ACTA FACULTATIS
MEDICAE NAISSENSIS

UDC: 004.85:[616.98:578.834
DOI: 10.5937/afmnai39-38354

Review article

Application of Machine Learning in the Fight against the COVID-19 Pandemic: A Review

Alem Čolaković, Elma Avdagić-Golub, Muhamed Begović, Belma Memić, Adisa Hasković Džubur

University of Sarajevo, Faculty of Traffic and Communications, Sarajevo, Bosnia and Herzegovina

SUMMARY

Introduction: Machine learning (ML) plays a significant role in the fight against the COVID-19 (officially known as SARS-CoV-2) pandemic. ML techniques enable the rapid detection of patterns and trends in large datasets. Therefore, ML provides efficient methods to generate knowledge from structured and unstructured data. This potential is particularly significant when the pandemic affects all aspects of human life. It is necessary to collect a large amount of data to identify methods to prevent the spread of infection, early detection, reduction of consequences, and finding appropriate medicine. Modern information and communication technologies (ICT) such as the Internet of Things (IoT) allow the collection of large amounts of data from various sources. Thus, we can create predictive ML-based models for assessments, predictions, and decisions.
Methods: This is a review article based on previous studies and scientifically proven knowledge. In this paper, bibliometric data from authoritative databases of research publications (Web of Science, Scopus, PubMed) are combined for bibliometric analyses in the context of ML applications for COVID-19.
Aim: This paper reviews some ML-based applications used for mitigating COVID-19. We aimed to identify and review ML potentials and solutions for mitigating the COVID-19 pandemic as well as to present some of the most commonly used ML techniques, algorithms, and datasets applied in the context of COVID-19. Also, we provided some insights into specific emerging ideas and open issues to facilitate future research.
Conclusion: ML is an effective tool for diagnosing and early detection of symptoms, predicting the spread of a pandemic, developing medicines and vaccines, etc.

Keywords: **machine learning, COVID-19 pandemic, COVID-19 datasets, artificial intelligence**

Corresponding author:
Alem Čolaković
e-mail: alem.colakovic@fsk.unsa.ba

INTRODUCTION

Machine learning (ML) algorithms and methods are increasingly used in various fields of medicine. ML-based tools provide valuable features for fighting the COVID-19 pandemic (officially known as SARS-CoV-2). ML-based approaches use computer algorithms to learn relationships among different data elements and events to obtain outcomes by identifying relationships between various events. For example, ML-based approach is useful for extracting the relationship between the number of COVID-19 patients and different variables such as local regulations and restrictions.

Over the past two years, many web-based and mobile applications that use ML algorithms have been developed to solve various problems related to the pandemic. Trends show that ML will play an increasing role in providing predictive and preventive health care in the future. This paper focuses on the latest research related to the application of ML for mitigating COVID-19. We briefly outlined the role of ML for a majority of aspects to manage the COVID-19 pandemic including diagnosis, prediction, medicine development, etc. Also, we highlighted some ML datasets and algorithms used to mitigate the COVID-19 pandemic in various prospects (e.g. early detection and diagnosis, projections of cases, development of medicine, etc).

ML is the branch of AI (Artificial Intelligence) used for developing various tools used in the fight against the COVID-19. The potentials of ML-based tools for this purpose are extensive (1, 2). Its significance for developing screening, predicting, forecasting, and contact tracing mechanisms to improve the decision-making in the fight against the COVID-19 pandemic has been proved. The existing clinical prediction models (e.g. Apache, Gleason, PASI, etc.) capture a relatively limited number of variables. Due to many infected people with different symptoms around the globe, it is necessary to collect and analyze a huge amount of data. ML surpasses the existing clinical prediction models in the features and performances while analyzing big data sets and variables. Some ML-based tools enable real-time analysis of data sets consisting of complex and diverse raw data. We can use ML algorithms to develop a virus spread model in a given area and to predict the future number of infections (3). Besides, ML helps to detect people with the highest risk of infection (4), develop inexpensive diagnostic tools (5), and to help frontline workers and decision-makers in various ways (6). Some ICT solutions (e.g. GPS) enable tracking the travel history of infected patients. This feature allows collecting data and monitoring the spread of the virus. Furthermore, ML-based solutions can use these data to search for contacts required to control the spread of the virus (7).

In the field of medical diagnostics, ML has enabled the development of software for detecting the signs of COVID-19 based on chest CT or CXR of the patient's chest. For example, ML enables active surveillance for COVID-19 based on speech or cough recognition techniques using a mobile application (8). Authors Hassan H, et al. (9) analyzed the previous review of AI-enabled COVID-19 CT imaging models. The first steps were the development of a revolutionary neural network-based system called COVID-Net, which aims to distinguish the cases of COVID-19 from other diseases by analyzing lungs on X-rays (10).

ML plays a significant role in the development of medicines or vaccines. It has a particularly significant role in researching the behavior of existing medicines for similar viral diseases (11). Some ML-based tools enable to personalize the treatment of patients and to automate some procedures, which in some cases is crucial due to shortening the response time. The collection of relevant data is a prerequisite for the application of ML. The quantity and quality of data determine the possibility of efficient use of ML algorithms. ML can use COVID-19 data sets in many ways to combat a pandemic. Therefore, the paper highlights this issue and presents the significance of collecting COVID-19 data from various sources. The second chapter explains the significance of ML and COVID-19 data sets. In this chapter, we compare some ML-based algorithms. The third chapter provides an overview of ML roles in the fight against the COVID-19 pandemic. Besides, we consider specific examples of the application of ML-based algorithms in current practice. The fourth chapter contains a brief overview of some ML-based solutions which use various data sources and types such as respiratory data, blood samples, radiography images, etc. The fifth chapter contains concluding remarks and some ideas for future research. Therefore, we provide insights into specific emerging ideas to facilitate future research.

MACHINE LEARNING

Machine Learning (ML) is a sub-domain of Artificial Intelligence (AI) that uses statistical techniques to learn from data samples to predict future outcomes without prior knowledge (12). ML methods are a powerful analytical tool for discovering and applying hidden knowledge in solving practical problems in many empirical disciplines. The process of ML-based tools application is complex. Most of the tasks in the development process are learning-related issues, so different ML algorithms can be applied to solve these problems. These algorithms are efficient when exact solving methods are unknown. Besides, a large number of implicit interdependencies are documented in databases. Therefore, COVID-19 tools must operate in an ever-changing environment. We can use different ML approaches to solve these problems. Figure 1 presents the differences between major categories of ML.

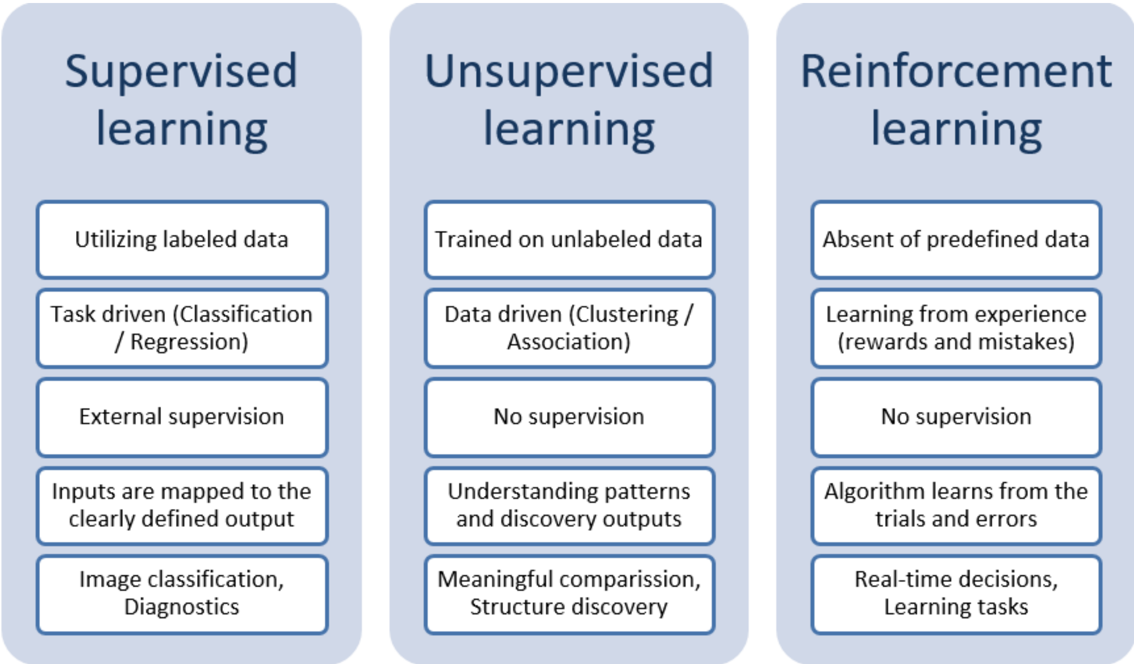

Figure 1. *Differences between major categories of machine learning*

Supervised learning is the easiest type of ML which is based on data labels and where the learning algorithm is data-dependent. (e.g. image classification is based on labeled images). The ML algorithm learns from considering the training set and then inferring the subject of an unknown image. This problem is easy to frame as a supervised learning problem. Therefore, this technique is a very efficient tool in the fight against COVID-19. For example, SVM (Support Vector Machine) or NN (Neural Network) as a supervised learning technique can be used for radiography image (CT, CXR) classification as well as for developing diagnostic tools. Therefore, supervised learning includes two groups of problems: classification and regression. The classification algorithm predicts categorical values and identifies the category of new observations. The most common ML algorithms used in practice for classification problems are Naive Bayes classifier, Decision Trees, Logistic Regression, K-Nearest Neighbours, Support Vector Machine (SVM), Random Forest, Neural Networks (NN), etc. The regression algorithms predict the outcomes for continuous values. Some of the most popular ML-based algorithms for regression problems are Simple Linear Regression (SLR), Multiple Linear Regression (MLR), Polynomial Regression (PR), Decision Tree Regression (DTR), Random Forest Regression (RFR), Ensemble Method, NN, etc. Unsupervised learning is a technique based on an unlabeled bunch of data where the algorithm creates groups based on discovered patterns in the data. The strategy of this approach is data-dependent, and the algorithm is based on clustering and association while finding covered structures of given data. For example, K-means is unsupervised technique that can be used for clustering of infected COVID-19 patients' vs non- COVID-19 cases. Therefore, Unsupervised learning is used to solve clustering problems and anomaly detection. The most popular ML algorithms for clustering are K-means, Gaussian Mixture, Hidden Markov, DBSCAN Clustering, Agglomerative Hierarchical Clustering, NN, etc.

Reinforcement learning is a crossbreed of the previous two approaches. It is based on time-delay labels (rewards). These rewards are given to an algorithm that learns to interact in an environment (learns from experience) that supports real-time decision-making. For example, MDP (Markov Decision Process) approach has the advantage that patterns are discovered only by performing clinical experimentation while utilizing cooperative involvements and evaluating inputs.

ML DATASETS FOR COVID-19

ML-based tools for COVID-19 mitigating require different data types stored in the COVID-19 data repository. These data include symptoms, laboratory tests, demographics, etc. Data extraction can be manual, automated or a combination of both (13). The data repository can be enhanced by pre-processing data that includes data cleaning, noise removal, feature extraction, etc. ML-based techniques enable the utilization of the COVID-19 data sets to help in the fight against irresistible illnesses like COVID-19.

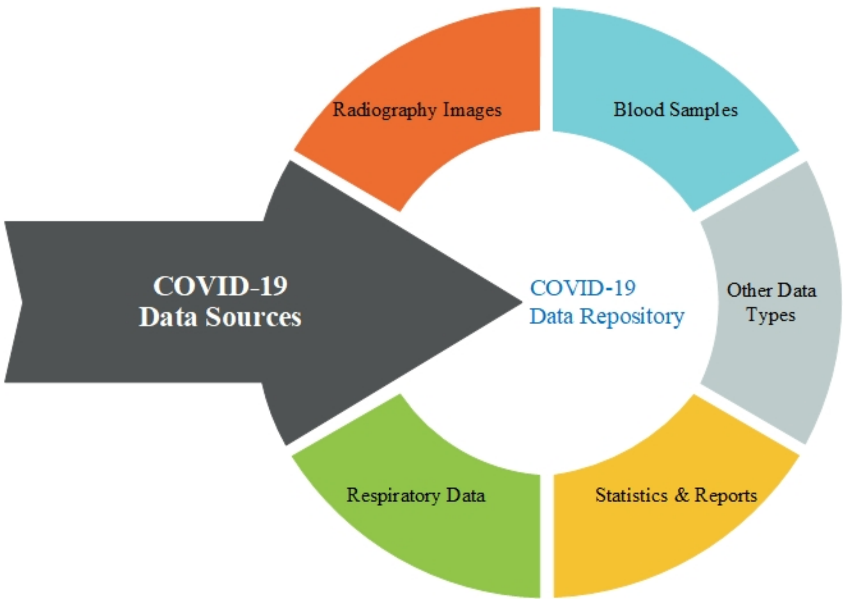

Figure 2. *The mostly used data sources for COVID-19 outbreak forecasting models*

ML enables to process a large amount of raw data. Mostly used COVID-19 data type are radiographic images (CT, CXR), blood samples, respiratory data (breathing and coughing information), and COVID-19 reports (Figure 2). Also, there are other data sources such as patient's health records, patients' self-reporting, government reports, social media (e.g. Twitter), Google Trends, etc. For example, authors Hou Z et al. in (14) collected data from social platforms used in Chine (e.g. Baidu search engine, Sina Weibo, Ali e-commerce). Baoquan et. al. (15) proposed a mathematical model based on public datasets to evaluate the effectiveness of some measures against COVID-19 transmission.

IoT (Internet of Things) enables effective tools to collect data. IoT technologies enable the automatic collection of large amounts of data generated from various sources including ubiquitous wearable devices equipped with sensors. IoT enables data collection required for first-line detection and tracking such as temperature, location, imaging, cough, etc. Ndiaye et al. (16) provide an overview of the IoT implementation targeted for mitigating the effects of COVID-19. They emphasize the significance of IoT

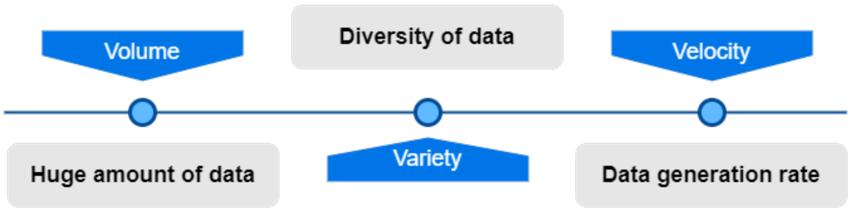

Figure 3. *Main characteristics of big data*

Table 1. *Most commonly used datasets for COVID-19*

Dataset	Type
Kaggle	Chest X-ray and CT images
Kaggle 2	Chest X-ray images
Cohen	Chest X-ray and CT images
Paul Mooney	Chest X-ray
COVIDx	CT images
ChestXray-8	Chest x-ray images
CheXpert	Chest radiographs

for monitoring patients and data collection. Also, a mobile phone equipped with biosensors and cameras can collect some personal data (e.g. images, cough sound, heart rate). In a study (17) authors. provide a review of recent progress in optical biosensors based on smartphone platforms. COVID-19 data sets consist of various data types required for the efficient application of ML-based tools for mitigating pandemic. The different data sources aggregation provides data asset characterized by high volume, velocity, and variety is known as big data (Figure 3). This process provides a huge potential to support the fight against COVID-19 (18).

Chest CT and CXR scans are frequently used data types for ML-based automatic COVID-19 diagnosis (19). Besides, some research efforts include other data types such as atmospheric temperature and humidity to analyze COVID-19 virus transmission (20). One analysis includes patients' geographical and travel data (21) or COVID-19 patients' statistics such as the total number of positive cases, recoveries, and fatal outcomes (22). COVID-19 data types can be in various formats including images (e.g. CT and CXR), speech (e.g. cough and breath), statistics, and textual data (e.g. COVID-19 case reports, geographical data, etc.).

Another role of COVID-19 data sets is to monitor the virus spread and to discover possible patterns between virus spread and parameters such as the enactment of regulations and rules (policy). For this purpose, authorities such as government institutions, health institutions, international organizations, etc., provide various COVID-19 data sets. These data sets are required to develop ML-based models for different purposes such as tracking the spread of COVID-19 (23).

The corresponding data sets is necessary for the efficient application of ML-based tools in the fight against COVID-19. For example, detection of COVID-19 symptoms from cough samples requires a set of data from healthy and infected people to perform a comparative analysis to distinguish the characteristics of an infected person from a healthy person. We can use different publicly available data sets to train the model and to mitigate the COVID-19 pandemic in various ways. The most of open database of COVID-19 cases include X-ray and CT images such as Cohen et al. (24). Table 1 presents the most commonly used datasets based on X-ray and CT images.

In this paper, we briefly describe three scientifically validated data sets. These data sets are open for public use and span the global level. However, several other data sets are based on restricted access. Some data sets consist of data from a limited geographic area (local sets) while others cover the global aspect. The most notable data set showing the COVID-19 pandemic globally was presented by John Hopkins University. A real-time interactive COVID-19 based web map is available from the beginning of 2020 (25). The interactive web map shows the number of COVID-19 cases, deaths, and recoveries grouped into regions of states/provinces. The data set contains more details for the USA, Canada, and Australia. The corresponding Github data warehouse is also provided by the Center for Systems Science and Engineering (CSSE) (25). Data collection is semi-automated with data collected from different sources being updated every 15 minutes. This data set aimed to provide the public, authorities, healthcare professionals, and researchers with a user-friendly tool to monitor, analyze and model the spread of COVID-19. For example, Pandey G et al. (26) used this data set, mathematical SEIR (Susceptible, Exposed, Infectious, Recovered) model and Regression model for some COVID-19 predictions.

Our World in Data (OWD) provides the COVID-19 data set and updates the data several times a day. Key data sources for this repository are various reports snd statistics provided by national governments and institutions. This data set includes data on positive COVID-19 cases, deaths, and statistics about testing. The list of data sources and all data for each country is available on the Github repository website (27). OWD maintains a complete COVID-19 dataset. Also, the collection of official COVID-19 data sets is available at the ECDC (European Center for Disease Prevention and Control) website (28). This data collection includes COVID-19 figures from over 500 sources in 196 countries. ECDC gathering data from the official websites of the Ministries of Health, public health institutes, governments, health statistics, World Health Organization, etc. Data from official social media accounts such as Twitter, Facebook, YouTube, or Telegram from national authorities are also included.

In the context of COVID-19 data sets, we need to emphasize the advantages of the Kaggle platform. Kaggle provides the ability to find datasets from all research areas, publish datasets, research, and model in a web-based environment, work with data science experts, and participate in competitions to solve various problems using data science. By simply searching the Kaggle platform for COVID-19 datasets, it is possible to get almost 40,000 different results.

ML ALGORITHMS COVID-19

ML helps to discover patterns and trends from a large set of data. It enables the identification of cause-and-effect relationships between various events. For example, algorithms such as SVM, Decision Tree, Random Forest, etc. support the process of discovering the pattern that occurs in large data sets. Stakeholders can make corresponding decisions based on these patterns (29). Efficiency and accuracy increase as the number of data increases (30). Figure 4 shows a workflow of generic ML-based COVID-19 mitigation tools.

The first step is to identify data sources and collect the required dataset including radiography images (CT, CXR), respiratory data (cough, sneezing, breathlessness, etc.), blood samples, various statistic data (positive cases, recovered patients, fatal outcomes, etc.), and other data types (e.g. location, user self-tests...). All data are stored in a data warehouse and pre-processed to prepare further processing, extractions, analysis, and visualization. Data preprocessing enhances the raw data collection and transforms it into more meaningful representations. Pre-processing techniques include data cleaning, finding missing values, feature extraction, segmentation, noise removal, feature analysis, etc. Data mining and ML techniques utilize the COVID-19 datasets for diagnosis, forecasting, and mitigating the impact of the pandemic. The data repository archives COVID-19 data sets and enables sharing among different stakeholders. The corresponding data set is required to train the ML model. This process is required to understand the patterns, rules, and features. Besides, testing the model enables checking its accuracy. ML-based algorithms provide further processing and retrieval of various outcomes. Model deployment means integrating an ML-based model into a production environment to return out-

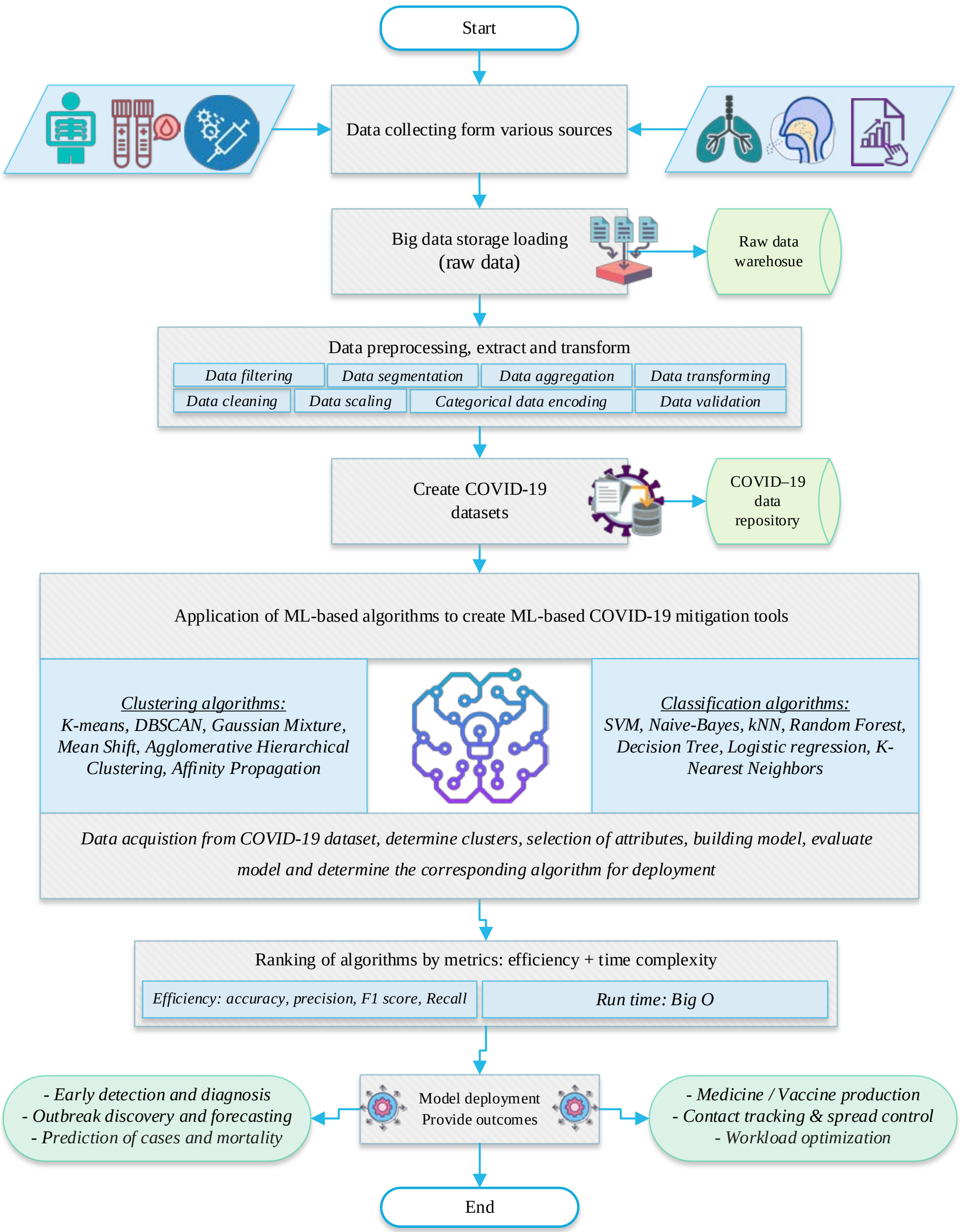

Figure 4. *A generic workflow of ML-based COVID-19 mitigation tools*

puts. These outcomes help the stakeholders take action and make the corresponding decisions in the fight against COVID-19. It is necessary to continuously monitor the model for errors, crashes, and time to ensure that model is maintaining the desired accuracy and performance.

ML techniques can clean the raw data for further processing. It processes various types of COVID-19 data such as images (e.g. X-ray images, CT images), blood samples, GPS, etc. The most commonly used ML algorithms are K-means, Naïve Bayes, K-nearest neighbors, Principal Component Analysis (PCA), Support Vector Machine (SVM), Random Forest (31). These algorithms have different functions and capabilities for data processing. Table 2 provides a comparative analysis of these ML algo-

Table 2. *Comparison of ML algorithms based on their advantages and disadvantages in the context of COVID-19*

Algorithm	Advantages	Disadvantages
K-means	Better for larger data sets because it makes stronger clusters	Determining the k-value is very difficult, especially for data of different densities
Naïve Bayes	Simple to implement, and does not require a large amount of data	The initial assumption is often wrong, and it is difficult to find the right one
K-nearest neighbors	It responds well to sparse data	The value for k must be correct, and this is difficult to assess
SVM	The data can be visualized in more than two dimensions, and the decision-making system is almost perfect	Choosing the kernel is not easy and also if the data set is large, it will take some time to train
PCA	If there is no data redundancy, the complexity of the solution is less	It is difficult to make a covariance matrix, even the simplest cases may seem impossible
Neural Networks	The ability to learn in the case of large data sets is fascinating especially for modeling nonlinear and complex relationships	Weight parameters are very difficult to calculate and a large data set is required
Random Forest	It is good for cases where there are thousands of input variables	It does not show good performance if there is a large number of irrelevant attributes in the data set

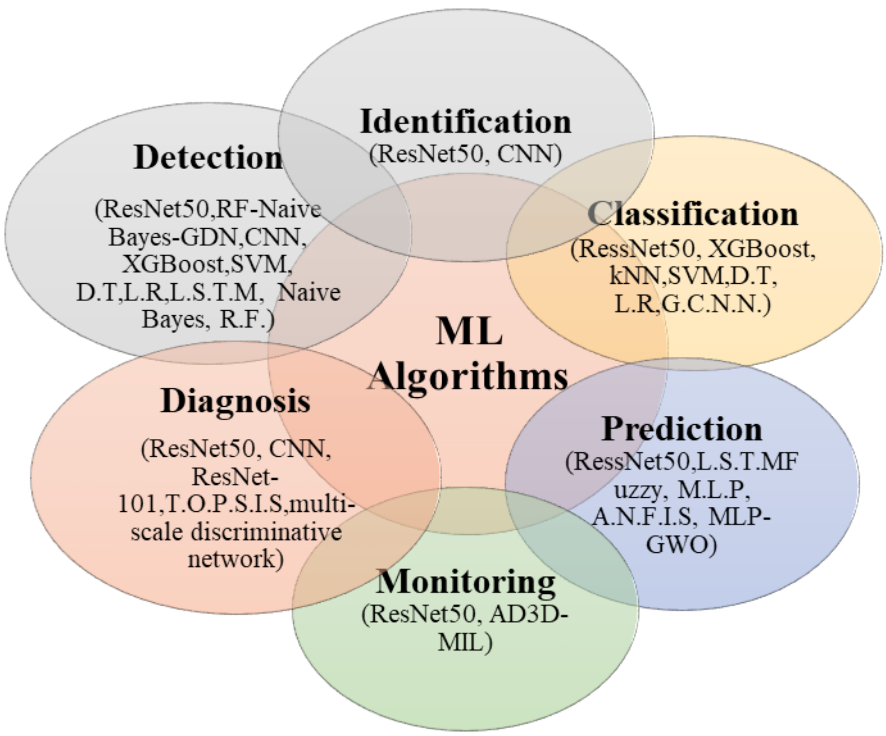

Figure 5. *Comparison of ML algorithms in COVID-19 applications*

rithms according to their advantages and disadvantages for processing large COVID-19 data sets.

The development of ML-based tools is demanding due to various challenges such as memory and processor limitations, diversity and unstructured data, simultaneous use of data from multiple sources, etc. Over the past few years, significant technological advances have been made to overcome some of these problems. For example, Qui, Wu. Q, Ding. G, Feng (32) made suggestions for more efficient application of the following algorithms: SVM, regression, and neural networks for big data processing, as well as for the problem of incomplete data. J. L. Berral Garcia (33) made some improvements to the classification as well as prediction process. Wu, W. Cheng, Hoffman, and Wang (34) presented a model of health analytics. However, this is an open topic for future research and further improvements (35). One of the trends in ML that has arisen due to the inability to choose the best algorithm is combining the results of multiple algorithms that achieve better accuracy. This trend is known as Ensemble Learning. There are four methods of combining multiple models: adding parameters, amplifying, stacking, and debugging one algorithm to another. This approach enables the development of new tools to combat the COVID-19 pandemic.

The problems of applying ML algorithms in the fight against the pandemic can be classified into six applications: identification, classification, prediction, monitoring, detection, and diagnosis. Different methods are used for each of these applications depending on the nature of the problem, as shown in the Figure 5. We notice that ResNet stands out the most as a deep learning method, followed by CNN, XGBoost, SVM, D.T. and others.

THE ROLE OF MACHINE LEARNING IN THE FIGHT AGAINST COVID-19

The key benefits of applying ML are powerful predictive ability, enhancing medical diagnosis, drug development, modeling the data to make the corresponding decisions, work automation. Figure 6 presents roles of ML-based tools in the fight against the COVID-19.

The results obtained by applying ML contribute to making appropriate decisions in the fight against the COVID-19 pandemic. Also, the application of ML allows a significant improvement in the methods of treatment and treatment of infected persons.

DETECTION AND DIAGNOSIS OF INFECTION

Reverse transcription-polymerase chain reaction (RT-PCR) is an established procedure to identify

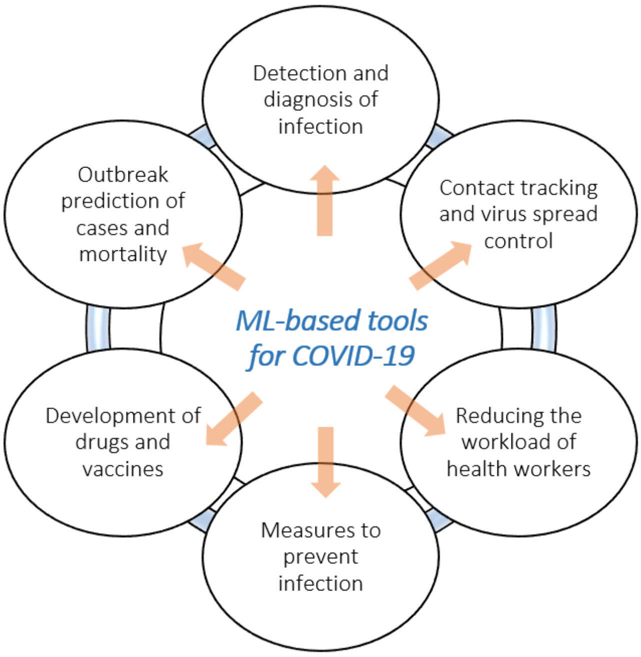

Figure 6. *Key roles of ML techniques and datasets to tackle the COVID-19*

the COVID-19 infection. Also, there are some efforts to improve the efficiency of this technique (36). However, these techniques require specific materials, equipment, and instrument. There are some auxiliary alternative tools to speed up and enhance the process of diagnosis. ML allows data to be quickly analyzed to detect potential symptoms and infection. Timely warnings of patients and health care facilities are often crucial to take appropriate actions. ML algorithms help develop new low-cost systems for detection, diagnosing, and assessing symptoms. Therefore, ML is an indispensable tool for making appropriate decisions faster. This feature helps to provide a better health care system, reduce the negative impacts of the COVID-19 pandemic, mitigate the negative impact on the economy, etc. For this purpose, scientists use different ML-based techniques such as CNN (Convolutional Neural Network) (37), RF (Random Forest) (38), etc.

ML-based COVID-19 diagnosis tools use various data sets such as radiography images (CT, CXR) and respiratory data such as cough and breath. In the past, ML techniques have been used to process and analyze X-rays to help predict infections (39). For example, COVID-Net is a revolutionary system based on a neural network (40). This system enables the differentiation of COVID-19 cases from other diseases by analyzing lung conditions on X-ray images. A system proposed in a study (39) uses X-rays for the COVID-19 case classification. This system achieved 92% and 89.6% accuracy, respectively. The schemes presented in a study (41) applied different CT scan diagnoses and obtained the highest accuracy of 94.98% to differentiate coronavirus symptoms. Also, a modified model for extracting the characteristics of COVID-19 using CT scanning with 89.5% accuracy was presented (41). The 3D system, based on the mechanism of observing CT images, achieved an accuracy of 86.7%. This system differentiates COVID-19 from viral pneumonia (42). Other ML-based systems for COVID-19 diagnosis usually use CT scan readings. For example, neural network-based systems make it possible to distinguish COVID-19 symptoms from other lung diseases (10), (43), (44).

Another example of the ML application is for the analysis of recorded voices to detect the symptoms of the COVID-19 virus. Voice detection applications use volunteer recordings of the user's cough, shortness of breath, and other respiratory samples. ML algorithms assess whether a user has some COVID-19 symptoms or not. Some mobile applications apply ML algorithms for processing these data for early detection of COVID-19 symptoms (47) (48). Table 3 provides an overview of some solutions for the COVID-19 diagnosis.

CONTACT TRACKING AND VIRUS SPREAD CONTROL

SIR (Susceptible-Infected-Removed) is a traditional established model to predict infectious diseases (49). However, this model has some shortcomings including the assumption that recovered patients will not become infected again, ignoring social distancing and quarantine policies and the time-varying nature of some parameters. For example, Dandekat et al. (50) discussed that the traditional SIR model does not include some of the parameters of the COVID-19 pandemic such as distance policies. Therefore, we need to adapt this model to COVID-19 or to develop a new model that is more suitable for this pandemic. There are some efforts to improve traditional SIR models such as Susceptible–Infectious–Quarantined–Recovered (SIQR) (51) or stochastic SIR model (52). Mbunges (53) provides an overview of some opportunities for integrating contact tracing with emerging technologies.

ML technique enables the construction of an intelligent platform for automatic monitoring and prediction of virus spread. For example, we can use the neural network to isolate the visual characteristics of the virus spreading and its impact on humans. Chen YC et al. (54) proposed a time-dependent SIR model to adapt to the COVID-19 pandemic. This model considers two types of patients including detectable and undetectable infected persons. This model enables the prediction of infected cases. The proposed systems in a study (55) applied an ML algorithm to real-time data to take appropriate steps in monitoring the outbreak of COVID-19. The ML application is particularly significant for automated contact retrieval (7). All this helps in the proper monitoring and treatment of patients and to apply the corresponding solutions. We can use ML techniques to monitor and predict the spread of COVID-19 infection over time and to analyze the potential symptoms of COVID-19. The application of ML algorithms can help analyze the level of infection in some geographic areas by tracking the contacts of individuals. In this way, we can identify and monitor the groups of people with highest risk. Furthermore, there are existing solutions to monitor the spread of the pandemic (55).

OUTBREAK PREDICTION OF CASES AND MORTALITY

ML-based tools enable to process a large amount of raw data required for the prediction of the COVID-19 virus spreading, the risks of infection, the projection of deaths in some regions, and some other events. Therefore, ML helps identify the most vulnerable people, countries, and regions. Projections of infection and virus spreading is one of the crucial factors for making appropriate decisions. This is very important to plan effective COVID-19 control strategies (56). There are various approaches and architectures used to outbreak prediction (e.g. project cases of infection and mortality). Many studies aim to establish an ML-based model to predict virus spread and mortality. For example, Kavadi et. al. (57) have proposed a partial derivative regression and non-linear ML model to predict virus transmission. Carrillo-Larco (58) has presented a model based on k-means and various data to show the significance of ML for feature prediction, involved risk, and the causes of an analogous epidemic. Chuanyu Hu et. al. (59) have proposed ML-based approaches (e.g. elastic net, logistic regression, random forest, partial least squares regression, bagged-FDA) for early prediction of mortality risk among COVID-19 patients. Bai et. al. established a model to predict mild patients with potential malignant progression to optimize treatment strategies (60). O.S. Albahri et. al. (61) proposed an ML-based rescue framework to identify severely patients that are suitable for the transfusion of the best convalescent plasma. Therefore, the results of applying ML show that it is an effective tool for virus analysis and prevention in the early stages.

MEASURES FOR PREVENTING COVID-19 INFECTION

ML tools provide updated results based on the real-time processing of large amounts of raw data from various sources. It enables to predict the probable locations of the infection that is useful for quality planning of the engagement of health workers, analysis of the need to open new support centers, planning new beds in hospitals, the introduction of lockdown measures, etc. Based on the discovery of cause-and-effect relationships between individual events, the characteristics and causes of the spread of infection can be discovered. In this way, ML provides the possibility of preventive measures. In the last two years, researchers have proposed various solutions to assess risks and trends at the level of a specific geographical area (62). The obtained results of ML-based data processing can be used to take appropriate steps to control the spread of the pandemic.

DEVELOPMENT OF MEDICINES

ML is a great tool to research medicines and vaccines required for immunization. It enables the deep analysis of available data on COVID-19 as a required action for understanding the virus. Also, this technology enables to speed up the testing of medicines and vaccines where standard testing would take a lot of time (63, 64). Thus, ML algorithms have become a powerful tool for the design of diagnostic tests and clinical trials during the development of vaccines and other medicines. In this field of medicines research, numerous studies provided a significant contribution (65). Numerous studies try to discover effective medicines for COVID-19 using ML-based techniques (66). For example, Rishikesh et. al. (67) use different ML-based algorithms (SVM, RF, MLP, XGBoost) to predict the possible inhibitory synthetic antibodies for SARS-CoV-2. In a study (68) authors used the GISAID database to extract the amino acid residues while seeking potent targets for developing vaccines. Banerjee et.al. (69) proposed a method to investigate the spike proteins of some coronavirus strains required for vaccine development. Besides, ML techniques were used to identify available medicines that may be beneficial against the COVID-19 virus (70).

PRIORITIZATION IN VACCINATION

A relatively new field of ML application in the fight against COVID-19 is vaccination. After the development of some vaccines, numerous studies began to investigate the application of ML to determine the priority of vaccinating patients. Romeo L et al. (71) perform vaccination priority classification using Classification eXtreme Gradient Boosting on a dataset of 17,000 patients of the Italian Federation of General Practitioners. In addition to setting priorities, ML methods have been used to determine the propensity for vaccination campaigns and future strategies around them (72). The vaccination chain involves several complex processes including the planning, procurement, and timely distribution of vaccines to patients. ML techniques can aid these processes.

REDUCING THE WORKLOAD OF HEALTH WORKERS

With the sudden and massive increase of COVID-19 patients, health workers spend much of their time in hospitals. These circumstances cause a heavy workload for health workers. An increased number of infected people led to degradation of the quality of the health care system. The application of ML techniques has contributed to making quality decisions that have resulted in reduced workload. This is achieved by providing more effective models of support for workers and patients (73). Also, ML helps to improve the training of health professionals and medical students who came to the rescue due to the overload of the entire health system (74).

EXAMPLES OF ML-BASED SOLUTIONS FOR COVID-19 MITIGATION

ML techniques and tools have offered an effective, rapid, and inexpensive solution for mitigating the COVID-19 pandemic. Many researchers used and adopted ML approaches to support the fight against this pandemic in various ways such as developing diagnostic tools, mortality assessments models, forecasting models, medicine assessments and discovery models, etc. Scientists have used different types of data for this purpose. For example, some ML approaches use radiography images (e.g. CT and CXR) for diagnostic tools. Other scientists and IT experts have developed inexpensive ML-based tools and applications to screen COVID-19 by using clinical blood samples. These tools are an alternative to expensive radiographic imaging machines (CT, CXR). Besides, some diagnostic tools use coughing data, breathing, other respiratory patterns, thermal videos, etc. Also, some research includes data types such as the total number of positive cases, recoveries, and deaths for COVID-19 outbreak forecasting models and to predict the possible outcome (e.g. patient recovery, infection severity, deaths, etc.). Table 4 include examples of application using other types of data such as statistic reports, weather variables, demographic and mobility data, social platforms, etc while Table 5 summarizes some applications where ML contributes to mitigating the COVID-19.

FUTURE RESEARCH

ML plays a significant role in mitigating the effects of COVID-19 through diagnostic tools, virus spread control, outbreak prediction, vaccine, medicine discovery, etc. However, the ML approaches for the COVID-19 applications face some challenges and impediments. For example, ML requires training data from various sources while there is a shortage of COVID-19 data repositories. Besides, many irregular and downpour data may cause wrong decisions. Collecting COVID-19 data needs to address the issue of privacy protection, security, the methods of collecting data such as the location and movement of people, the way of transferring data between applications, etc. Another challenge is to apply an administrative protocol and policies (local, national, regional, and global) to collect and store the diagnosis data collected from individuals. It is required to develop specific sensing and computing components such as system on chip – SoC for devices used for data collecting (e.g. smartphones) to overcome some of these issues. Therefore, all these issues are challenging tasks to be tackled in future research. We have grouped these challenges as shown in Figure 7.

CUSTOMIZATION OF ML-BASED ALGORITHMS

ML tools are used for various COVID-19 applications (e.g. diagnostic tools, outbreak pandemic spread, the vaccine and drug development, etc.). However, researchers need to examine the most superior algorithm for different applications. For example, the question is which algorithm is the best to find the potential antibody candidates for the fight against the COVID-19 virus. Also, there is a requirement to reduce the computational complexity of the training data. A possible solution for this issue is to use local devices (e.g. smartphones, local servers, etc.) to perform pre-processing COVID-19 data. Each device can train its model by using local data. Then, the device can send results to the central server or cloud for the aggregating process to make a global model. Further, this model of computational offloading can be distributed to all local devices for further processing. However, this initial idea is a very challenging task for future research.

COLLECTING RELEVANT DATA FROM VARIOUS SOURCES

One of the main challenges to ML application is the lack of widely available relevant data sets and computational resources. The corresponding data sets are necessary to discover trends and patterns of infections by specific criteria such as age group, gender, medical history, place of residence. Besides, there is a need for global coordination that implies research collaboration and data sharing among various entities such as countries, civil society, medical institutions, the research community, public and private sectors, etc. This approach enables some high-level decisions to mitigate the COVID-19 pandemic impact. However, this solution has an implementation problem which is particularly pronounced in developing countries where adequate ICT infrastructure is not available. Thus, the development of an infrastructure that enables sharing data and provides big and high-quality datasets is still an open issue.

LINKING ML AND MEDICAL KNOWELEDGE

All ML-based solutions used in the fight against COVID-19 require the corresponding data sets and communication technologies for data transferring. Using ML-based tools for real-time spatial analysis of collected data is a very interesting research topic. This issue requires the application of GIS, data collecting technologies, communication technologies, etc. Another example is using ML and big data repositories to examine the effects of meteorological conditions, living environment, other diseases, air pollution, etc. on the COVID-19 transmission. It is required to use the corresponding ML technique and various ICT and medical knowledge to discover some patterns. This multidisciplinary approach and technologies enable optimization of processes such as treatment of patients, hospital procedures, etc. The increased number of patients in hospitals overwhelmed the health systems across the globe. ML enables to find solutions for reducing the pressure and workload of medical staff. However, this is still a challenging task that needs to be considered by researchers. This issue implies a requirement for a multidisciplinary approach that includes the application of ML, ICT, medical knowledge, and intensive research collaboration and data sharing.

ACCURACY AND RELIABILTY

Many ML-based approaches have been proposed and tested by using different datasets. This issue is a problem because it is impossible to compare the accuracy of the results. The results are usually questionable and need to be compared with those of the direct test. The question of data reliability arises, on which the accuracy of the results depends the most. Therefore, it is challenging issue to compare ML-based approaches for various applications since various sizes and reliability of COVID-19 data sets. Thus, it is required to unify datasets to evaluate proposed approaches and models.

SECURITY AND PRIVACY

ML-based approaches require data from various sources such as personal data such as CT scans, diagnosis reports, GPS location, etc. The question is how these data are secure and private. There is a challenge to encourage people to provide their personal data regardless its purpose. Therefore, authorities should guarantee data privacy and security. One of the technologies with a promising future to solve this problem is blockchain. This technology provides high-reliable, secure, and accessible datasets. However, its application for this purpose is in an early stage of research.

ADOPTION AND HARMONIZATION OF POLICIES

In the last two years' various policies have been taken to control the COVID-19 pandemic. For example, the authorities have taken many measures such as lockdown, social distancing, travel control, massive testing, vaccination. Therefore, government authorities around the globe have a crucial role in defining these policies and harmonizing the approaches as well. This issue is a very challenging task because of the various socio-economic, political development, and other states of the countries.

Authorities need to encourage the involvement of all stakeholders such as patients and other residents, medical staff, scientists, industry, etc. to respect the rules and to engage the collaboration in activities that can improve the fight against the COVID-19 pandemic. WHO has a key role in overcoming this challenge. It encourages governments, health organizations, and industry for collaborative work to adapt and harmonize policies around the globe and to coordinate actions globally which is still an open issue. These are necessary activities to enable quality data sets and to exchange achievements such as the use of ML for mitigating the COVID-19 pandemic.

CONCLUSION

Since the outbreak of the COVID-19 pandemic, researchers around the world have been working intensively to find appropriate methods and tools to address the greatest challenge today. ML has proven to be an effective tool used for various purposes, such as predicting the spread of a pandemic, diagnosing and detecting symptoms, developing medicines and vaccines. ML-based approaches improve techniques for early detection of virus infection, which is necessary to take appropriate measures and adequate treatment. In addition, ML applications have enabled the processing of big data sets, which has helped to make the right decisions. Therefore, ML-based tools enable aggregating raw data from many sources required for predicting transmission. Also, ML plays a particularly significant role in medicine and vaccine development and studying the effectiveness of the existing medicine. It enables the development of tools for facilitating vaccine manufacturing. However, there are many open issues and various challenges. In this paper, we have presented the examples of ML-based tools.We also presented several ML-based approaches that have been proposed for mitigating the COVID-19. This paper is a structured overview of the current state-of-the-art that helps to identify some potentials of ML techniques in the fight against this pandemic. Besides, we highlighted some issues that require the attention of all stakeholders in the future.

Although all ML categories have the potential to be used in the fight against the COVID-19 pandemic, supervised learning categories are the most prevalent and useful. The reason is the fact that the most common applications of ML in COVID-19 require labelled data. In addition, supervised learning shows higher accuracy, which positively affects confidence in the obtained results. Since the emergence of new COVID-19 virus strains, it is necessary to continuously improve and develop new tools, vaccines, and other medicines. Time has shown that the COVID-19 virus mutates and new strains emerge. ML is emerging as an indispensable tool for the fight against the pandemic. However, some open issues require future research to tackle the COVID-19 pandemic. For example, some challenges include the development of new ML-based diagnostic tools, contact tracking systems, etc. Therefore, we need to adapt an existing ML-based tools that have proven good but insufficient in the fight against the COVID-19.

Acknowledgements

This work was supported by the Federal Ministry of Education and Science, Federation of Bosnia and Herzegovina.

References

1. 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. https://doi.org/10.1016/j.chaos.2020.110059

2. Nwanosike EM, Conway BR, Merchant HA, Hasan SS. Potential applications and performance of machine learning techniques and algorithms in clinical practice: A systematic review. Int J Med Inform 2022; 159: 104679. https://doi.org/10.1016/j.ijmedinf.2021.104679

3. Kucharski AJ, Russell TW, Diamond C et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020; 20(5): 553-558. https://doi.org/10.1016/S1473-3099(20)30144-4

4. Campbell TW, Wilson MP, Roder H et al. Predicting prognosis in COVID-19 patients using machine learning and readily available clinical data. Int J Med Inform 2021; 155: 104594. https://doi.org/10.1016/j.ijmedinf.2021.104594

5. Ai T, Yang Z, Hou H et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China. Radiology 2020; 296(2): E32-E40. https://doi.org/10.1148/radiol.2020200642

6. Rasheed J, Jamil A, Hameed AA et al. A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. Chaos Solitons Fractals 2020; 141: 110337. https://doi.org/10.1016/j.chaos.2020.110337

7. Xu B, Kraemer MUG. Open access epidemiological data from the COVID-19 outbreak. Lancet Infect Dis 2020; 20(5): 534. https://doi.org/10.1016/S1473-3099(20)30119-5

8. Verma A, Amin SB, Naeem M, Saha M. Detecting COVID-19 from chest computed tomography scans using AI-driven android application. Comput Biol Med 2022; 143: 105298. https://doi.org/10.1016/j.compbiomed.2022.105298

9. Hassan H, Ren Z, Zhao H et al. Review and classification of AI-enabled COVID-19 CT imaging models based on computer vision tasks. Comput Biol Med 2022; 141: 105123. https://doi.org/10.1016/j.compbiomed.2021.105123

10. Islam MZ, Islam MM, Asraf A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Inform Med Unlocked 2020; 20: 100412. https://doi.org/10.1016/j.imu.2020.100412

11. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet 2015; 16: 321–332. https://doi.org/10.1038/nrg3920

12. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health 2018; 8(2): 020303. https://doi.org/10.7189/jogh.08.020303

13. Yin AL, Guo WL, Sholle ET et al. Comparing automated vs. manual data collection for COVID-specific medications from electronic health records. Int J Med Inform 2022; 157: 104622. https://doi.org/10.1016/j.ijmedinf.2021.104622

14. Hou Z, Du F, Jiang H et al. Assessment of Public Attention, Risk Perception, Emotional and Behavioural Responses to the COVID-19 Outbreak: Social Media Surveillance in China. medRxiv 2020; 2020.03.14.20035956. https://doi.org/10.1101/2020.03.14.20035956

15. Chen B, Shi M, Ni X al. Visual Data Analysis and Simulation Prediction for COVID-19. International Journal of Educational Excellence 2020; 6(1) 95-114.

https://doi.org/10.18562/IJEE.055

16. Ndiaye M, Oyewobi SS, Abu-Mahfouz AM et al. IoT in the Wake of COVID-19: A Survey on Contributions, Challenges and Evolution. IEEE Access 2020; 8: 186821-186839. https://doi.org/10.1109/ACCESS.2020.3030090

17. Geng Z, Zhang X, Fan Z et al. Recent Progress in Optical Biosensors Based on Smartphone Platforms. Sensors 2017; 17(11): 1-19. https://doi.org/10.3390/s17112449

18. Chen CM, Jyan HW, Chien SC et al. Containing COVID-19 Among 627,386 Persons in Contact With the Diamond Princess Cruise Ship Passengers Who Disembarked in Taiwan: Big Data Analytics. J Med Internet Res 2020;22(5):e19540. https://doi.org/10.2196/19540

19. Shuja J, Alanazi E, Alasmary W, Alashaikh A. COVID-19 open source data sets: a comprehensive survey. Appl Intell 2021; 51: 1296–1325. https://doi.org/10.1007/s10489-020-01862-6

20. Malki Z, Atlam ES, Hassanien AE et al. Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. Chaos Solitons Fractals 2020; 138: 110137. https://doi.org/10.1016/j.chaos.2020.110137

21. Iwendi C, Bashir AK, Peshkar A et al. COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm. Front Public Health 2020; 8: 357. https://doi.org/10.3389/fpubh.2020.00357

22. Yadav M, Perumal M, Srinivas M. Analysis on novel coronavirus (COVID-19) using machine learning methods. Chaos Solitons Fractals 2020; 139: 110050. https://doi.org/10.1016/j.chaos.2020.110050

23. Zhao X, Liu X, Li X. Tracking the spread of novel coronavirus (2019-nCoV) based on big data. medRxiv 2020; 2020.02.07.20021196. https://doi.org/10.1101/2020.02.07.20021196

24. Cohen JP, Morrison P, Lan D. Covid-19 Image Data Collection. arXiv 2020; 2003.11597. https://doi.org/10.48550/arXiv.2003.11597

25. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020; 20(5): 533-534. https://doi.org/10.1016/S1473-3099(20)30120-1

26. Pandey G, Chaudhary P, Gupta R, Pal S. SEIR and Regression Model based COVID-19 outbreak predictions in India. arXiv 2020; 2004.00958. https://doi.org/10.48550/arXiv.2004.00958

27. Mathieu E, Ritchie H, Ortiz-Ospina E et al. A global database of COVID-19 vaccinations. Nat Hum Behav 2021; 5(7): 947-953. https://doi.org/10.1038/s41562-021-01122-8

28. European Centre for Disease Prevention and Control. Download COVID-19 datasets. [cited 2021 November 3]. https://www.ecdc.europa.eu/en/covid-19/data

29. Fetić A, Dželihodžić A. A practical implementation of machine learning in predicting breast cancer. Sci Eng Technol 2021; 1(2): 16–23. https://doi.org/10.54327/set2021/v1.i2.16

30. Feng S, Xu L. An intelligent decision support system for fuzzy comprehensive evaluation of urban development. Expert Syst Appl 1999; 16(1): 21-32. https://doi.org/10.1016/S0957-4174(98)00028-1

31. Likas A, Vlassis N, Verbeek J. The global k-means clustering algorithm. Pattern Recognit 2003; 36(2): 451-461. https://doi.org/10.1016/S0031-3203(02)00060-2

32. Qiu J, Wu Q, Ding G et al. Erratum to: A survey of machine learning for big data processing. Eurasip J Adv Signal Process 2016; 2016:85. https://doi.org/10.1186/s13634-016-0382-7

33. Berral-García JL. A quick view on current techniques and machine learning algorithms for big data analytics. 18th International Conference on Transparent Optical Networks (ICTON) 2016; pp. 1-4. https://doi.org/10.1109/ICTON.2016.7550517

34. Wu PY, Cheng CW, D Kaddi C et al. Omic and Electronic Health Record Big Data Analytics for Precision Medicine. IEEE. Trans. Biomed. Eng. 2017; 64(2): 263-273. https://doi.org/10.1109/TBME.2016.2573285

35. Wang L, Alexander CA. Machine Learning in Big Data. Int J Math Eng Manag Sci 2016; 1(2): 52–61. https://doi.org/10.33889/IJMEMS.2016.1.2-006

36. S. Fomsgaard A, Rosenstierne MW. An alternative workflow for molecular detection of SARS-CoV-2 – escape from the NA extraction kit-shortage, Copenhagen, Denmark, March 2020. Euro Surveill 2020; 25(14) :pii=2000398. https://doi.org/10.2807/1560-7917.ES.2020.25.14.2000398

37. J L G, Abraham B, M S S, Nair MS. A computer-aided diagnosis system for the classification of COVID-19 and non-COVID-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. Comput Biol Med 2022; 141: 105134. https://doi.org/10.1016/j.compbiomed.2021.105134

38. Tang Z, Zhao W, Xie X et al. Severity Assessment of Coronavirus Disease 2019 (COVID-19) Using Quantitative Features from Chest CT Images. arXiv 2020; 2003.11988. https://doi.org/10.48550/arXiv.2003.11988

39. Iqbal Khan A, Latief Shah J, Mudasir Bhat M. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Programs Biomed 2020; 196: 105581. https://doi.org/10.1016/j.cmpb.2020.105581

40. Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep 2020; 10(1): 19549. https://doi.org/10.1038/s41598-020-76550-z

41. Wang S, Kang B, Ma J et al. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). Eur Radiol 2021; 31: 6096–6104. https://doi.org/10.1007/s00330-021-07715-1

42. Xu X, Jiang X, Ma C et al. A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia. Engineering (Beijing) 2020; 6(10): 1122-1129. https://doi.org/10.1016/j.eng.2020.04.010

43. Jin C, Chen W, Cao Y et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nat Commun 2020; 11(1): 5088. https://doi.org/10.1038/s41467-020-18685-1

44. Panahi A, Askari Moghadam R, Akrami M, Madani K. Deep Residual Neural Network for COVID-19 Detection from Chest X-ray Images. SN Comput Sci. 2022; 3(2):169. https://doi.org/10.1007%2Fs42979-022-01067-3

45. Lahsaini I, El Habib Daho M, Chikh MA. Deep transfer learning based classification model for covid-19 using chest CT-scans. Pattern Recognit Lett 2021; 152: 122-128. https://doi.org/10.1016/j.patrec.2021.08.035

46. Li WT, Ma J, Shende N et al. Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2020; 20(1): 247. https://doi.org/10.1186/s12911-020-01266-z

47. Imran A, Posokhova I, N Qureshi H et al. Ai4covid-19: AI enabled preliminary diagnosis for covid-19 from cough samples via an app. Inform Med Unlocked 2020; 20: 100378. https://doi.org/10.1016/j.imu.2020.100378

48. Chamola V, Hassija V, Gupta V, Guizani M. A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing Its Impact. IEEE Access 2020; 8: 90225-90265. https://doi.org/10.1109/ACCESS.2020.2992341

49. Zhong X, Deng F, Ouyang H. Sharp Threshold for the Dynamics of a SIRS Epidemic Model With General Awareness- Induced Incidence and Four Independent Brownian Motions. IEEE Access 2020; 8: 29648-29657. https://doi.org/10.1109/ACCESS.2020.2972577

50. Dandekar R, Barbastathis G. Neural Network

aided quarantine control model estimation of COVID spread in Wuhan, China. arXiv 2020; 2003.09403.
https://doi.org/10.48550/arXiv.2003.09403

51. Crokidakis N. Modeling the early evolution of the COVID-19 in Brazil: Results from a Susceptible–Infectious–Quarantined–Recovered (SIQR) model. Int J Mod Phys C 2020; 31(10): 2050135.
https://doi.org/10.1142/S0129183120501351

52. Gaeta G. A simple SIR model with a large set of asymptomatic infectives. Math Eng 2020; 3(2): 1-39.
https://doi.org/10.3934/mine.2021013

53. Mbunge E. Integrating emerging technologies into COVID-19 contact tracing: Opportunities, challenges and pitfalls. Diabetes Metab Syndr 2020; 14(6): 1631-1636.
https://doi.org/10.1016/j.dsx.2020.08.029

54. Chen YC, Lu PE, Chang CS, Liu TH. A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons. arXiv 2020; 2003.00122.
https://doi.org/10.1109/TNSE.2020.3024723

55. Maghdid HS, Ghafoor KZ, Sadiq AS et al. A Novel AI-enabled Framework to Diagnose Coronavirus COVID 19 using Smartphone Embedded Sensors: Design Study. arXiv 2020; 2003.07434.
https://doi.org/10.1109/IRI49571.2020.00033

56. Giordano G, Blanchini F, Bruno R et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020; 26: 855–860.
https://doi.org/10.1038/s41591-020-0883-7

57. Kavadi PD, Patan R, Ramachandran M, Gandomi AH. Partial derivative Nonlinear Global Pandemic Machine Learning prediction of COVID 19. Chaos Solitons Fractals 2020; 139: 110056.
https://doi.org/10.1016/j.chaos.2020.110056

58. Carrillo-Larco RM, Castillo-Cara M. Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach. Wellcome Open Res 2020;5:56.
https://doi.org/10.12688/wellcomeopenres.15819.2

59. Hu C, Liu Z, Jiang Y et al. Early prediction of mortality risk among patients with severe COVID-19, using machine learning. Int J Epidemiol 2020; 49(6): 1918–1929.
https://doi.org/10.1093/ije/dyaa171

60. Bai X, Fang C, Zhou Y et al. Predicting COVID-19 Malignant Progression with AI Techniques. medRxiv 2020; 2020.03.20.20037325.
https://doi.org/10.1101/2020.03.20.20037325

61. Albahri OS, R.Al-Obaidi J, Zaidan AA et al. Helping doctors hasten COVID-19 treatment: Towards a rescue framework for the transfusion of best convalescent plasma to the most critical patients based on biological requirements via ml and novel MCDM methods. Comput Methods Programs Biomed 2020; 196: 105617.
https://doi.org/10.1016/j.cmpb.2020.105617

62. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals 2020; 135: 109864.
https://doi.org/10.1016/j.chaos.2020.109864

63. Haleem A, Vaishya R, Javaid M, Khan IH. Artificial Intelligence (AI) applications in orthopaedics: An innovative technology to embrace. J Clin Orthop Trauma 2020; 11(Suppl 1): S80-S81.
https://doi.org/10.1016/j.jcot.2019.06.012

64. Chen S, Yang J, Yang W et al. COVID-19 control in China during mass population movements at New Year. Lancet 2020; 395(10226): 764-766.
https://doi.org/10.1016/S0140-6736(20)30421-9

65. Ward D, Higgins M, Phelan JE et al. An integrated in silico immuno-genetic analytical platform provides insights into COVID-19 serological and vaccine targets. Genome Med 2021; 13(1): 4.
https://doi.org/10.1186/s13073-020-00822-6

66. Zhavoronkov A, Aladinskiy V, Zhebrak A et al. Potential 2019-nCoV 3C-like protease inhibitors

designed using generative deep learning approaches. ChemRxiv Cambridge: Cambridge Open Engage 2020; 11829102.v1. https://doi.org/10.26434/chemrxiv.11829102.v1

67. Magar R, Yadav P, Barati Farimani A. Potential neutralizing antibodies discovered for novel corona virus using machine learning. Sci Rep 2021; 11(1): 5261. https://doi.org/10.1038/s41598-021-84637-4

68. Ahmed SF, Quadeer AA, McKay MR. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. Viruses. 2020; 12(3): 254. https://doi.org/10.3390/v12030254

69. Banerjee A, Santra D, Maiti S. Energetics and IC50 based epitope screening in SARS CoV-2 (COVID 19) spike protein by immunoinformatic analysis implicating for a suitable vaccine development. J Transl Med 2020; 18: 281. https://doi.org/10.1186/s12967-020-02435-4

70. Beck BR, Shin B, Choi Y et al. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol 2020; 18: 784-790. https://doi.org/10.1016/j.csbj.2020.03.025

71. Romeo L, Frontoni E. A Unified Hierarchical XGBoost model for classifying priorities for COVID-19 vaccination campaign. Pattern Recognit 2022; 121: 108197. https://doi.org/10.1016/j.patcog.2021.108197

72. Su Z, McDonnell D, Wen J et al. Young adults' preferences for influenza vaccination campaign messages: Implications for COVID-19 vaccine intervention design and development. Brain Behav. Immun. 2021; 14: 100261. https://doi.org/10.1016/j.bbih.2021.100261

73. Gozes O, Frid-Adar M, Greenspan H et al. Rapid AI Development Cycle for the Coronavirus (COVID-19) Pandemic: Initial Results for Automated Detection & Patient Monitoring using Deep Learning CT Image Analysis. arXiv 2020; 2003.05037. https://doi.org/10.48550/arXiv.2003.05037

74. Haleem A, Javaid M, Vaishya R. Effects of COVID 19 pandemic in daily life. Curr Med Res Pract 2020; 10(2): 78-79. https://doi.org/10.1016/j.cmrp.2020.03.011

75. Rustam F, Ahmad Reshi A, Mehmood A et al. COVID-19 Future Forecasting Using Supervised Machine Learning Models. IEEE Access 2020; 8: 101489 - 101499. https://doi.org/10.1109/ACCESS.2020.2997311

76. Ye Y, Hou S, Fan Y et al. α-Satellite: An AI-driven System and Benchmark Datasets for Hierarchical Community-level Risk Assessment to Help Combat COVID-19. arXiv 2020; 2003.12232. https://doi.org/10.48550/arXiv.2003.12232

77. Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos Solitons Fractals 2020;140:110212. https://doi.org/10.1016/j.chaos.2020.110212

78. Kamal Paul S, Jana S, Bhaumik P. A Multivariate Spatiotemporal Model of COVID-19 Epidemic Using Ensemble of ConvLSTM Networks. J Inst Eng India Ser B 2021; 102: 1137–1142. https://doi.org/10.1007/s40031-020-00517-x

79. Tuli S, Tuli S, Tuli R, Singh Gill S. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet of Things 2020; 11: 100222. https://doi.org/10.1016/j.iot.2020.100222

80. Kang H, Xia L, Yan F et al. Diagnosis of Coronavirus Disease 2019 (COVID-19) With Structured Latent Multi-View Representation Learning. IEEE Trans Med Imaging 2020; 39(8): 2606-2614. https://doi.org/10.1109/TMI.2020.2992546

81. Zhou L, Li Z, Zhou J et al. A Rapid, Accurate and Machine-Agnostic Segmentation and Quantification Method for CT-Based COVID-19 Diagnosis. IEEE Trans Med Imaging 2020; 39(8): 2638-2652. https://doi.org/10.1109/TMI.2020.3001810

82. Abbasian Ardakani A, Acharya UR, Habibollahi S, Mohammadi A. COVIDiag: a clinical CAD system to diagnose COVID-19 pneumonia based on CT findings. Eur Radiol 2021; 31: 121–130. https://doi.org/10.1007/s00330-020-07087-y

83. Sethy PK, Behera SK, Ratha PK, Biswas P. Detection of Coronavirus Disease (COVID-19) based on Deep Features and Support Vector Machine. Int J Math Eng Manag Sci 2020; 5(4): 643-651. https://doi.org/10.33889/IJMEMS.2020.5.4.052

84. Yoo SH, Geng H, Chiu TL et al. Deep Learning-Based Decision-Tree Classifier for COVID-19 Diagnosis From Chest X-ray Imaging. Front Med (Lausanne) 2020; 7: 427. https://doi.org/10.3389/fmed.2020.00427

85. Toğaçar M, Ergen B, Cömert Z. COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. Comput Biol Med 2020;121:103805. https://doi.org/10.1016/j.compbiomed.2020.103805

86. Tuncer T, Dogan S, Ozyurt F. An automated Residual Exemplar Local Binary Pattern and iterative ReliefF based COVID-19 detection method using chest X-ray image. Chemometr Intell Lab Syst 2020; 203: 104054. https://doi.org/10.1016/j.chemolab.2020.104054

87. Yang HS, Hou Y, Vasovic LV et al. Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection Using Machine Learning. Clin Chem 2020; 66(11): 1396-1404. https://doi.org/10.1093/clinchem/hvaa200

88. Batista AFM, Miraglia JL, Donato THR, Chiavegatto Filho ADP. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. medRxiv 2020; 2020.04.04.20052092. https://doi.org/10.1101/2020.04.04.20052092

89. Wu J, Zhang P, Zhang L et al. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. medRxiv 2020; 2020.04.02.20051136. https://doi.org/10.1101/2020.04.02.20051136

90. Brinati D, Campagner A, Ferrari D et al. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. J Med Syst 2020; 44(8): 135. https://doi.org/10.1007/s10916-020-01597-4

91. Khanday AMUD, Rabani ST, Khan QR et al. Machine learning based approaches for detecting COVID-19 using clinical text data. Int J Inf Technol 2020; 12(3): 731-739. https://doi.org/10.1007/s41870-020-00495-9

92. Wang Y, Hu M, Zhou Y et al. Unobtrusive and Automatic Classification of Multiple People's Abnormal Respiratory Patterns in Real Time Using Deep Neural Network and Depth Camera. IEEE Internet of Things J 2020; 7(9): 8559-8571. https://doi.org/10.1109/JIOT.2020.2991456

93. Jiang Z, Hu M, Gao Z et al. Detection of Respiratory Infections Using RGB-Infrared Sensors on Portable Device. IEEE Sensors J 2020; 20(22): 13674-13681. https://doi.org/10.1109/JSEN.2020.3004568

94. Wang Y, Hu M, Li Q et al. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. arXiv 2020; 2002.05534. https://doi.org/10.48550/arXiv.2002.05534

95. Brown C, Chauhan J, Han J al. e. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. 26th ACM SIGKDD Int Conf Knowl Discov Data Min; 2020. p. 3474-3484. https://doi.org/10.1145/3394486.3412865

96. Despotovic V, Ismael M, Cornil M et al. Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results. Comput Biol Med 2021; 138: 104944. https://doi.org/10.1016/j.compbiomed.2021.104944

97. Pahar M, Klopper M, Warren R, Niesler T. COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features. Comput Biol Med 2022; 141: 105153. https://doi.org/10.1016/j.compbiomed.2021.105153

Article info

Received: June 9, 2022
Revised: July 28, 2022
Accepted: September 5, 2022
Online first: January 1, 2023

Primena mašinskog učenja u borbi protiv pandemije COVID-19 virusa

Alem Čolaković, Elma Avdagić-Golub, Muhamed Begović, Belma Memić, Adisa Hasković Džubur

Univerzitet u Sarajevu, Fakultet za saobraćaj i komunikacije, Sarajevo, Bosna i Hercegovina

SAŽETAK

Uvod. Mašinsko učenje (engl. Machine Learning – ML) ima značajnu ulogu u borbi protiv COVID-19 pandemije (SARS-CoV-2). ML bazirane tehnike omogućavaju brzo otkrivanje uzročno-posledičnih veza i trendova iz velikog uzorka podataka. Zbog toga ove tehnike pružaju efikasne metode za generisanje informacija i sticanje znanja iz struktuiranih i nestruktuiranih podataka. Ovo je posebno značajno u uslovima koji utiču na sve apsekte ljudskih života kao što je slučaj sa pandemijama. U ovakvim slučajevima je neophodno prikupljati veliku količinu podataka koja će omogućiti donošenje odgovarajućih mera za sprečavanje širenja pandemije, ranu dijagnostiku infekcije, pronalazak lekova, smanjenje negativnih posledica, itd. Savremene informacijske i komunikacijske tehnologije (IKT) omogućavaju i nove koncepte primene kao što je Internet stvari (engl. Internet of Things –IoT), a koji omogućava automatizirano i efikasno prikupljanje podataka iz različitih izvora. Ovo otvara mogućnosti za kreiranje efikasnih ML-baziranih mehanizama i prediktivnih modela potrebnih za donošenje odgovarajućih mera i odluka u specifičnim situacijama.

Metode. Ovaj pregledni rad je baziran na prethodnim studijama i znanstveno dokazanim saznanjima. U radu su korišteni bibliometrijski podaci iz referentnih baza podataka istraživačkih publikacija (Web of Science, Scopus, PubMed) kombinirani za bibliometrijske analize u kontekstu primjene mašinskog učenja za borbu protiv COVID-19.

Cilj. Rad prezentuje najkorištenije primene mašinskog učenja za smanjenje uticaja COVID-19 pandemije. Cilj je da se prikažu potencijali, rešenja i mogućnosti primene različitih tehnika, algoritama i skupova podataka (engl. datasets) u kontekstu borbe protiv navedene pandemije. Također, u radu su predstavljene određene ideje i otvorena pitanja za buduća istraživanja što može koristiti kao polazna tačka za buduća istraživanja.

Zaključak. ML je učinkovit alat za dijagnosticiranje i rano otkrivanje simptoma, predviđanje širenja pandemije, razvoj lijekova i vakcina, itd.

Ključne reči: **mašinsko učenje, COVID-19 pandemija, COVID-19 datasets, vještačka inteligencija**