



Coronavirus disease (COVID-19) cases analysis using machine-learning applications

Ameer Sardar Kwekha-Rashid¹ · Heamn N. Abduljabbar^{2,3} · Bilal Alhayani⁴

Received: 7 March 2021 / Accepted: 4 May 2021 / Published online: 21 May 2021
© King Abdulaziz City for Science and Technology 2021

Abstract

Today world thinks about coronavirus disease that which means all even this pandemic disease is not unique. The purpose of this study is to detect the role of machine-learning applications and algorithms in investigating and various purposes that deals with COVID-19. Review of the studies that had been published during 2020 and were related to this topic by seeking in Science Direct, Springer, Hindawi, and MDPI using COVID-19, machine learning, supervised learning, and unsupervised learning as keywords. The total articles obtained were 16,306 overall but after limitation; only 14 researches of these articles were included in this study. Our findings show that machine learning can produce an important role in COVID-19 investigations, prediction, and discrimination. In conclusion, machine learning can be involved in the health provider programs and plans to assess and triage the COVID-19 cases. Supervised learning showed better results than other Unsupervised learning algorithms by having 92.9% testing accuracy. In the future recurrent supervised learning can be utilized for superior accuracy.

Keywords COVID-19 · Artificial intelligence AI · Machine learning · Machine learning tasks · Supervised and un-supervised learning

Introduction

Recently, the world gained rapid progression in technology and it shows an important role in the developed countries. Nowadays all daily life sectors such as education, business, marketing, militaries, and communications, engineering, and health sectors are dependent on the new technology applications. The health care center

is a crucial field that strongly needs to apply the new technologies from defining the symptoms to the accurate diagnosis and digital patient's triage. Coronavirus-2 (SARSCoV-2) causes severe respiratory infections, and respiratory disorders, which results in the novel coronavirus disease 2019 (COVID-19) in humans who had been reported as the first case in Wuhan city of China in December 2019. Later, SARS-CoV-2 was spread worldwide and transmitted to millions of people and the world health organization (WHO) have announced the outbreak as a global pandemic since the number of infected people is still increasing day by day. As of 16th December 2020, the total (global) coronavirus cases were approximately 73,806,583 with reported deaths of 1,641,635 (Pasupuleti et al. 2021). The novel coronavirus appeared in December 2019, in the Wuhan city of China and the World Health Organization (W.H.O) reported it on 31st December 2019. The virus produced a global risk and W.H.O named it COVID-19 on 11th February 2020 (Wu 2020). Up to the present time, there was no specific medication that deals directly with this new generation of COVID-19 virus, but some of the companies produced several combination drugs that basically made up from ethanol, isopropyl alcohols, and hydrogen peroxides in different combinations

✉ Ameer Sardar Kwekha-Rashid
ameer.rashid@univsul.edu.iq

Heamn N. Abduljabbar
heamn.abduljabbar@su.edu.krd

Bilal Alhayani
bilalabed1978@gmail.com

- ¹ Business Information Technology, College of Administration and Economics, University of Sulaimani, Sulaimaniya, Iraq
- ² College of Education, Physics Department, Salahaddin University, Shaqlawa, Iraq
- ³ Department of radiology and imaging Faculty of Medicine and Health Sciences, Universiti Putra Malaysia UPM, Seri Kembangan, Malaysia
- ⁴ Electronics and Communication Department, Yildiz Technical University, Istanbul, Turkey

show a significant reaction to the novel virus and had been confirmed and accepted by WHO to be used in the world (Mahmood et al. 2020). The artificial intelligence and deep learning algorithm show the ability to diagnose COVID-19 in precise which can be regarded as a supportive factor to improve the common diagnostic methods including Immunoglobulin M (IgM), Immunoglobulin (IgG), chest x-ray, and computed tomography(CT) scan, also reverse transcription-polymerase chain reaction (RT-PCR) and immunochroma to graphic fluorescence assay. The developments of a potential technology are one of the currently used methods to identify the infection, such as a drone with thermal screening without human intervention, which needs to be encouraged (Manigandan 2020). The assessment of the research that had been produced whether it hits the target of the existing knowledge gaps or not can be done by applying an artificialintelligence/machine learning-based approach to analyze COVID-19 literature (Doanvo et al. 2020). Thus, the acceleration of the diagnosis and treatment of COVID-19 disease is the main advantage of these AI-based platforms (Naseem et al. 2020) which finally shows a huge potential to exponentially enhance and improve health care research (Jamshidi et al. 2020). Corona Virus Disease 2019 (COVID-19), has become a matter of serious concern for each country around the world. AI applications can assist in increasing the accuracy and the speed of identification of cases through data mining to deal with the health crisis efficiently, the rapid expansion of the pandemic has created huge health care disorders which as a result encouraged the real need for immediate reactions to limit the effects. Artificial Intelligence shows great applications in dealing with the issue on many sides (Tayarani-N 2020). The COVID-19 is an epidemic disease that challenged human lives in the world. The systematic reviews showed that machine learning ML training algorithms and statistical models that are used computers to perform various tasks without explicit commands (Bishop 2006). Currently, machine learning techniques are used internationally for predictions due to their accuracy. However, machine learning (ML) techniques, have few challenges such as the new poor database that is available online. For instance, the selection of the appropriate parameters is one of the challenges involved in training a model or the selection of the best Machine learning model for prediction. Depending on the available dataset researchers obtained predictions by using the best Machine learning model that suits the dataset (Shinde 2020). Machine learning techniques can be used to extract hidden patterns and data analytics (Khan 2020). The algorithms of Machine-learning are designed for identifying complex patterns and interfaces in the data, in the context of unknown and complicated correlation patterns among risk factors (Hossain 2019).

Related work

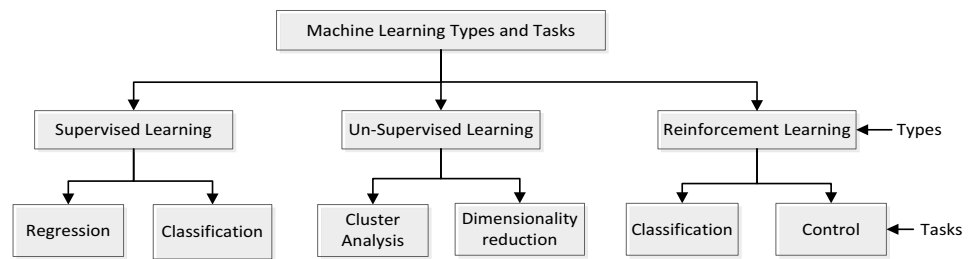
COVID-19

The contagion disease caused by the SARS-COV-2 virus named COVID-19 is requiring extraordinary responses of special intensity and possibility to more than 200 countries around the world, the first 4 months from its epidemic, the number of infected peoples ranged from 2 to 20 million, with at least 200,000 deaths. To manage the spread of the COVID-19 infection among people rapidly, all the governments around the world applied severe actions, such as the quarantine of hundreds of millions of citizens worldwide (Alimadadi 2020). Nevertheless, the difficulty of distinguishing between the positive and negative COVID-19 individuals depending on the various symptoms of COVID-19, all of these efforts are limited. Therefore, tests to detect the SARS-CoV-2 virus are believed to be critical to recognize the positive cases of this infection to limit the (Brinati et al. 2020). Radiology and imaging are some of the most beneficial and critical modalities used for diagnosis COVID-19 stage and hazards on the patient's lungs specifically by chest CT scan (Day 2020). Early diagnosis of COVID-19 is important to minimize human-to-human transmission and patient care. Recently, the separation and quarantine of healthy people from the infected or persons who suspect that they are carrying the virus is the most effective technique to avoid the spread of COVID-19 (Deng 2020). Machine-learning techniques role showed an important understandings of the COVID-19 diagnosis, such as lung computed tomography (CT) scan whether it can be regarded as the first screening or an alternative test for the real-time inverse transcriptase-polymerase chain reaction (RT-PCR), and the differences between COVID-19 pneumonia and other viral pneumonia using CT scan of the lungs(Kassani et al. (2004)).

Machine learning

Machine learning is one of the most promising tools in classification (Hossain 2019). In essence; machine learning is a model that aims to discover the unknown function, dependence, or structure between input and output variables. Usually, these relations are difficult to be existed by explicit algorithms via automated learning process (Zhang 2020a). Machine-learning methods are applied to predict possible confirmed cases and mortality numbers for the upcoming (Hastie et al. 2009). Machine learning can be divided into two parts. The first part is to define the optimal weight of data fusion of multi-node perception outcomes and eliminate unusable nodes based on the

Fig. 1 Overview of machine-learning types and tasks



genetic algorithm, while the second part is to find fault nodes through a fault recognition neural network (Ünlü and Namlı 2020). Machine learning is a subsection of Artificial Intelligence (AI), and it involves several learning paradigms, such as Supervised Learning (SL), Un-supervised Learning (UL), and Reinforcement Learning (RL) (Shirzadi 2018). Typical ML models consist of classification, regression, clustering, anomaly detection, dimensionality reduction, and reward maximization (Gao 2020). The ML algorithms are trained in the SL paradigm, on labeled data sets, meaning that they exist to a ground-truth output (continuous or discrete) for every input. Conversely, in UL (Bishop (2006)) there is no ground-truth output, and the algorithms normally attempt to discover patterns in the data. Reinforcement Learning aims to raise the cumulative reward so that it is more suitable for sequential decision-making tasks (Zhang 2020b). Supervised learning has regression and classification; unsupervised learning includes cluster analysis and dimensionally reduction, also Reinforcement Learning (RL) includes classification and control, as illustrated in Fig. 1.

COVID-19 with machine learning

Recently there are three different perspectives of work that had been done on edge computing and the detection of (COVID-19) Cases. The viewpoints are including the recognizing of (COVID-19) cases by machine-learning systems (Table 1). The algorithms for the recognition of activity from machine learning and the approaches which used in edge computing are considered the Imaging workflows that can inspire machine-learning methods that are able of supporting radiologists who search for an analysis of complex imaging and text data. For the novel COVID-19 there are models capable of analyzing medical

imaging and recognizing COVID-19 (Shirzadi 2018). Artificial intelligence AI has various types, machine learning (ML), is one of these applications, it had been applied successfully to different fields of medicine for detection of new genotype–phenotype associations, diagnosis, which showed effects on assessment, prediction, diseases classification, transcriptomic, and minimizing the death ratio(Gao 2020).

The technique of automatic classification of COVID-19 can be applied by comparing general deep learning-based feature extraction frameworks to achieve the higher accurate feature, which is an important module of learning, MobileNet, DenseNet, Xception, ResNet, InceptionV3, InceptionResNetV2, VGGNet, NASNet were selected among a group of deep convolutional neural networks CNN. The classification then achieved by running the extracted features into some of machine-learning classifiers to recognize them as a case of COVID-19 or other diseases (Bishop 2006). Progressive machine-learning algorithms can integrate and evaluate the extensive data that is related to COVID-19 patients to provide best understanding of the viral spread pattern, increase the diagnostic accuracy, improve fresh, and effective methods of therapy, and even can recognize the individuals who, at risk of the disease depending on the genetic and physiological features (Khanday 2020).

Literature searching strategy and article selection

This systematic review paper used articles from online digital databases, which include Science Direct, Springer, Hindawi, and MDPI databases, two independent authors started the search strategy from October 2020 until

Table 1 Search strategy and paper selection process

Source	After query search	After applying the selection criteria	After quality assessment	After full article reading
Science direct	1254	440	24	4
Springer	5008	1549	217	5
Hindawi	10,000	3027	134	1
MDPI	44	38	20	4
Total	16,306	5054	395	14

Table 2 Supervised and un-supervised machine learning for analyzing the COVID-19 disease that included articles with the related details of the Dataset, author name, country of publication, year of publication, the used method in the study, and their results

<i>n</i>	Author	Year	Country	Dataset	Method	Tasks and Algorithms	Result
1	Khanday et al. (2020)	2020	India	GitHub 212 reports	Supervised learning	Classification Logistic Regression and Naive Bayes	The findings showed that Logistic regression and multinomial Naïve Bayes are better than the commonly used algorithms according to 96% accuracy obtained from the findings
2	Burdick et al. (2020a)	2020	USA	United States health systems 197 patients	Supervised learning	Classification Logistic Regression	Their results showed that this algorithm displays a higher diagnostic odds ratio (12.58) for foreseeing ventilation and effectively triage patients than a comparator early warning system, such as Modified Early Warning Score (MEWS) which showed (0.78) sensitivity, while this algorithm showed (0.90) sensitivity which leads to higher specificity ($p < 0.05$), also it shows the capability of accurate identification 16% of patients more than a commonly used scoring system which results in minimizing false-positive results
3	Varun et al. (2020)	2020	USA	184,319 reported cases	Supervised learning	Classifications Convolutional Neural Networks (CNN)	In response to this crisis, the medical and academic centers in New York City issued a call to action to leverage their electronic medical record (EMR) data to better understand SARS-COV-2 patients. Due to the scarcity of ventilators and a reported need for a quick and accurate method of triaging patients at risk for respiratory failure, our purpose was to develop a machine-learning algorithm for frontline physicians in the emergency department and the inpatient floors to better risk-assess patients and predict who would require intubation and mechanical ventilation

Table 2 (continued)

n	Author	Year	Country	Dataset	Method	Tasks and Algorithms	Result
4	Luca et al. (2020)	2020	Italy	85 chest X-rays	Supervised Learning	Classification K-nearest neighbors classifier (k-NN)	In the paper, we propose a method aimed to automatically detect the COVID-19 disease by analyzing medical images. We exploit supervised machine-learning techniques building a model considering a data set freely available for research purposes of 85 chest X-rays. The experiment shows the effectiveness of the proposed method in the discrimination between the COVID-19 disease and other pulmonary diseases
5	Constantin et al. (2020)	2020	Germany	152 datasets of COVID-19 patients, 500 chest CTs	Supervised learning	Classifications Convolutional Neural Network (CNN)	The findings showed that the combining between machine learning and a clinically embedded software developed platform allowed time-efficient development, immediate deployment, and fast adoption in medical routine. Finally they achieved the algorithm for fully automated segmentation of the lung and opacity quantification within just 10 days was ready for medical use and achieved human-level performance even for complex cases
6	Lamiaa et al. (2020)	2020	Egypt	COVID-19 5000 cases	Supervised learning	Regression Linear Regression model	The result showed that the designed models, such as the exponential, fourth-degree, fifth-degree, and sixth-degree polynomial regression models are brilliant especially the fourth-degree model which will benefit the government to prepare their procedures for 1 month. Furthermore, they introduced a well-known log that will grow up the regression model and will result in obtaining the epidemic peak and the last time of the epidemic during a specific time in 2020. Besides, the final report of the total size of COVID-19 cases

Table 2 (continued)

n	Author	Year	Country	Dataset	Method	Tasks and Algorithms	Result
7	Dan et al. (2020)	2020	Israel	6995 patients in Sheba Medical Center	Supervised learning	Classifications Artificial Neural Network (ANN)	The most contributory variables to the models were APACHE II score, white blood cell count, and time from symptoms to admission, oxygen saturation, and blood lymphocytes count. Machine-learning models demonstrated high efficacy in predicting critical COVID-19 compared to the most efficacious tools available. Hence, artificial intelligence may be applied for accurate risk prediction of patients with COVID-19, to optimize patients triage and in-hospital allocation, better prioritization of medical resources, and improved overall management of the COVID-19 pandemic
8	Joep et al. (2020)	2020	Netherlands	319 patients	Supervised learning	Classification Logistic regression	Chest CT, using the CO-RADS scoring system, is a sensitive and specific method that can aid in the diagnosis of COVID-19, especially if RT-PCR tests are scarce during an outbreak. Combining a predictive machine-learning model could further improve the accuracy of diagnostic chest CT for COVID-19. Further candidate predictors should be analyzed to improve our model. However, RT-PCR should remain the primary standard of testing as up to 9% of RT-PCR positive patients are not diagnosed by chest CT or our machine-learning model

Table 2 (continued)

n	Author	Year	Country	Dataset	Method	Tasks and Algorithms	Result
9	Christopher et al. (2020)	2020	Germany	368 independent variables	Supervised learning	Classifications Naive Bayes	They focused on variables and factors that increase the COVID-19 incidence in Germany depending on the multi-method ESDA tactic which provides a unique insight into spatial and spatial non-stationaries of COVID-19 occurrence, the variables, such as built environment densities, infrastructure, and socioeconomic characteristics all showed an association with incidence of COVID-19 in Germany after assessment by the county scale Their research outcome suggests that implementing social distancing and reducing needless travel can be important methods for reducing contamination
10	Hoyt et al. (2020b)	2020	U.S	290 patients	Supervised learning	Classification Logistic Regression	The findings showed that there is no correlation between the mortality and treatment in the entire population as the hydroxychloroquine was associated with a statistically significant ($p = 0.011$) rise in survival the adjusted hazard ratio was 0.29, 95% with a confidence interval (CI) 0.11–0.75. Although the patients who were indicted by the algorithm the adjusted survival was 82.6% in the treated group and 51.2% in the group who were not treated, after machine-learning applications the algorithm detected 31% of improving among the COVID-19 population which shows the important role of the machine-learning application in medicine

Table 2 (continued)

n	Author	Year	Country	Dataset	Method	Tasks and Algorithms	Result
11	María et al. (et al. 2020)	2020	International	Food for each of the 170 countries	Unsupervised learning	Clustering K-means clustering	The research findings stated that countries with the highest death ratio were those who had a high consumption of fats, while countries with a lower death rate have a higher level of cereal consumption followed by a lower total average intake of kilocalories
12	Shinwoo et al. (2020)	2020	U.S.A	790 Korean immigrants	Supervised learning	Classifications Artificial Neural Network (ANN)	Their result showed The Artificial Neural Network (ANN) analysis, which is a statistical model and able to examine complex non-linear interactions of variables, was applied. The algorithm perfectly predicted the person's flexibility, familiarities of everyday discernments, and the racism actions toward Asians in the U.S. since the beginning of the COVID-19 pandemic which finally provides important suggestions for public health practitioners (Zhang 2020b)
13	Yigrem et al. (2020)	2020	Southern Ethiopia	244 samples	Supervised learning	Classification Logistic Regression	Results showed that more than half of the research participants were presented with perceived stress of coronavirus disease, which means that there is a strong correlation between the health care staff and perceived stress of COVID-19
14	Abolfazl et al. (2020)	2020	USA	US Centers for Disease and Control and Johns Hopkins University. Database of 57 candidate	Supervised learning	Classification Artificial Neural Networks (ANN)	Results showed that the presented model (logistic regression) shown that these factors and variables describe the presence/absence of the hotspot of the COVID-19 incidence which was clarified by Getis-Ord Gi ($p < 0.05$) in a geographic information system. As a result, the findings provided valuable insights for public health decision makers in categorizing the effect of the potential risk factors associated with COVID-19 incidence level

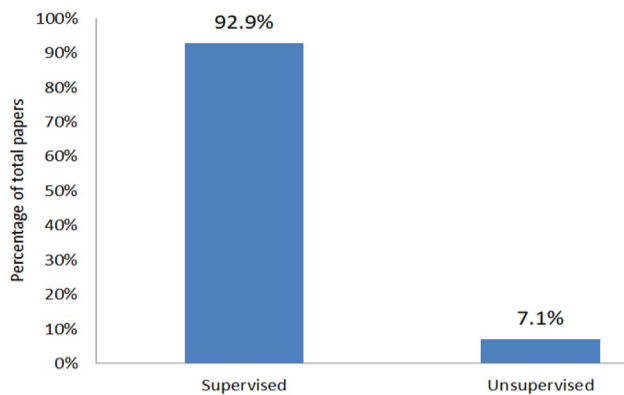


Fig. 2 Distribution of machine-learning types

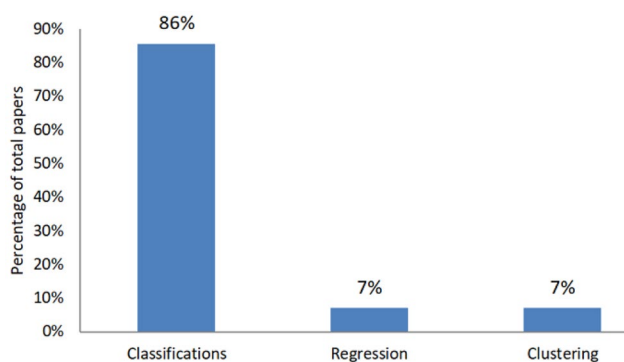


Fig. 3 Distribution of machine-learning tasks

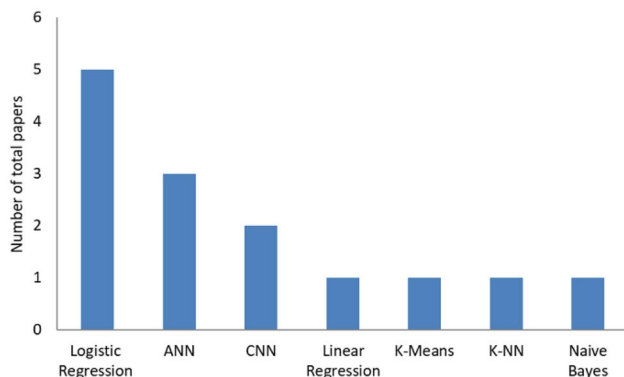


Fig. 4 Distribution of machine-learning algorithms

December 2020. The used keywords were “COVID-19; Machine Learning; Supervised Learning; Un-supervised Learning.” They were connected to the relevant articles using “and”, or “or” to find the studies that deals with human disease and COVID-19. The total number of the studies were (16,306) articles from all the databases, according to the inclusion and exclusion criteria this number was

limited. The limitation includes selecting the publication year (2019–2021), the articles type original articles that had been published as journal articles in English language only included. This selection strategy reduced the total number to 5054 articles, then after quality assessment of these studies there was 395 articles which remained, then finally the full text article reading minimized the last included articles to 14. The included articles are presented according to the author’s name, publication’s year, country, the used dataset, the applied method, and finally their results in (Table 2).

Machine-learning types applied

According to Fig. 2, supervised learning is the dominant machine-learning type applied for production lines. The majority of studies used both supervised learning methods which were (92.9%), whereas unsupervised learning was (7.1%).

Results

Machine-learning tasks addressed

Figure 3 shows that classification is the main task, which accounts for about (86%) of all selected papers. There are about (7%) of papers that applied for each of the regression and clustering.

Machine-learning algorithms used

Figure 4 shows that the logistic regression is largely applied in production lines. Logistic regression is the most frequently applied machine-learning algorithm, including five papers in 14 papers. Artificial neural network algorithm (ANN) and CNN (convolutional neural network) are in the second and third ranks which were three and two papers in 14 papers, respectively. Linear regression, K-Means, KNN (K-nearest neighbors), and Naive Bayes are the other algorithms applied for production lines.

Discussions and implications

The new transmitted virus was discovered and spread out from Wuhan city of China in December 2019 and affected more than (100) countries around the world in a very short time (Wu 2020). It was represented and introduced to the World Health Organization (W.H.O) on 31st December 2019. The virus was then termed COVID-19 by W.H.O on 11th February 2020, because it formed a global risk (Wu

2020). This family of viruses also includes SARS, ARDS. W.H.O confirmed this eruption as a public health emergency (Manigandan et al. 2020). Technology progressions have a fast effect on each field of life; the medical field is one of the important direct daily related to people's lives. Recently Artificial intelligence AI had been introduced to the medical field and it has shown promising outcomes in health care due to the high accuracy of data analysis which makes an exact decision making. Researchers all over the world tried to find a method to improve the clinical diagnosis and minimize the rapid spread of this virus so that they involved AI algorithms in the diagnosis of this disease. This review paper explains various AI algorithms that people used in their researches and will compare their results to demonstrate the best accurate method that shows the most improving in COVID-19 diagnosis. The total studies that used in this research are (14) original articles, all of them used supervised learning as the main method, but the algorithms were differed among them according to the research purpose.

A study recently published 2020 in India they extracted their dataset from GitHub which was 212 reports of 1000 cases, they used supervised learning as their main method in machine-learning application, and the algorithm that they applied was classification logistic regression and multinomial Naïve Bayes. The findings showed that Logistic regression and multinomial Naïve Bayes are better than the commonly used algorithms according to 96% accuracy obtained from the findings (Khanday 2020). Scientists in the USA published an article 2020 they relied on United States health systems to custom 197 patients as their data, the main method that they used was supervised learning, while the algorithm was classification logistic regression, their results showed that this algorithm displays higher diagnostic odds ratio (12.58) for foreseeing ventilation and effectively triage patients than a comparator early warning system, such as Modified Early Warning Score (MEWS) which showed (0.78) sensitivity, while this algorithm showed (0.90) sensitivity which leads to higher specificity ($p < 0.05$), also it shows the capability of accurate identification 16% of patients more than a commonly used scoring system which results in minimizing false-positive results (Burdick 2020a). Varun et al. (2020) used 184,319 reported cases as a dataset in his article in which he applied the same method supervised learning but with a different algorithm which was convolutional neural network CNN and their outcomes were in response to this crisis, the medical and academic centers in New York City issued a call to action to artificial intelligence researchers to leverage their electronic medical record (EMR) data to better understand SARS-COV-2 patients. Due to the scarcity of ventilators and a reported need for a quick and accurate method of triaging patients at risk for respiratory failure, our purpose was to develop a machine-learning algorithm for frontline physicians in the emergency

department and the inpatient floors to better risk-assess patients and predict who would require intubation and mechanical ventilation (Arvind 2020). Meanwhile, another study had been published in Italy by (Luca et al. 2020) who used also supervised learning in their methodology but they used a different algorithm this time called K-nearest neighbors classifier K-NN, their research results showed that the proposed method that aims to detect the COVID-19 disease by analyzing medical images by building a model allowing an easily data set availability for research purposes using 85 chest X-rays. The research shows the effectiveness of the proposed method in the discrimination between the COVID-19 disease and other pulmonary diseases (Brunese 2020). Constantin et al. (2020) published an article in Germany he depended on 152 datasets of COVID-19 patients and 500 chest CT scans, he also relied on supervised learning but using Neural Network Algorithm for analyzing these data. Their findings showed that the combining between machine learning and a clinically embedded software developed platform allowed time-efficient development, immediate deployment, and fast adoption in medical routine. Finally, they achieved the algorithm for fully automated segmentation of the lung, and opacity quantification within just 10 days was ready for medical use and achieved human-level performance even for complex cases (Anastasopoulos 2020). Far away from Europe and the USA, a study conducted by (Amar et al. 2020) in Egypt depended on 5000 COVID-19 cases as a dataset. They had chosen supervised learning as their method than using regression analysis as the selected algorithm. The result showed that the designated models, such as the exponential, fourth-degree, fifth-degree, and sixth-degree polynomial regression models are brilliant especially the fourth-degree model which will benefit the government to prepare their procedures for 1 month. Furthermore, they introduced a well-known log that will grow up the regression model and will result in obtaining the epidemic peak and the last time of the epidemic during a specific time in 2020. Besides, the final report of the total size of COVID-19 cases (Amar et al. 2020). Researchers in Israel presented research by (Dan et al. 2020) they extracted 6995 patient reports from Sheba Medical Center to be used as research data, they also used supervised learning as the main method, and then they selected the artificial neural network ANN as the used algorithm in their study, depending on the patient biography it had been demonstrated that APACHE II score, white blood cell WBC count, duration from symptoms to admission, oxygen saturation and blood lymphocytes count were the most related variables to the used models. The findings demonstrated that Machine-learning (ML) models showed high efficiency in predicting serious COVID-19 as compared to the other efficient tools available. Hereafter, the results suggested artificial intelligence be applied for accurate risk estimation of COVID-19 patients, to

enhance patient triage (Assaf 2020). In a study conducted by (Hermans et al. 2020) in the Netherlands, their article used 319 patients as the dataset and they selected supervised learning as their method, while the logistic regression was the selected algorithm. In this article, they depended on the patient's chest CT scan scores, and the RT-PCR test the results demonstrated that Chest CT, using the CO-RADS scoring system, is a specific useful method that can lead to accurate diagnosis of COVID-19, particularly if RT-PCR tests are uncommon during an epidemic. Also merging a predictive machine-learning model may more improve the diagnosis accuracy of chest CT scans for COVID-19 patients. Nevertheless, they recommended RT-PCR must remain as the primary standard of testing, because up to 9% of patients with positive RT-PCR were not identified by chest CT or the presented machine-learning model (Hermans 2020). In Germany, Christopher et al. (2020) used 368 independent variables as a sample size in their article which built its methodology on supervised learning, and the model was Bayesian machine-learning analysis. They focused on variables and factors that increase the COVID-19 incidence in Germany depending on the multi-method ESDA tactic which provides a unique insight into spatial and spatial non-stationaries of COVID-19 occurrence, the variables, such as built environment densities, infrastructure, and socioeconomic characteristics all showed an association with incidence of COVID-19 in Germany after assessment by the county scale. Their research outcome suggests that implementation social distancing and reducing needless travel can be important methods for reducing contamination (Scarpone 2020). Hoyt et al. (2020b) presented an article that depended on the data obtained from 290 patients to use supervised learning in their article and the logistic regression as the specific algorithm, to find the correlation between the treatment and the mortality in the entire 290 population that is infected by COVID-19 in the USA by detecting the hazards on the entire population the 290 patients who enrolled in their research and also on the subpopulation who prepared for the suitable treatment identified by the algorithm. The findings showed that there is no correlation between the mortality and treatment in the entire population as the hydroxychloroquine was associated with a statistically significant ($p=0.011$) rise in survival the adjusted hazard ratio was 0.29, 95% with a confidence interval (CI) 0.11–0.75. Although the patients who were indicted by the algorithm the adjusted survival was 82.6% in the treated group and 51.2% in the group who were not treated, after machine-learning applications the algorithm detected 31% of improving among the COVID-19 population which shows the important role of the machine-learning application in medicine (Burdick 2020b). Reichberg et al. (2020) used the international program food for 170 countries as a source of their research using unsupervised learning and specifically the

K-means clustering algorithm to find the association between obesity and mortality in the COVID-19 countries.

The research findings stated that countries with the highest death ratio were those who had a high consumption of fats, while countries with a lower death rate have a higher level of cereal consumption followed by a lower total average intake of kilocalories (García-Ordás, et al. (2020)). A study conducted to (Shinwoo et al. 2020) their research data were extracted from the immigrant Korean COVID-19 patients who were 290 cases from 12 states all of them older than 18 years, the study observed the ability to the prediction of discrimination-related variables, such as racism effects, and sociodemographic factors that influence the psychological distress level during the COVID-19 pandemic, they nominated the supervised learning as the method and then using the Artificial Neural Network ANN as the main algorithm, their result showed The Artificial Neural Network (ANN) analysis, which is a statistical model and able to examine complex non-linear interactions of variables, was applied. The algorithm perfectly predicted the person's flexibility, familiarities of everyday discernments, and the racist actions toward Asians in the U.S. since the beginning of the COVID-19 pandemic which finally provides important suggestions for public health practitioners (Choi 2020). During the same time, a study presented by (Yigrem et al. 2020) conducted a cross-study based on 244 of the healthcare providers in Dilla, Southern Ethiopia. Supervised learning was used in the methodology and then they analyzed the data by logistic regression algorithm to find the association between the perceived stress of COVID-19 and the health care providers. Results showed that more than half of the research participants were presented with perceived stress of coronavirus disease, which means that there is a strong correlation between the health care staff and perceived stress of COVID-19 (Chekole, et al. (2020)). Finally, the last article conducted by (Abolfazl et al. 2020) their study used 57 samples of COVID-19 cases from the USA to find out the relationship between the sociodemographic and environmental variables, other diseases, such as chronic heart disease, leukemia, and pancreatic cancer, also socioeconomic factors and the death ratio due to COVID-19 disease. Results showed that the presented model (logistic regression) shown that these factors and variables describe the presence/absence of the hotspot of the COVID-19 incidence which was clarified by Getis-Ord Gi ($p < 0.05$) in a geographic information system. As a result, the findings provided valuable insights for public health decision makers in categorizing the effect of the potential risk factors associated with COVID-19 incidence level (Mollalo et al. 2020).

Conclusion

This study focused on the articles that applied machine-learning applications in COVID-19 disease for various purposes with different algorithms, 14 from 16 articles used supervised learning, and only one among them used unsupervised learning another one used both methods supervised and unsupervised learning and both of them shows accurate results. The studies used different machine-learning algorithms in different countries and by different authors but all of them related to the COVID-19 pandemic, (5) of these articles used Logistic regression algorithm, and all of them showed promising results in the COVID-19 health care applications and involvement. While (3) of the articles used artificial neural network (ANN) which also shows successful results, the rest of the 14 articles used different supervised and unsupervised learning algorithms and all of the models showed accurate results. Our conclusion is ML applications in medicine showed promising results with high accuracy, sensitivity, and specificity using different models and algorithms. In general, the paper results explored the supervised learning is more accurate to detect the COVID-19 cases which were above (92%) compare to the unsupervised learning which was mere (7.1%).

Funding This study was self-funded.

Declarations

Conflict of interest There are no conflicts of interest.

References

- Alimadadi A et al (2020) Artificial intelligence and machine learning to fight COVID-19. American Physiological Society, Bethesda
- Amar LA, Taha AA, Mohamed MY (2020) Prediction of the final size for COVID-19 epidemic using machine learning: a case study of Egypt. *Infect Dis Model* 5:622–634
- Anastasopoulos C et al (2020) Development and clinical implementation of tailored image analysis tools for COVID-19 in the midst of the pandemic: the synergetic effect of an open, clinically embedded software development platform and machine learning. *Eur J Radiol* 131:109233
- Arvind V et al (2020) Development of a machine learning algorithm to predict intubation among hospitalized patients with COVID-19. *J Crit Care* 62:25–30
- Assaf D et al (2020) Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern Emerg Med* 15(8):1435–1443
- Bishop CM (2006) *Pattern recognition and machine learning*. Springer, Berlin
- Brinati D et al (2020) Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. *J Med Syst* 44:135
- Brunese L et al (2020) Machine learning for coronavirus COVID-19 detection from chest x-rays. *Proced Comput Sci* 176:2212–2221
- Burdick H et al (2020a) Prediction of respiratory decompensation in Covid-19 patients using machine learning: the READY trial. *Comput Biol Med* 124:103949
- Burdick H et al (2020b) Is machine learning a better way to identify COVID-19 patients who might benefit from hydroxychloroquine treatment? The Identify Trial. *J Clin Med* 9(12):3834
- Chekole YA et al (2020) Perceived Stress and Its Associated Factors during COVID-19 among Healthcare Providers in Ethiopia: a cross-sectional study. *Adv Public Health*. <https://doi.org/10.1155/2020/5036861>
- Choi S et al (2020) Predicting psychological distress amid the COVID-19 pandemic by machine learning: discrimination and coping mechanisms of Korean Immigrants in the US. *Int J Environ Res Public Health* 17(17):6057
- Day M (2020) Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ* 368:135
- Deng X et al (2020) A classification–detection approach of COVID-19 based on chest X-ray and CT by using keras pre-trained deep learning models. *Comput Model Eng Sci* 125(2):579–596
- Doanvo A et al (2020) Machine learning maps research needs in covid-19 literature. *Patterns* 1(9):100–123
- Gao K et al (2020) Julia language in machine learning: algorithms, applications, and open issues. *Comput Sci Rev* 37:100254
- García-Ordás MT et al (2020) Evaluation of country dietary habits using machine learning techniques in relation to deaths from COVID-19. *Healthcare* 8:371
- Hastie TR, Tibshirani JF (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer Science and Business Media, Berlin
- Hermans JJ et al (2020) Chest CT for triage during COVID-19 on the emergency department: myth or truth? *Emerg Radiol* 27(6):641–651
- Hossain B et al (2019) Surgical outcome prediction in total knee arthroplasty using machine learning. *Intell Autom Soft Comput* 25(1):105–115
- Jamshidi M et al (2020) Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. *IEEE Access* 8:109581–109595
- Kassani SH et al (2020) Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: a machine learning-based approach 10(4):1–18
- Khan MA et al (2020) Intelligent cloud based heart disease prediction system empowered with supervised machine learning. *CMC Comput Mater Cont* 65(1):139–151
- Khanday AMUD et al (2020) Machine learning based approaches for detecting COVID-19 using clinical text data. *Int J Inf Technol* 12(3):731–739
- Mahmood A et al (2020) COVID-19 and frequent use of hand sanitizers; human health and environmental hazards by exposure pathways. *Sci Total Environ* 742(44):140561
- Manigandan S, Ming-Tsang W, Vinoth KP, Vinay BR, Arivalagan P, Kathirvel B (2020) A systematic review on recent trends in transmission, diagnosis, prevention and imaging features of COVID-19. *Process Biochem* 98(11):233–240. <https://doi.org/10.1016/j.procbio.2020.08.016>
- Mollalo A, Rivera KM, Vahedi B (2020) Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States. *Int J Environ Res Public Health* 17(12):4204
- Naseem M et al (2020) Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: a Scoping review. *J Primary Care Community Health* 11:1–11

- Pasupuleti RR et al (2021) Rapid determination of remdesivir (SARS-CoV-2 drug) in human plasma for therapeutic drug monitoring in COVID-19-Patients. *Process Biochem* 102(3):150–156
- Reichberg SB, Mitra PP, Haghmagad A, Ramrattan G, Crawford JM, Northwell COVID-19 Research Consortium et al (2020) Rapid emergence of SARS-CoV-2 in the greater New York metropolitan area: geolocation, demographics, positivity rates, and hospitalization for 46 793 persons tested by northwell health. *Clin Infect Dis* 71(12):3204–3213
- Scarpone C et al (2020) A multimethod approach for county-scale geospatial analysis of emerging infectious diseases: a cross-sectional case study of COVID-19 incidence in Germany. *Int J Health Geogr* 19(1):1–17
- Shinde GR et al (2020) Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. *SN Comput Sci* 1(4):1–15
- Shirzadi A et al (2018) Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* 18(11):3777
- Tayarani-N, M.-H (2020) Applications of artificial intelligence in battling against Covid-19: a literature review. *Chaos Solitons Fractals* 142:110338
- Ünlü R, Namlı E (2020) Machine learning and classical forecasting methods based decision support systems for COVID-19. *CMC Comput Mater Cont* 64(3):1383–1399
- Wu F et al (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798):265–269
- Zhang C et al (2020a) Applying feature-weighted gradient decent k-nearest neighbor to select promising projects for scientific funding. *CMC Comput Mater Cont* 64(3):1741–1753
- Zhang Y et al (2020b) Overview on routing and resource allocation based machine learning in optical networks. *Opt Fiber Technol* 60:102355

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.