International Journal of Innovation in Computer Science and Information Technology
Vol.2, NO.1 , P: 14 - 24
Received : 14 January 2020
Accepted : 20 March 2020

# Investigation and Comparison of Effective Machine Learning Algorithms in order to Improve the Prediction of Corona Virus Behavior

**Bardia Alaedini**
Bachelor's Student of Shamsipour Technical University, Tehran,
IranBardiaa1985@gmail.com
**Amir Hossein Jorsaraei**
Bachelor's Student of Shamsipour Technical University, Tehran, Iran
Amir77hero@gmail.com
**Sasan Berehlia**
Teacher of Shamsipour Technical University, Tehran, Iran
Sberehlia@tvu.ac.ir
**Hamed Sepehrzedeh**
Technical and Vocational University, Tehran, Iran
hsepehrzadeh@tvu.ac.ir

## Abstract

Covid-19 virus has been affecting people's lives as an acute respiratory disease since late 2019. For this reason, it has become a key topic for IT .professionals. Therefore, the concept of machine learning and deep learning can help a lot in controlling this virus. However, different methods of machine learning and deep learning patterns for predicting viral behavior such as mortality data and CT images of the scanned disease have been investigated. In this paper, according to the review of algorithms and work done in this field, the most optimal algorithm for predicting viral behavior in the human body has been identified. Also, these algorithms are compared and categorized based on virus detection. The results show that most of the data used were CT scan images of corona disease patients. Also, these researches have been analyzed in order to use machine learning algorithms, deep learning and neural networks. In addition, in the field of pattern recognition in this area, the most optimal algorithms are related to classical machine learning. Finally, experiments show that the best algorithm for diagnosing corona disease behavior is Naïve Bayes and SVM.

## Keywords

Machine Learning, Deep Learning, Neural networks, Corona virus, Pattern Recognition.

## I. INTRODUCTION

Machine learning actually refers to using the system to reach the student and integrate knowledge through large-scale observation and improve and expand themselves by learning new knowledge instead of planning with that knowledge. Machine learning techniques are used in smart instructors to gain new knowledge about students, learn their skills and learn new approaches. They improve the department by observing how students work and amending laws related to the field or student [1, 2]. Deep learning is a machine learning method that teaches computers. You see this technology in

autonomous vehicles. This enables the vehicle to distinguish between different objects on the road and enables the vehicle to stop when it sees a red light. An autonomous vehicle can propel or hold time to make sure. An artificial neural network learning algorithm, or a neural network, or just a neural network, is a computational learning system that uses a network of functions to understand and translate the log for a desired output, specifically in another form. The concept of an artificial neural network is inspired by human biology and how the neurons of the human brain function to make inputs to the human senses. Neural networks are just one of the tools and approaches used in the algorithm machine. The neural network may be used as a component in various machine learning algorithms to process given inputs in the space where your computer can operate [3, 4].

Coronavirus 2019 (CoV) or SARS-CoV-2, as it is now called, is rapidly spreading from its origin in Wuhan City, Hebei Province, China. As of 05/03/2020, the world has reported about 96,000 cases of coronavirus 2019 (COVID-19) and 3300 deaths. India has reported 29 cases so far. Fortunately, until now, children have rarely been affected and have died. But the future trend of this virus is not clear. This article gives you a brief overview on this new virus. Readers should be asked to update you regularly as the student completes the virus quickly [5, 6].

In this article, the effect of machine learning algorithms in predicting and treating Quid 19 disease is investigated. For this reason, work has been done to review the work done in the field of corona disease in line with the machine learning algorithm, deep learning and neural network. To use this science, data and machine learning have been examined, patterns have been mentioned to review and revise. In this article, we use the algorithms used in the diagnosis of COVID-19 behavior to conclude which of the algorithms plays your role in the diagnosis of virus 19. The results also show that Naïve Bayes and SVM algorithms perform better for diagnosis, analysis, pattern marketing and learning of COVID-19 disease.

In the continuation of this article, the second part is to review the research literature in the direction of machine learning, deep learning and algorithms related to its usage. COVID-19 disease and risk factors for this disease have also been reported. Chapter 3 reviews previous research on coronary heart disease and machine learning. Chapter 4 presents the work done and suggested to improve the performance of research related to machine learning and corona virus. Finally, the fifth chapter concludes the article.

## II. LITERATURE REVIEW

This section examines the concept of machine learning and deep learning. Application algorithms in this field have also been studied. Finally, COVID-19 disease is investigated.

### A. Machine Learning

F Machine learning does exactly what its title implies. Machine learning is actually the way in which a computer program can be learned and automatically improved from experience without being directly programmed. Machine learning is the machine learning of the artificial intelligence (AI) disciplines that systems Enables them to learn and improve themselves automatically and through experience without planning. The focus of this technology is on the development of computer programs that have access to the data and can use this data to learn for themselves. Imagine for a second that you are a web design training instructor and You want to try to determine the best level of learning for your trainees and Decide what

the solution will be in the shortest possible time. This is exactly what is needed in the case of machine learning [2, 7, 8].

Every machine learning program needs an "training data set" to teach it what kind of information it can expect and to understand what kind of information the programmer is looking for. Machine learning algorithms can process large amounts of data and predict outcomes and patterns based on this information. Over time, as the program modifies itself, the predictive model becomes more accurate and accurate, without the need for external manipulation. Algorithms have three main categories, which are defined by the type of training data set they are given. These three categories are: supervised, unsupervised and semi-supervised. Each of these approaches has its advantages and disadvantages, depending on what the program is supposed to achieve [9, 10].

Supervised learning: Machine learning algorithms are trained based on datasets in which a hypothetical input, based on a mapping function, leads to a specific output [11].

Unsupervised learning: Unsupervised machine learning has no correct output for hypothetical input. Unlike supervised machine learning, there is no expected response and no teacher. Only the program itself proceeds based on the data. The purpose of this type of machine learning is to analyze the data in general and to discover facts about the underlying structure [12].

Semi-observational learning: Semi-observational learning is usually a combination of the first two types [11, 12].

Each of these categories has algorithms to improve its performance. Table 1 examines machine learning algorithms by type of learning [10, 11, 12, 13, 14].

**Table 1. Machine learning algorithms**

| Name | Category | Description |
|---|---|---|
| Linear Regression | Supervised-regression | Linear regression algorithm uses data points to find the best suitable line for data modeling. |
| SVM (Support Vector) Machine | Supervised-classification | The algorithm will separate points of data using a line. This dividing line is selected so that it is the closest line selected between the two categories. |
| KNN (K-Nearest) Neighbors | Supervised-classification | It is a simple algorithm that predicts unknown data points with its nearest k neighbors. Determines the closest distance by calculating distances through basic functions such as Euclidean geometry. |
| Decision Tree | Supervised-classification | This algorithm classifies a population for several sets based on some selected characteristics (independent variables) of a population. This algorithm is usually used to solve classification problems. |
| Random Forest | Ensemble Method | The random forest can be identified as a set of decision trees. Each tree tries to estimate a classification, and this is called a "vote". Ideally, |

| | | |
|---|---|---|
| | | we consider each vote from each tree and select the most votes. |
| Naïve Bayes | Supervised-classification | This algorithm assumes that there is independence between the "predators". In fact, a New Business classification assumes that a feature in one class is unrelated to another. |
| Logistic Regression | Supervised-classification | Logistic regression is used when a clean output is given and some events are expected to occur (for example, predicts whether rain will occur or not). |

### B. Deep Learning

Deep learning is a subset of machine learning that uses algorithms that simulate the human brain. These algorithms are called artificial neural networks. Deep learning is a machine learning method that teaches computers to do what humans usually do: learning by example. Deep learning is the main technology of making cars; Deep learning gives cars the ability to recognize signs, or to distinguish electric light poles from humans. In-depth learning is the key technology used in the audio interface of mobile phones, tablets, TVs and handsfree. Deep learning has been gaining traction for some time; Because it has brought with it results that have never been possible before [3].

In-depth learning, a computer model learns to execute classification commands directly from image, text, or sound. Deep learning models can achieve the highest level of accuracy; So that sometimes they do better than humans. Deep learning models are trained using large batches of data and multi-layered neural networks [3, 15].

### C. Corona Virus

Coronary heart disease is a disease that has become prevalent these days and has occupied the minds of the general public.

There is a lot of talk about this disease and its severity, and this has caused a lot of concerns to spread in the community. Corona virus are a type of virus that affects the respiratory system of mammals, including humans. These viruses can cause colds, pneumonia and respiratory tract infections (SARS). Coronary artery disease also affects and can affect this organ of the body. The human coronavirus (HCoV) was first discovered in the 1960s in the nose of a patient suffering from the common cold. OC43 and 229E viruses from the corona family cause the common cold [5].

The name of this virus is derived from its shape. The appearance of this virus has crown-like corners, which is why it is called Corona. The word in Latin means "crown" or "halo". In humans, infections are often more common during the winter or early spring. It is very common to catch a cold (caused by the Corona virus) and get it again after four months. The reason for the recurrence of this disease may be that the antibodies that the body produces to fight the coronavirus do not remain in the body for a long time. Another reason for the recurrence of the disease is that these antibodies are produced for one type of virus and are ineffective for another [6].

**Table 2. Risk factors for Covid-19 [16]**

| Feature type | Feature name |
|---|---|
| Basic demographic | Age |
| | Weight |
| | Height |
| | Sex |
| | Smoker |
| Most common symptoms | Fever |
| | dry cough |
| | Tiredness |
| Less common symptoms | aches and pains |
| | sore throat |
| | Diarrhea |
| | Conjunctivitis |
| | Headache |
| | loss of taste or smell |
| | a rash on skin, or discoloration of fingers or toes |
| Serious symptoms | difficulty breathing or shortness of breath |
| | chest pain or pressure |
| | loss of speech or movement |
| Background diseases | Heart disease |
| | Lung disease |
| | Liver disease |
| | kidney disease |
| | Diabetes |

Covid-19 is an infectious disease caused by severe acute respiratory syndrome corona virus 2 (SARS-CoV-2). It was first identified in December 2019 in Wuhan, Hubei, China, and has resulted in an ongoing pandemic. The first confirmed case has been traced back to 17 November 2019 in Hubei. As of 6 August 2020, more than 18.7 million cases have been reported across 188 countries and territories, resulting in more than 706,000 deaths. More than 11.3 million people have recovered. Common symptoms include fever, cough, fatigue, shortness of breath, and loss of smell and taste. While the majority of cases result in mild symptoms, some progress to acute respiratory distress syndrome possibly precipitated by cytokine storm, multi-organ failure, septic shock, and blood clots. The time from exposure to onset of symptoms is typically around five days, but may range from two to fourteen days. People with mild symptoms who are otherwise healthy should manage their symptoms at home. On average it takes 5–6 days from when someone is infected with the virus for symptoms to show, however it can take up to 14 days. Table 2 shows the risk factors for Covid-19 disease [6, 16].

## III. REALTED WORKS

In this section, the work done in relation to the use of machine learning and deep learning to improve and treat the COVID-19 virus is reviewed. Finding solutions through machine learning and deep learning algorithms can greatly help in predicting and treating corona disease in the future. In the following, various researches have been reviewed.

Hosseinzadeh Kassani et al (2020) Examined the automatic diagnosis of coronavirus in X-rays and machine learning-based CT images. In this paper, machine learning and deep learning algorithms (MobileNet, DenseNet, Xception, ResNet, InceptionV3, InceptionRes-NetV2, VGGNet, NASNet) are used to diagnose the disease or use CT scan images. And by comparing the selected methods, they came to the conclusion that the denseNet121 algorithm using the Bagging tree classifier gives the best result with 99% accuracy [17].

Iwendi et al (2020), Examined COVID-19 patient health prediction using augmented random forest algorithm. In this paper, several machine learning algorithms were used to diagnose Covid-19 disease on patient data, including Decision Tree Classifier, Support Vector Classifier, Gaussian Name Bayes Classifier, and Boosted Random Forest Classifier. In the meantime, the efficiency of naïve bayes was higher than other algorithms [18].

James Fong and associates (2020), researched finding an accurate initial prediction model of a small data set based on the prevalence of the Corona virus. This article is about examining a new and hybrid algorithm that includes PNN and CF algorithms. The PNN algorithm along with the Grooms model can play a useful role in predicting the end of an epidemic, and at this point the most important part can be the criterion of error or accuracy of any model with low data [19].

Khanday et al (2020), examined machine-based approaches to detect COVID-19 using clinical text data. One way to predict covid-19 disease is to use artificial intelligence tools. This paper uses machine learning algorithms and compares them. According to the results, Naïve Bayes method has given the best efficiency with 96% accuracy [20].

Yae et al (2020), Examined the severity of COVID-19 disease using a machine learning model based on blood and urine tests. In this paper, using the results of blood and urine tests and modeling them, they were able to detect the corona virus and the algorithms used in this paper are LR, RF, SVM, KNN, ADABOOST. In examining the modeling result of the experiments, the SVM algorithm has concluded with a higher percentage of accuracy than the other algorithms and has obtained the desired results [21].

Qianqian et al (2020), examined an in-depth learning approach to describe the 2019 coronavirus in chest CT images. In this article, in-depth learning is used to automatically diagnose coronavirus disease in the chest using CT scan images. Out of 96 patients, 88 had lung problems and 8 had no problems. Machine learning algorithms detect covid-19 disease with high accuracy and do it faster than CT scan specialists [22].

In 2020, Xiaowei and associates, proposed an in-depth learning system for screening for corona virus. In this article, deep learning science has been used to compare and evaluate convulsive neural networks and classify CT scan samples based on covid-19, influenza, pneumonia virus or without any infection. Out of 1710 samples, 357 were Covid-19, 390 were influenza and 963 were not related to lung infection. And the interesting thing is that for each CT scan (each of which contains 70 layers), the

average time it takes to get a result is less than 30 seconds [23].

Al-Karawi et al (2020), analyzed machine learning CT scan images of the chest as a complementary digital test of corona virus patients. In this article, 1014 CT scan images of the chest are examined using machine learning algorithms that can detect covid-19 disease. In this analysis, the SVM algorithm is used, which has recorded the best result with 90.51% accuracy [24].

## IV. Analysis and Comparison

Comparing the algorithms used in the articles, we came to the conclusion that on the subject of reviewing CT scan images of patients with corona disease, the best and most optimal algorithms with the lowest error rate are related to classical machine learning algorithms. Due to the lack of data and the lack of a clear pattern for virus behavior, it is not possible to get an optimal output from other algorithms, we now turn to the algorithms used in each category.

Classic Machine Learning: In the machine learning algorithms section, whose main function is statistical and numerical analysis, this section uses four algorithms to examine information, including Naïve Bayes, Logistic Regression, SVM, and Decision Tree. Among these algorithms, Naïve Bayes and SVM have the best and most optimal output and the most consumption among researchers, because with the lack of data, the most optimal method is to recognition patterns using classical machine learning algorithms is shown in Figure 1.

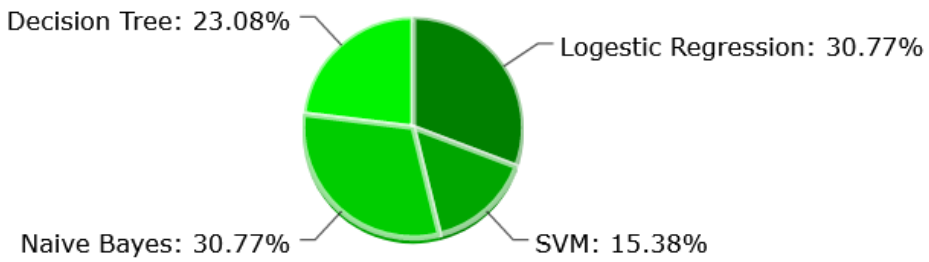


**FIGURE 1. Examining the algorithms used in classical machine learning**

Neural Networks: In the algorithms used in this field, there are about 9 algorithms that PNN and DenseNet are the best algorithms in this field and have the lowest error rate. Although this field may be of lower quality than classical machine learning, Because the main efficiency of this field is modeling from incomplete data, which of course is justified by the lack of corona data. Also, when more data is used and used for prediction, neural networks are a better option than classical machine learning, as they provide better output. Figure 2 shows the effective neural network-based algorithm.

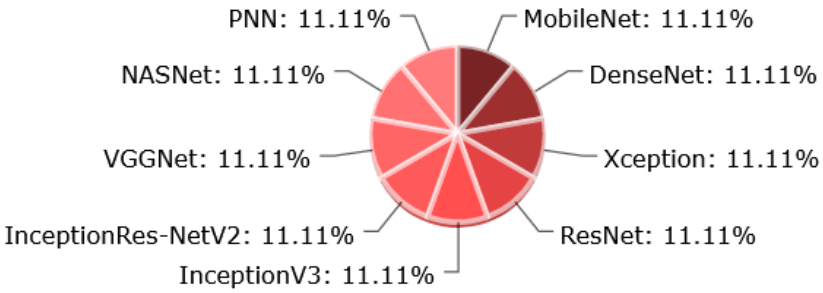


**FIGURE 2. Investigation of algorithms used in neural networks**

Ensemble methods: This method uses two algorithms that do not have a high percentage of accuracy, which include AdaBoost and Random Forest, because like neural networks, as long as the data is low, it does not provide us with optimal output.Deep learning: from modeling deep learning in reviewing these images by building an image processing model, we have good outputs for predicting virus behavior and in the future, we can use these outputs for processing with neural network algorithms and more efficient processing.

Proposed model: To use algorithms in the future, neural networks are a better option to replace the classical machine learning algorithms, because by modeling and predicting virus behavior, the reactions of viruses that may be transmitted to humans in the future can be predict and control, but at a time like now when it's hard to know virus behaviors, classical machine learning algorithms provide us with better output than neural networks.

According to Figure 3, most of the algorithms currently in use have been neural network algorithms, as modeling and prediction are more important to researchers. But due to the lack of a suitable model, the optimal output is not included in this section.

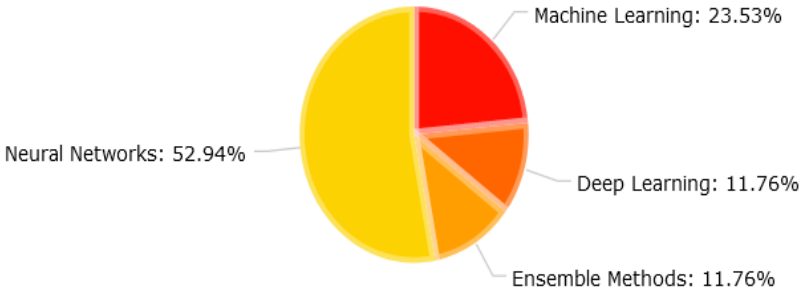


**FIGURE 3. Algorithms used to improve cognition of corona virus behavior**

According to Figure 4, the most optimal output is related to the classic machine learning category, which has been better used with low data.

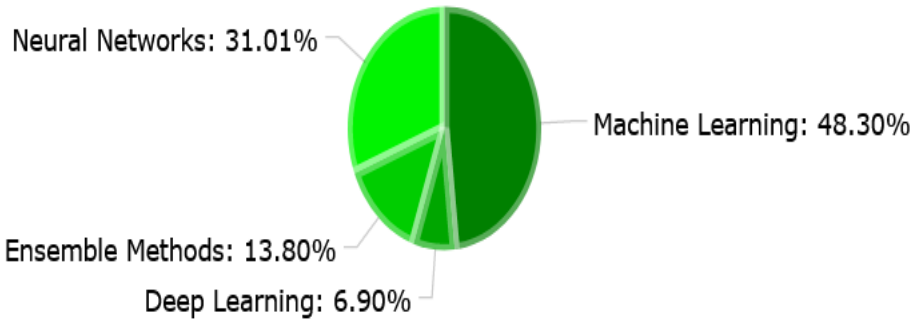


**FIGURE 4. Algorithms used to improve the corona virus with a lower error rate**

According to Figure 5, in the final part of the pattern recognition discussion, the best algorithms were Naïve Bayes and SVM algorithms, because in pattern recognition in this issue, it needed algorithms like these two that could provide a precise model for Researchers prepare.

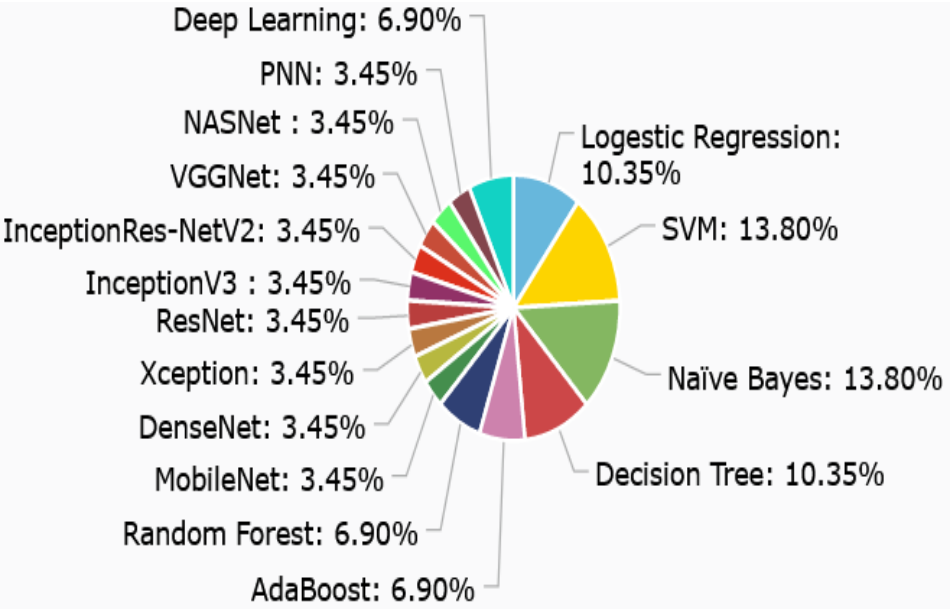


**FIGURE 5. The best algorithms for examining pattern recognition in the Corona virus**

## V. CONCLUSION

Machine learning is used to predict different activities to predict future behaviors and events. Therefore, for high accuracy of this prediction, different algorithms such as: supervised learning, neural networks and deep learning are used. Optimal use of machine learning in a subject such as the COVID-19 virus is to predict future behaviors and symptoms of the virus and to be able to deal better and more effectively with the virus. Analysis In various studies, machine learning algorithms have been used more than deep learning and neural network algorithms. The results show that SVM and Naïve Bayes algorithms with the lowest error rate can have the best performance in detecting coronavirus behavior. Also, in

research, these algorithms have been used more than other algorithms. Furthermore; Using neural network algorithms, deep learning and group learning to improve the coronavirus behavior prediction can be helpful. Also, SVM and Naïve Bayes algorithms can work best in improving pattern recognition due to their low error rate. In the future, machine learning algorithms can be used to more accurately predict and treat corona virus.

**References**


[1]V. Dhar, Data science and prediction. Communications of the ACM. 2013 Dec 1;56(12):64-73.

[2]E. Alpaydin, Introduction to machine learning. MIT press; 2020 Mar 17.

[3]L. Deng, D. Yu D, Deep learning: methods and applications. Foundations and trends in signal processing. 2014 Jun 30;7(3–4):197-387.

[4]MA. Nielsen MA, Neural networks and deep learning. San Francisco, CA: Determination press; 2015 Sep 25.

[5]G. Li, Y. Fan, Y. Lai, T. Han, Z. Li, P. Zhou, P. Pan, W. Wang, D. Hu, X. Liu, and Q. Zhang, Coronavirus infections and immune responses. Journal of medical virology. 2020 Apr;92(4):424-32.

[6]JM. Sharfstein, SJ. Becker, Mello MM. Diagnostic testing for the novel coronavirus. Jama. 2020 Apr 21;323(15):1437-8.

[7]M. Fatima, M. Pasha, Survey of machine learning algorithms for disease diagnostic. Journal of Intelligent Learning Systems and Applications. 2017;9(01):1.

[8]A. Singh, N. Thakur, and A. Sharma, A review of supervised machine learning algorithms. In2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) 2016 Mar 16 (pp. 1310-1315). Ieee.

[9]A. Singh, N. Thakur, and A. Sharma, A review of supervised machine learning algorithms. In2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) 2016 Mar 16 (pp. 1310-1315). Ieee.

[10] R. Choudhary, HK. Gianey, Comprehensive review on supervised machine learning algorithms. In2017 International Conference on Machine Learning and Data Science (MLDS) 2017 Dec 14 (pp. 37-43). IEEE.

[11] FY. Osisanwo, JE. Akinsola, O. Awodele, JO. Hinmikaiye, O. Olakanmi, and J. Akinjobi, Supervised machine learning algorithms: classification and comparison. International Journal of Computer Trends and Technology (IJCTT). 2017 Jun;48(3):128-38.

[12] N. Amruthnath, T. Gupta, A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In2018 5th International Conference on Industrial Engineering and Applications (ICIEA) 2018 Apr 26 (pp. 355-361). IEEE.

[13] S. Uddin, A. Khan, ME. Hossain, and MA. Moni, Comparing different supervised machine learning algorithms for disease prediction. BMC Medical Informatics and Decision Making. 2019 Dec;19(1):1-6.

[14] M. Khanum, T. Mahboob, W. Imtiaz, HA. Ghafoor, and R. Sehar, A survey on unsupervised machine learning algorithms for automation, classification and maintenance. International Journal of Computer Applications. 2015 Jan 1;119(13).

[15] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, MP. Reyes, ML. Shyu, SC. Chen, and SS. Iyengar, A survey on deep learning: Algorithms, techniques, and

applications. ACM Computing Surveys (CSUR). 2018 Sep 18;51(5):1-36.

[16] Y. Zhou, Z. Zhang, J. Tian, and S. Xiong, Risk factors associated with disease progression in a cohort of patients infected with the 2019 novel coronavirus. Annals of palliative medicine. 2020 Mar 17;9(2):428-36.

[17] SH. Kassani, PH. Kassasni, MJ. Wesolowski, KA. Schneider, and R. Deters, Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning-Based Approach. arXiv preprint arXiv:2004.10641. 2020 Apr 22.

[18] C. Iwendi, AK. Bashir, A. Peshkar, R. Sujatha, JM. Chatterjee, S. Pasupuleti, R. Mishra, S. Pillai, and O. Jo, COVID-19 Patient health prediction using boosted random forest algorithm. Frontiers in public health. 2020 Jul 3;8:357.

[19] SJ. Fong, G. Li, N. Dey, RG. Crespo, and E. Herrera-Viedma, Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak. arXiv preprint arXiv:2003.10776. 2020 Mar 24.

[20] AM. Khanday, ST. Rabani, QR. Khan, N. Rouf, and MM. Din, Machine learning based approaches for detecting COVID-19 using clinical text data. International Journal of Information Technology. 2020 Sep;12(3):731-9.

[21] N. Zhang, R. Zhang, H. Yao, H. Xu, M. Duan, T. Xie, J. Pan, J. Huang, Y. Zhang, X. Xu, and F. Zhou, Severity Detection For the Coronavirus Disease 2019 (COVID-19) Patients Using a Machine Learning Model Based on the Blood and Urine Tests.

[22] Q. Ni, ZY. Sun, L. Qi, W. Chen, Y. Yang, L. Wang, X. Zhang, L. Yang, Y. Fang, Z. Xing, and Z. Zhou, A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images. European radiology. 2020 Jul 2:1-1.

[23] C. Butt, J. Gill, D. Chun, and BA. Babu, Deep learning system to screen coronavirus disease 2019 pneumonia. Applied Intelligence. 2020 Apr 22:1.

[24] D. Al-Karawi, S. Al-Zaidi, N. Polus, and S. Jassim, Machine Learning Analysis of Chest CT Scan Images as a Complementary Digital Test of Coronavirus (COVID-19) Patients. medRxiv. 2020 Jan 1.