The most affected countries are the USA, China, Italy, and Spain. Originally, hospitals, emergency medical systems and the public health departments in Italy were harassing to comply with the affected patients due to its surge. Regions of China showed the increasing fatality rate of the Coronavirus in the outbreak. On 22nd January 2020, the National Health Commission of China reported its first 17 deaths. Again on 25th January 2020, China reported 56 deaths [4]. By 13th January 2020, 41 cases were confirmed in the city of Wuhan, upon consideration from the early case in December 2019 to the growth of new cases outside. An epidemiologic analysis reveals that this human-to-human transmission has happened due to the close interaction. Outside China, Thailand reported the first case on 13 January 2020, because of the travel history by a Wuhan resident to Thailand. In January, from outside Wuhan City, the cases reported were 19 [8].

Around 86 percent among 37,269 confirmed cases were due to the peoples residence or travel history to Wuhan or close contact through persons who had been to Wuhan. The infection range of Coronavirus varies widely from asymptomatic condition to severity and serious pneumonia accomplished with more numbered fatality rates. As per the report received from China's Disease Control Centers, the majority of the infected people were classified as a mild symptom to moderate illness. Around 13.8% of the infected people were in severe condition and 4.7% of the infected people were treated in critical condition. For confirmed cases of Coronavirus, it was detected that, male patients reported a higher mortality rate than female patients. Also, there was a high risk for death in both cases with rising age. The fatality rate was maximum for the peoples aged 80 and above [9].

## 2. Prevention Measures to Control Sars-Cov-2

WHO adopted various preventive measures such as wearing of facial masks, recurrent hand wash, usage of sanitizers, avoiding touching the eyes, nose and mouth in the face region, wearing hand gloves and social distancing as precautions that people can take to control the epidemic.

Recent report says that, due to transmission or household spread, infected cases were reported as 78%. Recently, a clinical report says that 5% of the hospital acquired infections from their medical staff and patients. Studies recommend lessening person to person contact is more important to control the virus spread until the suitable vaccine is made available.

A study highlights the symptoms of Coronavirus from travel history rather than chest's radiography. Moreover, it is very important to isolate Coronavirus patients for its early treatment and detection. Initiatives were taken to lessen the disease's wild spread and deliver enhanced healthcare systems to the common public. There is a necessity to typify Coronavirus in order to monitor and help general public health recommendations, develop vaccines, therapeutics, and afford suitable diagnostics.

With the advancement of the digital era, there are improved chances for propagation of knowledge. Meanwhile, there is more potential for the development and spread of bogus news or miscommunication over the internet channel. Due to this reason, to provide accurate information and clarify miscommunication related to Coronavirus infection, the people's government needs to take charge to help the common public.

## LITERATURE REVIEW

This article commensurate the assessments done by various authors in solving a clinical problem related to the machine learning domain.

Being a branch of AI, Machine Learning learns the system from the data obtained previously to recognize patterns and create decisions with the minimal intrusion of humans. Some examples of such algorithms are Support Vector Machine (SVM), logistic regression, clustering, etc.

Thiyagarajan C., Anandha Kumar K., Bharathi K., [10] has described the various machine learning approaches in detecting diabetes levels.

Barstugan M., Ozkaya U., and Ozturk S., [11] has adopted the machine learning methods for Coronavirus detection in early stage with the aid of CT Scan images. The proposed method has been adopted for 53 cases obtained from 150 numbered abdominal images CT scans and 150 numbered images of chest CT images. In this study, Images were cropped and the required features have been extracted using various feature extraction techniques. The classifier, Support Vector Machine classifies the features extracted. The finest outcome was obtained from the feature extraction techniques of GLSZM consisting of an accuracy rate of 99.68%.

Naseer Qreshi K, Din S., & Jeon G [12] has proposed a precise, vibrant, smart M-Health system. Their work invokes a predictive model that works in the machine learning domain supporting collection of data, pre-processing the data, partitioning the data, learning the algorithm and making decisions based on the trained datasets.

Lalmuanawma, S., Hussain, J., & Chhakchhuak, L [13] highlights the applications of various machine learning approaches for the outbreak situation of Coronavirus. Their work reveals various measures to address the pandemic.

Shafaf N., & Malek H [14] encompasses the uses of machine learning approaches in emergency medicine. Their study compiles and evaluates the approaches in existing studies to recent years in the field of artificial intelligence related to emergency medicine.

Christopher Tack [15] has labeled the vital applications of machine learning in musculoskeletal medicine.

Schwartz, J.M., Moy, A.J., Rossetti, S.C., Elhadad, N., & Cato, K D [16] describes the support rendered to the clinicians through machine learning systems in analyzing electronic medical record data and carrying their diagnosis and treatment process.

Garg A., & Mago V [17] suggests the various machine learning approaches supportive to the medical field.

Hang Lai, Huaxiong Huang, Karim Keshavjee, Aziz Guergachi & Xin Gao [18] have made a comparison study on the algorithms such as GBM and Logistic Regression models and finally has arrived at the result that, they outperforms the methods such as random forest and decision tree models.

Above works from distinct authors suggest numerous ideas in identifying the disease using various machine learning techniques. Though the research done is large in number, the lack of appropriate methods for identifying the disease encourages us to crack out new methods.

## **FUTURE STUDY**

### *1. Support Vector Machine (SVM)*

Support Vector Machine is a machine learning algorithm that belongs to supervised learning and suitable for dealing with problems related to classification and regression. The algorithm generates the best decision boundary called hyper plane for segregation of n-dimensional space into classes to put new data points in the correct category. Hyper plane is created by choosing the extreme vector called the support vector.

### *2. Random Forest Classifier*

Random Forest Classifier is a machine learning algorithm that supports supervised learning favoring problems related to classification and regression problems. In solving complex problems, this algorithm implies the feature of ensemble learning to combine multiple classifiers. Also this approach improves the performance of the system. This classifier improves the prediction accuracy of a particular dataset by computing the average of decision trees on various subsets for the given dataset.

### *3. K-Nearest Neighbour Classifier*

A non-parametric and most suited machine learning algorithm for classification problems related to supervised learning. The algorithm works by making a similarity assumption amongst new and available cases, categorizing by locating the new case that is most similar to the available categories. Based on similarity, the algorithm classifies the new data point from the available data. The algorithm is otherwise called a lazy learner algorithm as it does not learn from the training set immediately. Instead, during classification, from the dataset stored, it implements action. During the training phase, the algorithm just stores the dataset. At the time of facing new data, data will be classified and categorized based on the similarity with the new data.

### *4. Gradient Boost Algorithm (GBM)*

Gradient Boosting Machine algorithm generates final predictions by combining predictions from multiple decision trees. Decision trees are constructed from the weak learners. Every node in the decision tree takes features of different subsets in order to select the best split. This feature allows capturing different signals from the data. Furthermore, every new tree counts the errors committed by the previous trees and builds a successive decision tree. This act builds the trees sequentially.

### *5. Logistic Regression Model*

This algorithm also supports machine learning concepts belonging to the supervised learning category. This model predicts the output of a categorical dependent variable obtained from the independent variables.

### *6. Grey Wolf Optimization (GWO)*

This is a population-based, meta-heuristics algorithm that pretends the leadership hierarchy and the hunting mechanism of grey wolves in nature. Grey wolves are measured as apex predators bagging the top of the food chain. Grey wolves prefer to live in groups (packs), each group containing 5-12 individuals on average. In the group, each individual follows a firm social dominance hierarchy.

### *3.7. Bat Algorithm (BA)*

BA is a recent, swarm based meta-heuristic algorithm for global optimization built on the echolocation trait of microbats. The algorithm can be applied for problems related to missing data estimation.

### *3.8. Firefly Algorithm (FA)*

FA is a recent, meta-heuristic algorithm built on the trait of fireflies and their patterns of flashing. The algorithm can be applied for problems related to missing data estimation.

## **RESULT ANALYSIS**

Many studies and reviews related to detection of the fatal disease, Coronavirus 2019 have been made. Majority of the work has implemented the various machine learning algorithms for its high accuracy standards.

## CONCLUSIONS AND DISCUSSION

Until now, there is no proper remedy available for the treatment of patients affected from Coronavirus. As the cases increase rapidly, most urgent medication is needed in treating the patients worldwide. Also, we can reduce the spread of the disease by earlier detection and prediction of Coronavirus cases. Coronavirus data gathered from the patients worldwide is a great source for R&D departments. Researchers working in the domain of artificial intelligence and machine learning, rely on the current data for development of a suitable tool that suits automatic diagnosis, therapeutic strategy against Coronavirus patients, and for similar types of pandemics for the upcoming future. Moreover, many tests of diagnosis are obtained by involving clustering and classifying large scale data. The results may cause complications in diagnosis and difficulty in obtaining the end results. This study summarizes the various machine learning approaches that can be adopted for the recent spread of Coronavirus.

## REFERENCES

- [1] Wilson D.J., Falush D., McVean G. Germs, genomes and genealogies. *Trends Ecol. Evol.* 2005; 20(1):39–45.
- [2] Tumpey T.M., Garcia-Sastre A., Taubenberger J.K., Palese P., Swayne D.E., Pantin-Jackwood M.J., Schultz-Cherry S., Solorzano A., Van Rooijen N., Katz J.M., Basler C.F. Pathogenicity of influenza viruses with genes from the 1918 pandemic virus: functional roles of alveolar macrophages and neutrophils in limiting virus replication and mortality in mice. *J. Virol.* 2005; 79(23): 14933–14944.
- [3] Erkoreka A. The Spanish influenza pandemic in occidental Europe (1918–1920) and victim age. *Influenza Other Respi. Viruses.* 2010; 4(2): 81–89.
- [4] Rothan H.A., Byrareddy S.N. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J. Autoimmun.* 2020; 109(February): 102433.
- [5] Lai C.C., Shih T.P., Ko W.C., Tang H.J., Hsueh P.R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob. Agents.* 2020; 55(3).
- [6] Wang Z., Tang K. Combating COVID-19: health equity matters. *Nat. Med.* 2020; 26(April): 2019-2021.
- [7] “Department of Biomedical Engineering, Zonguldak Bulent Ecevit University, 67100, Zonguldak, Turkey.”
- [8] Sun J., He W.T., Wang L., Lai A., Ji X., Zhai X., Li G., Suchard M.A., Tian J., Zhou J., Veit M., Su S. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. *Trends Mol. Med.* 2020; 26(5): 483–495. doi: 10.1016/j.molmed.2020.02.008.
- [9] “Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study,” *Bmj*, 1295(March), p. m1295, 2020.
- [10] C. Thiyagarajan, K. Anandha Kumar, A. Bharathi “A Survey on Diabetes Mellitus Prediction using Machine Learning Techniques,” *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 11, Number 3 (2016) pp 1810-1814
- [11] M. Barstugan, U. Ozkaya, and S. Ozturk, “Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods,” 5, 1–10, 2020.
- [12] Naseer Qureshi K, Din S, Jeon G, Piccialli F. An accurate and dynamic predictive model for a smart M-Health system using machine learning. *Inf. Sci.* 2020; 538: 486–502. doi: 10.1016/j.ins.2020.06.025.
- [13] Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos Solitons Fractals.* 2020; 139: 110059. doi: 10.1016/j.chaos.2020.110059.
- [14] Shafaf, N., Malek, H.: Applications of Machine Learning Approaches in Emergency Medicine; a Review Article. *Arch. Acad. Emerg. Med.* 7, (2019).
- [15] Tack C. Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy. *Musculoskelet. Sci. Pract.* 2019; 39: 164–169. doi: 10.1016/j.msksp.2018.11.012
- [16] Schwartz JM, Moy AJ, Rossetti SC, Elhadad N, Cato KD. Clinician involvement in research on machine learning–based predictive clinical decision support for the hospital setting: A scoping review. *J. Am. Med. Inform. Assoc.* 2021; 28: 653–663. doi:10.1093/jamia/ocaa296
- [17] Garg A, Mago V. Role of machine learning in medical research: A survey. *Comput. Sci. Rev.* 2021; 40: 100370. doi:10.1016/j.cosrev.2021.100370
- [18] Hang Lai, Huaxiong Huang, Karim Keshavjee, Aziz Guergachi & Xin Gao, “Predictive models for diabetes mellitus using machine learning techniques”, *BMC Endocrine Disorders*, Article Number. 101, 2019.