Retraction Notice

The Editor-in-Chief and the publisher have retracted this article, which was submitted as part of a guest-edited special section. An investigation uncovered evidence of compromised peer review, which may have led to acceptance of this out-of-scope article. The Editor and publisher no longer have confidence in the results and conclusions of the article.

KY and MJE do not agree with the retraction.

REVIEW

Review of artificial intelligence-assisted COVID-19 detection solutions using radiological images

Kolsoum Yousefpanah[a] **and Mohammad Javad Ebadi**[b,c,*]
[a]University of Guilan, Department of Statistics, Rasht, Iran
[b]Chabahar Maritime University, Department of Mathematics, Chabahar, Iran
[c]International Telematic University Uninettuno, Section of Mathematics, Roma, Italy

Abstract. A new virus called coronavirus has been affecting the world since 2019 and has killed millions of people. Even though vaccines for the virus have been developed and the mortality rate is decreasing across the world, many countries are still struggling with the pandemic. Artificial intelligence (AI) methods have been regarded as fast techniques and powerful tools for screening this disease. We reviewed papers that used AI-based systems for the diagnosis of COVID-19 using radiological images, such as X-rays and computed tomography (CT) images. This survey focuses specifically on deep learning (DL)-based systems for screening COVID-19 patients. Privacy and accuracy of diagnosis are of paramount importance in a clinical environment. In most surveys, the privacy issue is not taken into consideration. In this regard, we categorize recent work into three taxonomies: federated learning (FL) models (privacy-guarding methods), ensemble machine learning (ML) models, and other ML and DL models. A summary of the selected articles is presented; parameters such as the modality, experimental tools, data sources, number of classes, and positive and nonpositive aspects of each model and work, as well as evaluation measures, are depicted. In fact, we compare papers and their experimental results to find more accurate and privacy-guarding methods. Also, the type of data and tools that are helpful for more accurate prediction were investigated. Finally, we refer to some limitations of ML methods and provide useful insight for future researchers. In this survey, 45% and 41% of papers used X-ray and CT images for experiments, respectively. Using multiple datasets was the preference of 61% of researchers, and 45% of papers considered binary classification. The average accuracy of 95.71%, 97.09%, and 93.38% was obtained for federated ML, ensemble ML, and other ML models, respectively. To sum up, X-ray images were the favorite of most articles. Also, most researchers employed multiple databases for their experiments, and binary classification was the method of choice for most of them. Among the three categories, ensemble learning-based systems demonstrated the best performance in terms of all evaluation metrics. Therefore, these systems can be used to screen COVID-19 patients. [DOI: 10.1117/1.JEI.32.2.021405]

Keywords: artificial intelligence; connected healthcare; deep learning; federated learning; IoT; machine learning.

Paper 220463SS received May 4, 2022; accepted for publication Sep. 1, 2022; published online Sep. 22, 2022.

1 Introduction

Coronavirus spread all over the world from Wuhan, China, in late December of 2019, leading to fatal respiratory infections.[1] Initially, the rate of spread was high in various countries such as Iran, Italy, India, and the United States. Several symptoms of this virus include respiratory problems, general weakness, cough, and fever. To cure and control the COVID-19 disease, early diagnosis and detection are crucial. Because the virus spreads rapidly, automatic COVID-19 detection studies using artificial intelligence (AI) methods are more effective compared with manual diagnosis. Although as a standard test, reverse transcription-polymerase chain reaction (RT-PCR) is employed to diagnose the disease, this test fails in many cases and in the early

*Address all correspondence to Mohammad Javad Ebadi, ebadi@cmu.ac.ir

1017-9909/2022/$28.00

phases of the disease.[2,3] To detect the disease by RT-PCR, normally 2 days is required, and sometimes it suffers from some inherited limitation. Therefore, X-rays and computed tomography devices could serve as a reliable, valuable, and rapid technique in the detection and evaluation of COVID-19.[4]

AI plays an important role in management science and operational research. Intelligence is often defined as the ability to gather information to tackle complex issues. In the not-too-distant future, intelligent machines will take over many human functions.[5] The study and design of intelligent machines and software that have the capability of collecting knowledge, thinking, learning, interacting, operating, and seeing objects is known as AI. AI places high attention on computing, so it differs from psychology.[5] Also, AI differs from computer science due to its emphasis on perception, reasoning, and action. It improves the intelligence and utility of machines. Artificial neurons and scientific theorems are used to make it work.[5,6]

AI systems can perform a given task such as visual perception, speech recognition, scene interpretation, object detection, decision making, and translation, similar to a specialist. One thing about AI that has never changed is that, as these tasks are executed, problem-specific features are extracted and processed. Initially, simpler algorithms were used to perform an automated decision-making process via features including edge information, frequency changes, and plane differences.[3] Then, the computing capability of computers grew, and more advanced algorithms appeared to extract and classify features. AI technologies have drawn the attention of the medical world and have entered a new age with the advent of deep learning (DL).[3] Much work has been conducted in the medical domain with AI-based systems.[7–9] For example, many studies have focused on feature selection and feature extraction methods. More information about how appropriate feature selection and extraction techniques affect the efficiency of models can be found in Refs. 10 and 11.

AI methods play an important role in the detection of COVID-19 disease. In tracing the speed rate of the virus, and identifying its growth rate, AI technology is highly helpful and assists us in recognizing the risk and severity of COVID-19 disease. Moreover, AI helps to predict the probability of death by analyzing prior patient data. Generally, AI is a robust tool in the world for fighting the virus via testing individuals, medical assistance, data, and information.[6] Different machine learning (ML) and DL systems, specifically convolutional neural networks (CNN), have been suggested for the classification of the X-ray and CT samples, the outbreak forecast, and prediction. Computer vision[12] has also had a positive contribution to the decline in the severity of this outbreak. Further, Internet of things (IoT),[13,14] big data,[15,16] and smartphone technology[17,18] were widely effective for combating the spread of COVID-19.[19]

Some review papers have been conducted on using AI-based systems for the detection and analysis of COVID-19 disease using X-ray and chest CT images. Shoeibi et al.[20] provided a review of DL methods-based papers for automated segmentation of lung images and detection of COVID-19 with a focus on papers that employed CT and X-ray images. Moreover, papers about the prediction of COVID-19 predominance in various areas of the world by DL methods were reviewed. They focused on several matters such as classification, segmentation, and prediction of coronavirus disease via DL models. Some challenges that researchers have confronted and future works were explained; however, taxonomy was not considered for the reviewed papers.

Islam et al.[19] investigated the recently developed methods for detection of COVID-19 via DL models from chest CT and X-ray images. For taxonomy in the paper, the reviewed methods were assorted based on customized DL techniques and pretrained systems through deep transfer learning. The most essential systems that are applied for the detection of coronavirus disease were investigated with emphasis on the applied data in experiments, the data splitting method, and the evaluation parameters. The challenges of existing DL-based systems and future works were also presented. Nevertheless, the paper selection process was not clear.

Alafif et al.[6] surveyed ML and DL methods for the diagnosis and treatment of COVID-19. In this paper, the ML- and DL-based techniques, tools, datasets, and performance were summarized. The authors discussed details of ML- and DL-based approaches for classifying chest CT and X-ray images. The severity of COVID-19 was also shown. Moreover, the challenges and potential guidance were provided. Nevertheless, the paper selection strategies were not taken into account.

Ozsahin et al.[21] provided a review on AI methods for diagnosis of COVID-19 via chest CT samples. The assortment of papers was based on the classification cases: COVID-19/non-COVID-19, COVID-19/non-COVID-19 pneumonia, COVID-19/normal, and intensity. The performance evaluations were presented for each method; however, the future trends were neglected.

Ilyas et al.[22] reviewed papers that investigate the various AI-based strategies to detect COVID-19 by chest X-ray images. Various DL models such as DenseNet-20, AlexNet, InceptionV3, GoogleNet, ResNet-101, ResNet-50, ResNet-18, VGG-16, XceptionNet, and InceptionNetV2 were discussed. These methods identify infected people and show whether the pneumonia is caused by COVID-19 or another fungal attack. The accuracy of methods is provided; nonetheless, a few papers were studied without any taxonomy. The paper selection strategies were also ignored.

Alghamdi et al.[23] provided a considerable review of the various DL systems for the detection of COVID-19 using CNNs and other DL architectures via CXR images. This study made a good comparison based on the number of different types of models used to diagnose COVID-19, their performance analysis, and other factors. Nonetheless, the negative aspect of each reviewed paper was not considered.

Ghaderzadeh and Asadi[24] surveyed papers dealing with ML methods for the screening of COVID-19 patients using radiology modalities. This paper highlighted the importance of various DL models in the field of COVID-19 radiologic image processing for the diagnosis. However, the taxonomy was not taken into account.

The main goal of this survey is to review various ML methods with a concentration on CNN-based methods for the detection of COVID-19 disease using chest CT and X-ray images. In the following, we provide our main contributions.

1. A taxonomy is provided for classifying the reviewed papers as ML and DL models (category 1), more specifically CNN models; federated ML models (category 2); and ensemble ML models (category 3).
2. Some parameters of papers are considered and discussed; these include the number of samples and classes, data sources, and positive and nonpositive aspects of models and papers, as well as evaluation metrics.
3. The challenges of AI-based methods in the reviewed papers are discussed, and a projection of future works is also provided.

We used some common research databases, such as Elsevier, Google Scholar, Springer, IEEE, and ArXiv, to find the articles related to our subject. The keywords that we searched included "federated learning (FL)," "CNN," "ML," "DL," "ensemble learning," "chest X-ray," "chest CT images," "COVID," and "coronavirus." After reviewing abstracts of almost 100 papers, we chose 29 studies. In particular, we tried to consider articles that used DL methods, more specifically CNNs, to detect and evaluate COVID-19 using chest CT and X-ray images. Therefore, we ignored the papers that used other AI methods and data samples or investigated irrelevant issues, such as the influence of COVID-19 on the economy of a country.

The remainder of this paper is organized as follows. Section 2 includes important preliminary concepts. Section 3 presents the recently developed methods for diagnosis of COVID-19 via DL models and ML models with FL and ensemble learning. In Sec. 4, results, discussion, challenges, and future trends are presented. Finally, the conclusion of the paper is provided in Sec. 5.

2 Preliminary Concepts

In this section, we provide some basic definitions of DL, CNN, FL, and ensemble learning, which are helpful for understanding this paper.

2.1 *Deep Learning and Deep Convolutional Neural Network*

There is a type of artificial neural network with more than one hidden layer after the input layer and before the output layer called a deep neural network. It has been called deep because of its

multiple layers for data transformation. The deeper the network is, the more layers exist. The traditional deep neural network has fully connected layers with a bunch of neurons and an activation function. Each neuron has an associated set of weights in which each is multiplied by an input to the neuron. Deep neural networks often require considerable amounts of data for promising performance.[25]

A CNN is an artificial neural network. Generally, CNN is composed of convolution layers (CLs), pooling layers (PLs), nonlinear activation functions (NLAFs), and fully connected layers (FCLs).[26] CNN is a deep network that has many hidden layers and mimics the processing and recognition of images by the visual cortex of the brain. In the training process, this network includes feature extractors with special kinds of neural networks, the weights of which are specified via the training process. To be more specific, CNN extracts features of the input image in the first step and classifies the extracted features in the second step. In the feature extraction step, the images are converted by the convolution layer. After that, the dimension of the image is decreased by a pooling layer.[27] Figure 1 shows the architecture of a CNN. As it can be observed in this architecture, the network includes many convolution layers followed by pooling layers in the feature extraction stride. Convolutional filters, which are indicated by the dark blue blocks, carry out the convolution process to create the output feature maps from the input images or prior layer feature maps. The pooling filters are applied to lessen the data dimension of the input feature maps. Finally, the output of the last pooling layer is flattened. After the feature extraction stride, one or more dense layers are employed to classify the image using extracted features. Due to decreasing overfitting and establishing nonlinearity, an activation function and dropout layer are usually added.[28,29]

To train a DL model, several steps should be considered. First, a lot of real and relevant data should be acquired for training. Data preprocessing is the next step in creating a DL model. It is quite common for real-world data to show errors, inaccuracies, or outliers, as well as a lack of specific values or trends associated with certain characteristics. In this regard, data preprocessing can help to clean, format, and organize the raw data, so it is ready for training.[1,30–34]

The neural network mimics the brain's mechanism. A neural network consists of connections between nodes, which represent neurons of the brain, just as the brain is made up of connections between neurons. Information is stacked in the deep neural network as weights. Consequently, it is necessary to change the weights of the network to train it with new information. An algorithm for modifying weights systematically according to the available information is referred to as a learning rule. The learning rule is a critical part of neural network research because training is the only method by which the neural network can store information systematically. Training in a supervised learning model involves several steps, such as initializing the weights, taking the "input," and entering it into the neural network.[27,35–37] Then, the output from the network is obtained to estimate the error from the proper output. After achieving the error reduction by adjusting weights, some steps are repeated until the final output is reached.[27]

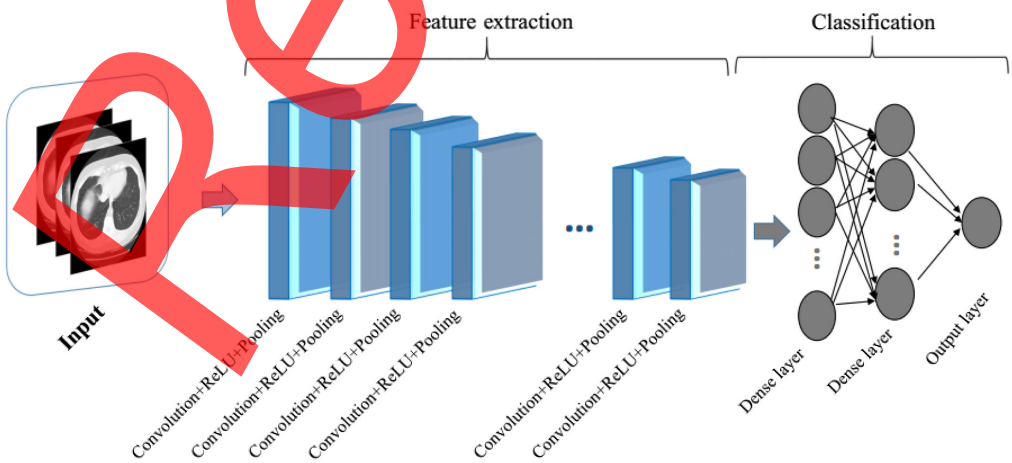

Fig. 1 CNN architecture.

2.2 *Federated Learning*

FL was introduced in 2017.[38] In the FL approach, several centers collaborate with each other for healthcare projects with no patient data sharing.[39] Its general framework includes several participants and a central server that delineates the computation rounds and selects the devices for the clients. Finally, the central server forms a global model by gathering the respective model updates of all participants.[40] Non-FL-based systems gather the data in the central server and train the model. This manner leads to leaking the data information out of the system, which is a violation of the policy and security rules and contracts. However, the FL models are intrinsically privacy-guarding. They keep the data in the devices where they are produced and only bring the global model to the devices.[41] The general framework of FL can be seen in Fig. 2.

Although FL was initially designed for mobile edge devices, it has attracted increasing attention in the healthcare field. Its privacy-preserving capacity makes it suitable for preserving private patient information and medical data, from free-text clinical reports to high-dimensional medical images.[40]

2.3 *Ensemble Learning*

In ensemble learning models, multiple learners are trained to solve the same problem. In contrast to other ML techniques that try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them for use. In fact, by combining various ML approaches, ensemble learning yields weak predictions that rely on features derived from a variety of data projections and merges outcomes with diverse voting tools to reach a more promising performances than that acquired from any single algorithm.[42] Figure 3 shows the framework of ensemble learning.

As can be seen in Fig. 3, there are various base learners that catch the patterns in the data in an independent manner. The ensemble learning model significantly decreases the error of the individual model; as a result, a better generalization can be done.[43] To train a complex individual neural network that has a significant number of layers and parameters, a great amount of time and memory is required. Instead, the structure can be decomposed into smaller and simpler single-base models, which present more accuracy. As a result, less time and memory are required for training compared with complex models. In the ensemble learning systems, patterns are captured in various areas of the input space using single base classifiers or learners, which leads to overall performance satisfaction.[42,43] To put together a satisfactory collection, it is generally recommended that the base learners be as accurate and diverse as possible. Learning accuracy can

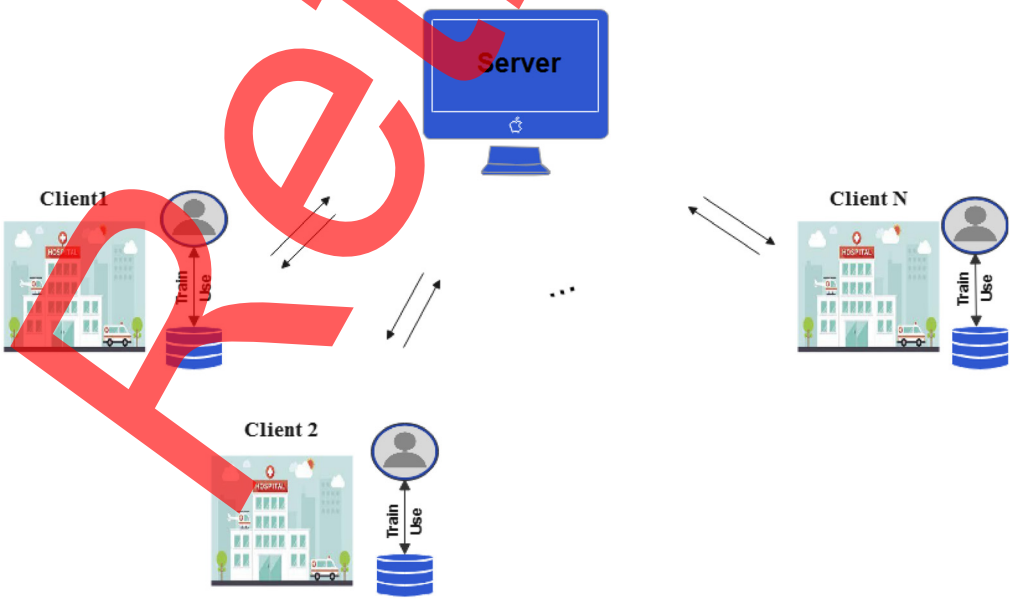

Fig. 2 FL framework.

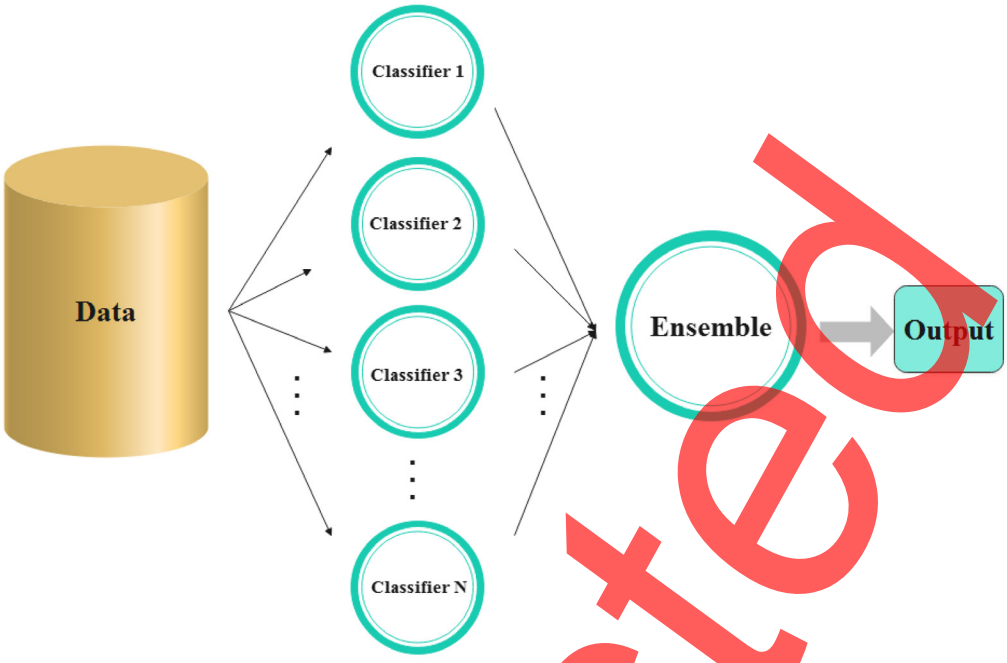

Fig. 3 Ensemble learning framework.

be determined in a number of ways, including cross-validation, hold-out tests, and other methods. Moreover, it is possible to introduce diversity into the base learners in a variety of ways, such as introducing randomness, subsampling the training examples, modifying outputs, altering the attributes, and employing more than one mechanism at once. Ensemble learning models generally consist of stacking, boosting, bagging, and so on. For the most part, the generalization of the learning system is improved remarkably by ensemble learning.[44]

2.4 *Important Evaluation Metrics*

In this section, we introduce several important evaluation metrics that are often used in studies to assess the performance of methods. The incorrect forecast proportion of the positives is called false positive (FP), and the correct classification proportion for the positive class is called true positive (TP). Also, the incorrect forecast proportion of the negatives is called false negative (FN), and the correct classification proportion of the negative class is called true negative (TN). Receiver operating characteristic curve (ROC-curve) is the graph of the TP proportion versus the FP proportion [TPR versus FPR defined by Eqs. (1) and (2)]

$$\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})}, \tag{1}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{(\mathrm{FP} + \mathrm{TN})}. \tag{2}$$

The area under the ROC curve shows AUC, which displays the cumulative measurement of all possible classification thresholds.[20]

Accuracy is a primary measure for evaluating the performance of classification problems and is defined as follows:

$$\mathrm{Accuracy} = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP})}. \tag{3}$$

Recall (sensitivity) is defined as Eq. (4) and identifies any infected patient by COVID-19 virus:

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \tag{4}$$

Specificity is defined as Eq. (5) and identifies any noninfected patient by COVID-19 virus:

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})}. \tag{5}$$

There is another important evaluation metric named precision that computes how the model performs precisely. Precision is defined as Eq. (6):

$$\text{Precision (PPV)} = \frac{\text{TP}}{(\text{TP} + \text{FP})}. \tag{6}$$

Discovering an equilibrium and balance between precision and sensitivity is the aim of $F1$-score, which is defined as

$$F1 - \text{score} = \frac{(2\text{TP})}{(2\text{TP} + \text{FN} + \text{FP})}. \tag{7}$$

To evaluate the linear inter-rater reliability, Kappa statistics (reliability measure) is used by the mathematical formula [Eq. (8)]. The reliability measure refers to this type of measurement, which takes into account the expected value after subtracting classification success from it. Rater reliability is important because it illustrates the extent to which the data gathered in the investigation are proper representations of the measured variables

$$\text{Kappa} = \frac{(\text{total accuracy} - \text{random accuracy})}{(1 - \text{random accuracy})}. \tag{8}$$

Finally, MCC is the abbreviate of the Mathew correlation coefficient and is defined as Eq. (9). Assuming that positive and negative cases are equally important, the MCC metric is used to sum models' performance in a single value. A high score can only be obtained if most of the predicted negative data samples and most of the predicted positive data samples are accurate.[20,45]

$$\text{MCC} = \frac{(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \tag{9}$$

We choose criteria because we need to somehow know how good or bad the model that we trained is. On the other hand, suppose that we want to examine whether the model suffers from overfitting or not. If there is no standard, we cannot check. Accuracy is the simplest possible model that we made for evaluation. Precision and recall are more complicated, but they are the most useful in medical work. However, the problem is that we have two of them for each class, so we have to aggregate them in a way to reach the $F1$-score. These criteria are all obtained from the confusion matrix, and there are other items as well.[46] Consider that we train the model with n data, but in the end, we report the accuracy with only one number. So it is impossible to reflect well on some aspects of accuracy. Therefore, this makes us focus more on the optimization of that evaluation criterion and not on optimizing the accuracy of the model.[46]

3 Literature Review

This section reviews the selected works in the field of COVID-19 detection and prediction using AI methods. We categorized the papers into three main categories: ML and DL models (Sec. 3.1), FL-based methods (Sec. 3.2), and ensemble learning-based methods (Sec. 3.3). The taxonomy of our study about ML and DL research works for COVID-19 diagnosis is presented in Fig. 4.

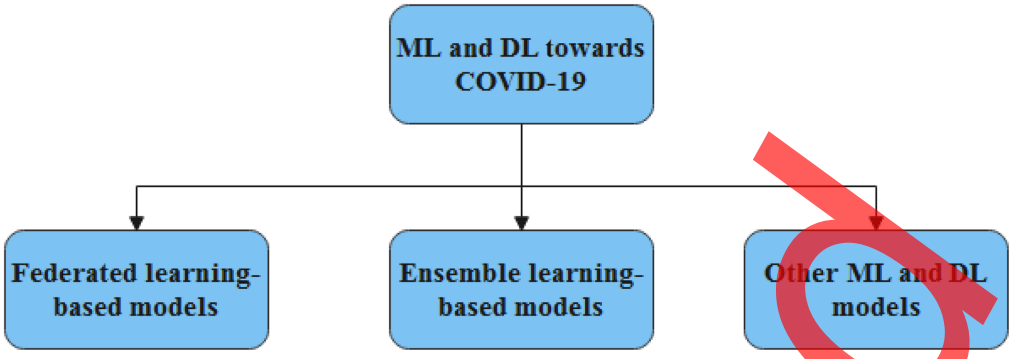

Fig. 4 Category of research works for COVID-19 detection.

3.1 *Machine Learning and Deep Learning Models*

To diagnose COVID-19, researchers have noticed the imaging patterns on the CT and X-ray sample images, which are very helpful in the detection process. Dilbag Singh et al.[45] proposed a DL method using multiobjective differential evolution (MODE) and CNNs to classify COVID-19 patients from their chest CT images. The proposed technique had good performance in terms of accuracy, *F*-measure, sensitivity, specificity, ROC, and Kappa statistics. Also, the results showed a low ratio of false-positive and false-negative for the proposed framework. However, the source of the dataset was not specified, and an important parameter called MCC was neglected for evaluation.

Özkaya et al.[3] developed a method called convolutional support vector machine (CSVM) for classification of chest CT images. To generate feature maps, linear support vector machine weights were used in the role of convolutional filters, which learn weights via a feed-forward learning method. In this paper, the CSVM model displays fast and high performance with 96.09% sensitivity, 94.03% accuracy, and 92.01% specificity. Also, the model leads to a reduction in the amount of work required by medical specialists. Moreover, unlike many CNN models, this method uses few parameters. Nevertheless, data augmentation methods were not considered. In addition, the specificity and precision rates of the proposed method are less than those of GoogleNet, VGG-16, DenseNet-201, and ResNet-50.

Khan et al.[4] introduced a new CNN-based technique to analyze the COVID-19 anomalies in chest CT samples with classification and segmentation stages. First, by performing classification, infected CT samples are separated from noninfected ones. After that, the segmentation stage of infected regions is implemented on the images. This can be useful in quantifying the spread of the infection. The suggested model (CoV-CTNet) presented high-performance evaluation values such as *F*-score (99%) and MCC (98%). Therefore, the infected samples can be effectively recognized by the proposed technique with a low number of FPs. Also, in the paper, a region approximation using the semantic segmentation technique (CoV-RASeg) was proposed to recognize and analyze infected regions on the images. The proposed segmentation method presented good performance. The low number of FPs and reduction in the search space to learn infectious patterns on CT images are the advantages of this method. However, it took too much time for training (3 days) for all networks.

Jaiswal et al.[47] utilized a deep transfer learning model with DenseNet-201 for extracting features from chest CT-scans using its own learned weights. The training, testing, and validation accuracy of the proposed model were achieved as 99.82%, 96.25%, and 97.40%, respectively. The proposed method showed low false-negative and false-positive rates. The over-fitting issue had the smallest effect on this method. However, this method needs more sophisticated feature extraction techniques. Also, MCC was neglected to further evaluate model performance.

Attallah et al.[48] used the computer-aided diagnosis (CAD) system "MULTI-DEEP" to detect COVID-19 and classify if patients are infected or not. The proposed method relies on the combination of several CNNs and uses various models such as GoogleNet, ResNet-18, Shuffle-Net, and AlexNet, and the framework comprises four scenarios. Four pre-trained models were employed in the first scenario to recognize whether the patient is infected or not. The high

accuracy and AUC were obtained by ResNet-18 with the values of 78.29% and 83.86%, respectively. The sensitivity and specificity values were 76.90% and 79.90%, respectively. For other CNNs (AlexNet, GoogleNet, and ShuffleNet), the accuracy, AUC, sensitivity, and specificity fluctuated within the ranges (71.99% − 75.89%), (77.54% − 81.66%), (68.70% − 76.20%), and (72.40% − 76.80%), respectively. In the second scenario, to train SVM classifiers separately, the features extracted from every pretrained CNN were used. Compared with other CNNs, the highest accuracy and AUC were achieved by ResNet-18 as 92.5% and 97%, respectively. Also, the sensitivity and specificity were obtained 93.30% and 91.80%, respectively. In the third scenario, the features extracted from every pre-trained model were applied with principal component analysis (PCA). After selecting some number of principal components from each deep feature set, the components were used to train SVM classifiers exclusively. A feature set with 50 elements from each of ShuffleNet, ResNet-18, and GoogleNet, and 150 elements from AlexNet led to an accuracy of 94%, a sensitivity of 94.90%, a specificity of 93.20%, and an AUC value of 98%. In the fourth scenario, the features extracted from every single pretrained model were combined to study the exerted impact on the performance and efficiency of the SVM classifier by this combination. For this scenario, the accuracy was obtained at 94.70%, which was slightly increased compared with the accuracy in the previous scenario with the value of 94%. The proposed method decreases the computational cost; however, for more evaluation of binary classification, the MCC value was not considered.

Shah et al.[49] proposed a self-developed model called CTnet-10 based on DL systems for COVID-19 detection by CT sample images. In this work, a comparison was drawn between the proposed model and other DL models, such as ResNet-50, VGG-16, DenseNet-169, VGG-19, and InceptionV3. For the CTnet-10, more accuracy was obtained than the ResNet-50 network and InceptionV3 with the value of 82.10%. The proposed method had satisfactory training, testing, and execution time. Nonetheless, the accuracy was unsatisfactory, and some vital evaluation measures, such as precision, specificity, and sensitivity, were ignored.

Ouchicha et al.[50] developed CVDNet, a deep CNN model, which is based on the residual neural network. To capture global and local features of the chest X-ray images, the model was built via two parallel levels with various kernel sizes. A recall of 96.84%, an accuracy of 96.69%, a precision of 96.72%, and an $F1$-score of 96.68% were achieved by the proposed model for three-class classification (normal, COVID-19, and viral pneumonia). The proposed technique showed a satisfactory performance on a small number of datasets. It diagnoses and detects COVID-19 infection in the shortest possible time. Nevertheless, slightly tangible misclassification was observed in some folds.

E. E.-D. Hemdan et al.[51] implemented a new CVOIDX-Net structure for automated diagnosis of COVID-19 using X-ray images. The new model comprises Xception, ResNetV2, VGG-19, InceptionV3, DenseNet-121, MobileNetV2, and InceptionResNetV2. DL systems based on the proposed framework (COVIDX-Net) demonstrated good classification performance. The classification performance of the VGG-19 and DenseNet models was notable with 91% and 89% $F1$-scores values for COVID-19 and normal classes, respectively. This method had satisfactory computational speed; nonetheless, data augmentation was ignored.

A. R. Martinez[52] used a multisource transfer learning (MSTL) to identify COVID-19 patients via chest CT samples. Moreover, an unsupervised label creation process was implemented in this work and led to improvement in deep residual networks' performance. The multisource transfer learning process includes three stages: source step, transition step, and target step. The recall was 89.70%, and the accuracy was 89.30%. The proposed method was highly sensitive. However, the preprocessing of data was very computationally expensive because the slice selection process for decomposing of three-dimensional scans into relevant two-dimensional slices was not automated.

Li et al.[53] developed a 3D DL framework, the COVID-19 detection neural network (COVNet), to draw out visual features of chest CT samples. In the extraction process, two-dimensional local and three-dimensional global representative features were extracted. The structure of the model includes RestNet-50 as the backbone for taking a series of CT slicing for input and generating features for the relevant slices. Then, by a max-pooling operation, selected features from all slices are fused. Finally, the last feature map is supplied to make a probability score for each type (COVID-19, CAP, and nonpneumonia). The sensitivity and

specificity were 90% and 96%, respectively, and the value of AUC was 96%. Therefore, the model presented good sensitivity, specificity, and AUC rates; nevertheless, it is not possible to identify what or which imaging features were applied for computing the output. The heatmap that was used in this paper was sufficient for visualizing the main areas in the images and identifying which one feature was utilized by the model.

Amyar et al.[54] proposed a multitask DL system for the detection of COVID-19 disease and segmentation using chest CT samples. For various datasets, the combination of three learning tasks: segmentation, classification, and reconstruction was performed. The new model demonstrated highly promising results. The model can enhance segmentation results even if there are not many segmentation ground truths. The AUC value was more than 97%. Nevertheless, the precision and $F1$-score were neglected to further evaluate model performance.

Li et al.[55] proposed a contrastive multitask CNN (CMT-CNN) model including two tasks. Diagnosis of coronavirus from normal cases and from other pneumonia is the main task, and encouraging local aggregation through a contrastive loss is the auxiliary task. First, each image is converted by a series of augmentation techniques. After that, the model is optimized for embedding augmented images. The simple auxiliary task presented powerful supervision for improving generalization. It was demonstrated that contrastive learning leads to accuracy enhancement for DL methods on X-ray and CT samples. So, there was no need for additional annotations. A generalization improvement was observed with this model. However, some important evaluation metrics such as precision and $F1$-score were not evaluated for validating the proposed method.

3.2 *Federated Machine Learning Models*

Qayyum et al.[56] proposed a clustered FL (CFL)-based collaborative learning framework. The model processes visual data at the edge by training a multimodal ML method. The model is able to detect COVID-19 in both X-ray and ultrasound imagery. Recall, precision, and $F1$-Score values for CFL (multimodal) reached 82%, 71%, and 76% for the COVID-19 class, and 97%, 94%, and 96% for the healthy class, respectively. The proposed method presented good performance in the presence of divergence in the data distribution from different sources (i.e., X-ray and ultrasound imagery); nevertheless, efficiency, security issues, and the optimization of CFL parameters are difficult. Additionally, there is a trade-off in the model performance for models trained with central data versus models trained with distributed data using FL.

Feki et al.[39] proposed a FL system based on ResNet-50 and VGG-16 for detection of COVID-19 via chest X-ray samples. The whole contribution of the paper is the comparison of two scenarios: classical centralized learning and FL. The suggested system showed remarkable performance in terms of sensitivity, accuracy, and specificity rates without sharing or centralizing private and sensitive data. However, several other important evaluative parameters, such as the $F1$-score, were not discussed.

Yang et al.[57] implemented a simple effective algorithm with FL on medical datasets using partial networks (FLOP) that share only a partial model between the server and the clients. The FLOP algorithm is applied to different model architectures (three-layer CNN, VGG-11, CovidNet, ResNet-50, MobileNet-v2, and ResNetXt). The mentioned algorithm led to a reduction in privacy and security risks. Experimental outcomes on both real-world medical datasets (COVIDx, Kvasir) and benchmark datasets (Fashion-MNIST, CIFAR-10) confirmed the utilities of the algorithm. FLOP achieved high global testing and local testing accuracy for different datasets. Nonetheless, the training of the models needs to be accelerated.

Kumar et al.[58] proposed a framework based on the DL models. The framework utilizes up-to-date data, enhances the diagnosis of disease via CT samples, and diffuses the data between centers as the privacy is preserved. As the data is acquired from diverse sources, a data normalization method was proposed to perfectly train the FL model. In addition, capsule network-based segmentation and classification were employed along with a method that can train a global model in a collaborative manner utilizing block-chain methods with FL. A high accuracy and sensitivity rate were achieved for the proposed blockchain-based FL framework. However, some other evaluation measures such as $F1$-score were neglected.

Yan et al.[59] implemented the FL framework for COVID-19 data training via X-ray samples. Then, a comparison was drawn between the performance of four DL models (CovidNet,

ResNeXt, MobileNet-v2, and ResNet-18) with the FL framework and without the framework. FL models showed great sensitivity and a high accuracy rate; nevertheless, the loss convergence rate caused by the use of FL decreased slightly. Also, some important evaluation metrics such as precision were not taken into account.

Zhang et al.[60] utilized a novel dynamic fusion relying on the FL method. First, the structure of the FL method based on dynamic fusion was presented for the analysis of medical diagnostic images. Then a decision-making mechanism was designed for clients to decide on each round's participation based on the local model performance. After that, an aggregation scheduling technique was presented to choose the clients in a dynamic manner in terms of each contributor's training time. The suggested architecture reduced the training time of ResNet-50 and ResNet-101 and performed well in terms of communication efficiency and accuracy. Nevertheless, the applied system only reduces training time for models with a large number of parameters; therefore, the GhostNet training time remained unchanged.

Xu et al.[61] implemented unified CT-COVID AI diagnostic initiative (UCADI) relying on the FL method. This new framework is a decentralized architecture. In this work, first, an initial AI CT method was used based on acquired data from three Tongji hospitals in Wuhan; a sensitivity of 97.50% and 72% were obtained for the Tongji CT test data and Wuhan Union Hospital test data, respectively. Then, a publicly available UCADI framework was developed to construct a federated system. Although the federated model behaved similarly to the initial model for the Tongji test data, it achieved a sensitivity value of 98% on the WU test data. This method presented good performance in terms of accuracy and sensitivity. However, improving the technical execution of the system, such as private information leakage from gradients and non-IID and unbalanced data distribution, is still required. Also, the number of local training iterations prior to global parameters needs to be updated and improved. Moreover, for more reassurance, some other parameters such as precision and $F1$-score need to be considered.

3.3 *Ensemble Machine Learning Models*

Das et al.[62] proposed an ensembling method based on CNN for automatic COVID-19 detection using X-ray image. The new method consists of DenseNet-201, Resnet-50V2, and Inceptionv3. The classification accuracy and sensitivity of the proposed model reached 95.7% and 98%, respectively. The suggested model showed high rates of sensitivity, accuracy, and $F1$-score; nonetheless, the specificity, MCC, and kappa were not considered to further evaluate model performance.

Zhou et al.[44] used an ensemble DL system (EDL-COVID) for diagnosis of COVID-19 via chest CT samples. ResNet, GoogleNet, and AlexNet models were trained using transfer learning, and then initialization parameters were determined. The softmax function was applied as the classification algorithm for building three-component classifiers: GoogleNet-Softmax, ResNet-Softmax, and AlexNet-Softmax. The relative majority vote algorithm was utilized for building the ensemble classifier EDL-COVID. A high accuracy, specificity, and sensitivity rate were achieved with a detection speed rate of 342.92 s for the proposed model. However, the training time was increased compared with single classifiers. Also, slightly tangible misclassification can be seen.

Rajaraman et al.[63] proposed an iteratively pruned DL model for the detection of COVID-19 using chest X-rays. In this model, after training and evaluation of a custom CNN and some pretrained models, the learned information is transmitted and fine-tuned to enhance the efficiency and generalization of the classification model. Models with the best performance are pruned in an iterative manner for reduction of complexity and enhancement of memory proficiency. The predictions of pruned models with outstanding performance are then fused through various ensemble methods. The weighted average of the best-performing pruned models improved the model's efficiency, with AUC and accuracy of 99.72% and 99.01%, respectively. The proposed model had a high rate for the evaluation metrics. However, some items, such as dataset size and inherent mutability, and the computational resources needed for successful deployment and use limit the success of the method.

Gianchandani et al.[64] implemented ensemble deep transfer learning models relying on CNNs to detect COVID-19 using X-ray images. For two class-classification and three-class

classification, in terms of AUC, F-measure, accuracy, recall, and precision, the proposed ensemble learning system showed the best performance compared with other individual DL models (VGG-16, ResNet-152V2, InceptionResNetV2, and DenseNet-201). Also, the model showed good generalization performance with a decline in the false predictions. Although most performance metrics were considered in this paper, MCC was neglected to further evaluate model performance.

Wiysobunri et al.[42] utilized an ensemble learning system relying on a majority voting strategy and X-ray images. The developed architecture includes DenseNet-201, ResNet-34, VGG-19, ResNet-50, and MobileNet-V2. The performance accuracy of the proposed model was 99%. The proposed method performed well in terms of several evaluation measures, as well as reducing the error rate of misclassification. Nevertheless, several valuable performance metrics such as MCC and kappa statistics were ignored for further evaluation and reliability.

Upadhyay et al.[65] utilized an ensemble learning method to detect coronavirus disease by feature boosting from X-ray sample images. In this method, three sets of images were created: normalized RGB X-ray images as the first case, normalized X-ray images in HSV colorspace as the second case, and the Prewitt edge images paralleling each image of the first set as the third case. In the proposed framework, three similar backbone models called VGG-16 were trained. Additionally, the outputs of these base models were fused via meta-model (logistic regression). The fused model (meta-model) was stronger than every single model (base models 1 to 3). It showed good results for all three types of classifications: normal and abnormal, pneumonia and COVID-19, and the third type of normal, COVID-19 and pneumonia. Furthermore, after being combined, the evaluation measures achieved were at a maximum rate compared with the three individual models and were above 90%. Also, the training time was satisfactory. However, the real picture of the model will appear if more test data is available. Moreover, considering some other evaluation metrics such as MCC would provide additional reassurance. Also, a much more comprehensive handcrafted feature-based image analysis should be done.

Tang et al.[66] designed EDL-COVID, an ensemble DL system via X-ray modality relying on the COVID-Net (open-sourced network architecture). In the model, several snapshot models of COVID-Net are generated and combined by a suggested weighted averaging ensembling strategy. The sensitivity and positive predictive value (PPV) were 96% and 94.10%, respectively. The model showed excellent accuracy, sensitivity, and PPV rate for the COVID-19 class. Nonetheless, compared with some other models, a lower sensitivity and PPV rate were acquired for normal class and pneumonia, respectively.

Saha et al.[67] developed an automated diagnosis method called EMCNet. For the extraction of deep and high-level features from X-ray images, a CNN model was applied with a concentration on the simplicity of the model. Binary ML classifiers (support vector machine, random forest, decision tree, and AdaBoost) were used to diagnose COVID-19 disease. Eventually, outputs of classifiers were fused to design an ensemble model. An accuracy of 98.91%, precision of 100%, recall of 97.82%, and $F1$-score of 98.89% were obtained. Although the proposed model presented good classification performance, it has still some misclassification, such as labeling some COVID-19-positive items as negative.

Öksüz et al.[68] proposed an end-to-end DL method called ensemble-CVDNet. In this method, ShuffleNet, EfficientNet-B0, and SqueezeNet were combined at various depths. For the suggested model, 98.30% accuracy, 97.78% sensitivity, and 97.61% $F1$-score were attained using 5.62M parameters. Processing and prediction times and the number of model parameters were relatively satisfactory. However, to decrease the number of parameters, the spatial resolution of input for the model was set to $224 \times 224 \times 3$, which may bring about the dropping of several fine-grained features.

Chowdhury et al.[69] designed the ECOVNet framework based on CNN with the class activation maps from the COVIDx dataset. In the proposed architecture, EfficientNet (EfficientNet B0 to B5 base models) was applied as feature extractors, and fine-tuned pretrained weights were assigned for relevant COVID-19 diagnosis. An ensemble learning strategy, particularly soft ensemble, enhanced the results with 97% accuracy and 100% recall and precision. However, as deeper base models are considered, the recall progressively rises, but it declines by 4% from ECOVNet-B0 to ECOVNet-B1. A similar situation can be observed for the $F1$-score (with a fall of 1%).

4 Discussion and Open Issues

In this work, 29 papers that used ML models, especially DL models, for the detection of COVID-19 diseases, were reviewed. In particular, CNN models were widely used. We took into consideration the custom DL and pretrained models. CT scans, X-ray images, and ultrasound were considered for implementation in the reviewed articles. We discussed various features including the number of classes, data sources, positive aspects, and nonpositive aspects of each model and the paper, and the evaluation parameters for the performance of models. We summarized these features, which are presented in different tables. In this study, we classified papers into three taxonomies: federated ML models, the ensemble of ML models, and other ML and DL models. Some features of each paper such as modality, experimental tools, and positive and negative aspects of each paper and proposed models are presented in Tables 1–3. Some information about data sources and the number of classes are available in Tables 4–6. Moreover, evaluation metrics such as sensitivity, accuracy, precision, specificity, AUC, $F1$-score, etc. are given in Tables 7–9.

Figure 5 shows the percentage of each modality used in reviewed papers. It can be seen that most researchers only consider X-ray or CT images as useful for the COVID-19 data samples. In fact, CT images were used in 12 papers, and X-ray images were used in 13 papers. Both CT and X-ray images were used in 3 papers, and X-ray and ultrasound were used in 1 paper. Therefore, X-ray images were preferred in most articles for experiments, especially articles that used ensemble learning. Among the papers that mentioned their data sources, as can be seen in Tables 4–6, and Fig. 6, 39% of papers used a single source of data, and 61% of papers used multiple sources of data. Also, data augmentation was implemented only in some papers. Although large numbers of image datasets have been used for some systems, limited datasets for coronavirus disease remained a major challenge for researchers. In terms of the number of classes, according to Fig. 7, most of the suggested models only took into consideration binary classification (45%), and some others considered multiple classes (35% three classes, 17% both binary and multiple classes, and 3% four classes). Figure 8 presents the programming language tools used in the reviewed items. It can be seen that a maximum number of researchers preferred to use Python (14% Keras and Tensorflow library, 10% PyTorch, and 14% did not mention the library). 17% of researchers used MATLAB for their experiments.

It can be observed in Tables 7–9 that the accuracy, precision, $F1$-score, specificity, sensitivity, and AUC of most reviewed papers reached above 90%. The highest accuracy and sensitivity rate are associated with the ensemble ML category with the values of 99.21% and 100%, respectively. In each category, we computed the average accuracy and sensitivity (two of the most important evaluation metrics) of their models and compared their classification performance for added assessment. Their results are shown in Figs. 9 and 10. They showed the average accuracy of 93.38%, 95.71%, and 97.09% for category 1 (ML and DL models), 2 (federated ML models), and 3 (ensemble ML models), respectively. Moreover, the average sensitivity of 94.82%, 93.7%, and 97.83% were obtained for categories 1–3, respectively. Although the number of papers that considered the accuracy metric was not enough for category 2, its performance measures are close to their counterpart in category 1. Moreover, the ensemble ML models present the best performance in terms of all evaluation metrics compared with the two other categories. Thus, ensemble learning-based systems can be widely applied for the diagnosis of coronavirus disease.

We also investigated some review papers on using AI-based systems for the detection and analysis of COVID-19 disease using X-ray and chest CT images. Table 10 displays related review papers, and some parameters such as the paper selection process, future research, taxonomy, experimental tools, and negative aspect of the work are depicted. As a case study, FL and ensemble learning models are not examined in the related surveys. Therefore, our study just surveyed the FL and ensemble learning-based papers. It is obvious that the most studied surveys did not consider the negative aspects of proposed methods and work. Also, some of the articles were reviewed without any taxonomy. In addition, the paper selection process and experimental tools were not clear in some articles. Our study reviews papers related to COVID-19 diagnosis based on AI methods and provides information on the paper selection process, experimental tools, and positive and negative aspects of each work or developed method. Future trends and classifications of articles are clear.

Table 1 Summary of ML and DL classification methods.

Work	Modality	Proposed method	Positive aspect	Negative aspect	Tools/toolbox
45	CT	MODE-based CNN	Good accuracy rate, less false-negative, and false-positive values	Not specified dataset, neglected MCC	MATLAB 2019a
3	CT	Convolutional support vector machine	Fast and high-performance framework with reduced workload of medical experts, low number of parameters, low training time	Without data augmentation methods, low specificity, and precision rate	MATLAB 2020a
4	CT	Deep CNN framework for classification and segmentation	Minimum number of FPs, reduction in the search space	Too much training time (3 days)	MATLAB 2019b
47	CT	DenseNet-201 based deep transfer learning model	Low FN and FP rate, smallest effect from the over-fitting issue, satisfactory rates of performance metrics	Need more datasets and more sophisticated feature extraction techniques, neglected MCC	Not mentioned
48	CT	Multideep system based on the combination of several CNN (ResNet-18, GoogleNet, ShuffleNet, and AlexNet)	Reduction in the computational cost by ~32%	Ignored MCC	Not mentioned
49	CT	A self-developed model named CTnet-10 based on CNNs	The low training, testing, and execution time	Unsatisfactory accuracy, ignored precision, specificity, and sensitivity	Not mentioned
50	X-ray	CVDNet, a deep CNN system based on the residual neural network	A satisfactory performance on a small number of data, diagnosis and detection of COVID-19 infection in a shortest possible time	Tangible misclassification in some folds	Tensorflow
51	X-ray	COVIDX-Net based on seven different architectures of CNNs	Relatively satisfactory training and testing time	Ignored data augmentation	Python, Keras
52	CT	A multisource fine-tuning approach (ResNet-50V2, ResNet-101V2, DenseNet-121, and DenseNet-169)	Highly sensitive	Expensive computational process	Tensorflow's Keras
53	CT	A three-dimensional DL framework (COVNet, RestNet-50 as backbone)	Good sensitivity, specificity, and AUC rate	The lack of transparency and interpretability	R
54	CT	A multitask DL method for segmentation and classification task	Improving segmentation results, high AUC rate	Neglected precision and *F*1 -score	Keras
55	CT and X-ray	A contrastive multitask convolutional neural network (CMT-CNN), (VGG-19, ResNet-50, and EfficientNet)	Improvement in generalization without additional annotations	Neglected some evaluation metrics such as precision and *F*1-score	PyTorch

Table 2 Summary of federated ML models.

Work	Modality	Proposed method	Positive aspect	Negative aspect	Tools/toolbox
56	X-ray and ultrasound	A CFL-based collaborative learning framework, [a clustered FL (multimodal)]	Good performance in presence of divergence in the data distribution	Difficulty in efficiency, security issues, and the optimization of CFL parameters, a trade-off in the model performance	TensorFlow
39	X-ray images	FL framework using deep CNNs (VGG-16 and ResNet-50 as backbone) with four clients	Remarkable performance in terms of accuracy, sensitivity, and specificity	Neglected *F*1-score	Not mentioned
57	X-ray or CT images	FLOP algorithm for different architectures (three-layer CNN, CovidNet, ResNet-50, MobileNet-v2, ResNetXt)	Strengthening the data protection, high global testing and local testing accuracy for different datasets	Slightly slow speed for training the model	PyTorch
58	CT images	Blockchain based FL (a proposed data normalization technique, capsule network-based segmentation and classification)	Improving the recognition of CT images, high accuracy and sensitivity rate	Neglected *F*1-score	Not mentioned
59	X-ray images	Four different networks (MobileNet, ResNet-18, ResNeXt, and COVIDNet) with FL framework	Good performance in terms of accuracy, and sensitivity	Slightly reduced loss convergence rate by the use of FL, some important evaluation metrics such as precision, and *F*1-score were neglected	PyTorch
60	CT and X-ray images	A novel dynamic fusion based-FL (GhostNet, ResNet-50, ResNet-101 as backbone)	Significant reduction in the training time for poor models with large amounts of parameters, good performance in terms of accuracy and communication efficiency	No change on GhostNet training time	Python
61	CT images	An FL-based Unified CT-COVID AI diagnostic initiative (UCADI)	High performance in terms of accuracy and sensitivity	The technical execution in the system needs to be enhanced, some evaluation parameters such as precision and *F*1-score were neglected	Not mentioned

Table 3 Summary of ensemble ML models.

Work	Modality	Proposed method	Positive aspect	Negative aspect	Tools/toolbox
62	X-ray images	Ensemble learning with CNNs (DenseNet-201, ResNet50v2, and Inceptionv3)	Inexpensive, high rate of sensitivity, accuracy, and *F*1-Score	Neglected specificity, MCC, and kappa	Python 3.7 and TensorFlow 2.2.0
44	CT images	EDL-COVID: ensemble DL model (AlexNet, GoogleNet, and ResNet)	Fast detection speed, high accuracy, specificity, and sensitivity rate	Increase in training time compared with individual models, slightly tangible misclassification	MATLAB2019a
63	X-ray images	Ensemble pruned DL models (VGG-16, VGG-19, and Inception-V3 models as backbone for multiclass classification)	High performance metrics, a considerable decline in the number of parameters of the pruned models compared with unpruned models	Limiting success of the method by the size of dataset and inherent mutability, and computational resources needed for successful deployment and use	Not mentioned
64	X-ray images	An ensemble deep transfer learning model	Good rates of performance metrics, good generalization and minimizing the false predictions	Neglected MCC	Not mentioned
42	X-ray images	An ensemble DL model based on majority voting strategy by DenseNet-201, ResNet-34, ResNet-50, VGG-19, and MobileNet-V2	High performance metrics, decrease in the error-rate of misclassification	Neglected MCC, and kappa statistics	Not mentioned
65	X-ray images	An ensemble learning-based method (VGG-16 as base-learners)	Evaluation measures above 90%, low training time	Neglected MCC	Not mentioned
66	X-ray images	An ensemble DL model (EDL-COVID model)	Excellent evaluation parameter rate for sensitivity, accuracy, and PPV for COVID-19 class	Low sensitivity and PPV rate for normal class and pneumonia, respectively	TensorFlow
67	X-ray images	EMCNet (A CNN system consisting of 20 layers)	High rate of classification performance metrics	Wrong classification of some COVID-19-positive cases as negative	Not mentioned
68	X-ray images	A DL based end-to-end architecture using ensembles of networks (ensemble-CVDNet)	Relatively satisfactory processing and prediction times, and number of parameters, high rate for evaluation measures	Missing some fine-grained features	MATLAB
69	X-ray images	ECOVNet: The ensemble of deep CNNs relying on EfficientNet B0 to B5	High performance measures in terms of accuracy, precision, and recall	Decrease in the recall of base models by 4% from ECOVNet-B0 to ECOVNet-B1, the similar situation for *F*1-score	Python/Keras

Table 4 Datasest information of ML and DL methods.

Work	No. of classes	Dataset information
45	Two (COVID + and COVID −)	Not mentioned
3	Two (COVID-19 and non-COVID)	SARS-CoV-2 CT scan dataset[70] including 2492 CT images (1262 COVID-19 and 1230 non-COVID)
4	Two (infected and healthy)	Standard CT images dataset[71]
47	Two [COVID-19 (−) and COVID (+)]	SARS-CoV-2 CT scan[70] including 2492 CT-scans
48	Two (non-COVID-19 and COVID-19)	Ref. 72 including 347 COVID-19 images and 397 non-COVID-19 images
49	Two (COVID positive and COVID negative)	738 CT scan images from an open-source dataset
50	Three (COVID-19, normal, and pneumonia)	An open source,[73] 2905 chest X-ray images, (1341 normal samples, 219 COVID-19 samples, and 1345 viral pneumonia samples)
51	Two (positive COVID-19, negative COVID-19)	The public datasets[74,75]
52	Two (COVID-19 and normal)	Source dataset: ILSVRC dataset (the ImageNet large scale visual recognition challenge)[76,77] Transition dataset: The 2015 SPIE-AAPM-NCI lung nodule classification challenge,[78] including 22,489 CT samples belonging to 70 patients Target dataset: An open-source COVID-19 CT samples.[72] 397 non-COVID-19 or normal CT images and 349 COVID-19
53	Three (COVID-19, CAP, and nonpneumonia)	Dataset acquired from six medical centers, 4352 3D CT samples belonging to 3322 patients, 40% (1735) CAP (community-acquired pneumonia) images, 30% (1292) COVID-19 images, and 30% (1325) nonpneumonia abnormalities samples
54	Three (COVID-19 +, normal, and others)	Three different datasets acquired from various hospitals including 1369 CT images, First:[72] Consist of 347 COVID-19 samples and 397 non-COVID samples with diverse types of pathologies Second:[79] Includes 100 COVID-19 CT images with ground truths lesions segmentation, Third: Acquired from HBCC (the Henri Becquerel Cancer Center), Rouen in France, consist of 98 lung cancer samples and 425 CT samples of normal patients
55	Two (COVID-19, other conditions), three (COVID-19, other pneumonia, and normal control)	A CT image dataset and an X-ray image dataset acquired from open datasets and data collected in a hospital, 4758 CT samples and 5821 CT samples

There are various challenges and limitations for researchers to use AI models for screening the COVID-19 virus. In this section, we refer to some limitations of ML methods and provide useful insight for future researchers.

4.1 *Data Volume*

As DL models are data-hungry, the insufficient dataset was one of the common and important challenges for all reviewed papers. Indeed, DL models do not show successful performance with finite data. Sometimes tons of network parameters are required for correct estimation, and this

Table 5 Dataset information of federated ML models.

Work	No. of classes	Dataset information
56	2 (COVID-19 and healthy)	Two datasets from different sources, one containing chest X-ray[80] (223 COVID-19, 1341 healthy), chest ultrasound images[81] (399 COVID-19, 146 healthy)
39	Two (COVID-19 and non-COVID-19)	The COVID-19 X-ray images available at Ref. 74 and 108 normal X-ray samples are randomly chosen from.[82] 108 X-ray sample from 76 patients with confirmation as infected with COVID-19, and 108 normal X-ray samples from healthy patients, $K = 4$ clients
57	COVIDx: three (normal, pneumonia, and COVID-19)	COVIDx (The open-access benchmark dataset) consists of 1579 for testing (885 normal, 594 pneumonia, and 100 COVID-19 images) and 13,954 images (7966 samples of normal, 517 samples of COVID-19, and 5471 samples of pneumonia) for training, different datasets were used in this paper (Kvasir dataset, FashionMNIST, and CIFAR-10). We only consider the results of COVIDx
58	Three (normal, COVID-19, and viral Pneu.)	Reference 83 include 34,006 CT samples from 89 patients, 28,395 CT scan slices belonging to positive COVID-19 patients, the dataset collected from the three different hospitals, and third party dataset[72,84]
59	Three (normal, pneumonia, and COVID-19)	COVIDx: An open access dataset, 15,282 images in this dataset
60	CT: two (COVID-19 and negative), X-ray: three (COVID-19, negative, and viral pneumonia)	Three datasets[73,85,86] 746 CT images: 349 samples as COVID-19, 397 samples as non-COVID-19/negative cases X-ray images: 274 samples as COVID-19 case, 1341 samples as non-COVID-19/negative cases, and 1345 samples as viral pneumonia, three clients
61	Four (COVID-19, pneumonia, bacterial pneumonia, and healthy)	5732 CT images from 1276 individuals acquired from several locations of Tongji Hospital including Tongji Hospital Main Campus, Tongji Optical Valley Hospital, and Tongji Sino-French New City Hospital (432 individuals infected by COVID-19, 76 patients with different viral pneumonia, 350 patients with bacterial pneumonia, and 418 individuals with clinical symptoms of respiratory system)

goal is achievable by only a considerable amount of data. Moreover, compared with other tasks including image or speech recognition, the screening of diseases and their variability is much more cumbersome, and a much greater amount of medical data is required. Therefore, this makes it difficult for researchers to claim with certainty that the proposed models can detect the COVID-19 disease and classify the images with high-performance metrics. We hope that, in the future, more datasets will be available for addressing this issue.

4.2 *Data Quality and Variety*

Healthcare data samples are tremendously noisy, incomplete, heterogeneous, and ambiguous. Hence, due to the vulnerability of real-life data to noise and other factors, annotated data need to be taken into account for reassurance of the model performance. Additionally, the training and test sets need to be varied and independent in these studies to obtain more accurate performance of models. Thus, the variety of datasets needs more consideration for future works.

4.3 *Privacy and Security*

Another issue in applying ML models is data privacy protection and security. Although FL can solve such problems, this technique was implemented in only a few studies. Therefore, it is

Table 6 Dataset information of ensemble ML models.

Work	No. of classes	Dataset information
62	Two (COVID +ve and −ve cases)	Different open sources,[80,86–91] 468 samples of COVID −ve patients, and 538 samples of COVID +ve
44	Three (normal lungs, lung tumors, and COVID-19)	2933 chest CT samples of COVID-19 patients acquired from public databases, prior publications, and authoritative media reports. 2500 high-quality CT images were acquired after preprocessing. 2500 lung CT as normal case and 2500 CT samples of lung tumor from general hospital of Ningxia Medical University in China
63	Two (normal and COVID-19 pneumonia), three (COVID-19, bacterial pneumonia, and normal), ensemble learning only was implemented for the multiclass classification	(1) Pediatric CXR dataset,[92] (2) RSNA CXR dataset,[93] (3) Twitter COVID-19 CXR dataset,[87] and (4) Montreal COVID-19 CXR dataset[94]
64	Three (COVID-19, normal, and pneumonia), two (COVID positive and COVID negative)	1. The first dataset from Kaggle datasets resource.[95] 2. The second dataset from the University of Dhaka and Qatar University along with medical practitioners and collaborators[96] (1203 X-ray samples of normal, COVID, and pneumonia subjects), this dataset is used for multiclass classification. Image samples from Ref. 97 is also added for the COVID class
42	Two (positive COVID-19 and normal)	First dataset from the Kaggle website[73] (1341 X-ray samples as normal case, 219 positive X-ray samples of COVID-19 case, and 1345 samples as viral pneumonia) The second dataset from the Github repository[98] (100 chest X-ray samples. 50 cases for COVID-19 positive and 50 cases for normal)
65	Type-I: 2 (normal/abnormal), Type-II: 2 (pneumonia/COVID-19) and Type-III: 3 (normal COVID-19/pneumonia)	1. 191 X-ray samples for COVID-19 class from updated datasets of Cohen et al.[80] 2. 5863 X-ray samples for pneumonia and normal classes from public database available on Kaggle (191 pneumonia and 382 normal samples were opted) 3. Also, proposed algorithm was tested on six locally gathered X-ray samples of COVID-19
66	Three (normal, pneumonia, and COVID-19)	COVIDx dataset,[99] includes 15,477 CXR samples (6053 pneumonia, 573 COVID-19 cases, and 8851 normal case)
67	Two (COVID-19 and normal)	First: 660 positive COVID-19 X-ray samples by Cohen et al.[80] (Just 500 positive samples of COVID-19 class were applied) Second: 1800 COVID-19 X-ray samples from,[100] TCIA,[101] SIRM database,[102] and Mendeley[103,104] Third: 2300 images of normal case of X-ray samples from[90] and NIH X-ray samples,[105] a total of 4600 images
68	Three (COVID-19, normal, and viral pneumonia cases)	The COVID-19 radiography database[96] on Kaggle with 1341 X-ray samples for healthy case, 219 samples for COVID-19 case, and 1345 samples for viral pneumonia class
69	Three (COVID-19, normal, and pneumonia)	COVIDx[106]

necessary to be taken into account that the information of patients should not be leaked out of the system and data should not be shared directly. Ensemble learning models in the reviewed papers presented the best performance in terms of evaluation metrics. Thus, they can be used as powerful tools for the diagnosis of COVID-19 disease. However, security and privacy issues of data were not considered in the mentioned models. As a case study, the combination of ensemble ML models and an FL framework can be taken into account for future research.

Table 7 Performance evaluation of ML and DL models.

Work	Performance metrics
3 [CSVM (7×7, 3×3, 1×1)]	Accuracy (%): 94.03, sensitivity (%): 96.09, specificity (%): 92.01, precision (%): 92.19, *F*-score: 94.10, MCC (%): 88.15, kappa (%): 88.07, training time: 25 min 36 s
4	Accuracy (%): 98.80, specificity (%): 99.0, precision (%): 99.0, AUC (%): 99.0, *F*-score (%): 99.0, recall (%): 99.0 MCC (%): 98.0, training time: 3 days
47	Accuracy (%): 96.25, specificity (%): 96.21, precision (%): 96.29, AUC (%): 97.0, recall (%): 96.29, F-measure (%): 96.29
48	Accuracy (%): 94.70, sensitivity (%): 95.60, specificity (%): 93.70, AUC (%): 98.0, precision (%): 93.40, *F*1-score (%): 94.50, time (s): 33.765
49	Accuracy (%): 82.10, training time: 130 s, testing time: 900 ms, execution time: 12.33 ms
50	Accuracy (%): 96.69, precision (%): 96.72, *F*1-score (%): 96.68, recall (%): 96.84
51	Accuracy (%): 90.0, *F*1-score (%): 91.0, recall (%): 100, precision (%): 83.0
52	Accuracy (%): 89.30, recall (%): 89.70, *F*1-score (%): 89.70, precision (%): 89.70
53 (COVID-19 class)	Sensitivity (%): 90.0, specificity (%): 96.0, AUC (%): 96.0
54	Accuracy (%): 94.67, sensitivity (%): 96.0, specificity (%): 92.0, AUC (%): 97.0
55	CT binary Cls: accuracy (%): 93.46, sensitivity (%): 90.57, specificity (%): 90.84, AUC (%): 89.22
	CT ternary Cls: accuracy (%): 91.45
	X-ray binary Cls: accuracy (%): 97.23, sensitivity (%): 92.97, specificity (%): 91.91, AUC (%): 92.13
	X-ray ternary Cls: accuracy (%): 93.49

Table 8 Performance evaluation of federated ML models.

Work	Performance metrics
56	X-ray, COVID-19 class: precision (%): 71.0, recall (%): 82.0, *F*1-score (%): 76.0
	Ultrasound, COVID-19 class: precision (%): 93.0, recall (%): 95.0, *F*1-score (%): 94.0
39	FL-VGG-16: accuracy (%): 93.57, sensitivity (%): 95.03, specificity (%): 92.12
	FL-VGG-16 + data aug: accuracy (%): 94.40, sensitivity (%): 96.15, specificity (%): 92.66
	FL-ResNet-50: accuracy (%): 95.4, sensitivity (%): 96.03, specificity (%): 94.78
	FL-ResNet-50 + data aug: accuracy (%): 97.0, sensitivity (%): 98.11, specificity (%): 95.89
57	Local testing accuracy of FLOP: COVID-Net (%): 91.52 ± 0.13, MobileNet-v2 (%): 91.90 ± 0.63, ResNet-50 (%): 94.51 ± 0.27, ResNeXt (%): 94.15 ± 1.06
	Global testing accuracy: COVID-Net (%): 88.51 ± 0.26, MobileNet-v2 (%): 89.65 ± 3.63, ResNet-50 (%): 91.45 ± 0.25, ResNeXt (%): 91.20 ± 0.81
58	Accuracy (%): 98.68, sensitivity (%): 98
61	Tongji test data sensitivity (%): 97.5, Wuhan Union Hospital test data sensitivity (%): 98
59	COVID-Net sensitivity: 89.17% ± 0.015%
	MobileNet v2 sensitivity: 86.83% ± 0.017%
	ResNet18 sensitivity: 91.26% ± 0.014%
	ResNeXt sensitivity: 90.37% ± 0.015%

Table 9 Performance evaluation of ensemble ML models.

Work	Performance metrics
62	Accuracy (%): 95.70, sensitivity (%): 98.0, *F*1-Score (%): 96.20
44	EDL-COVID classification: Detection speed: 342.92 s, accuracy (%): 99.05, sensitivity (%): 99.05, specificity (%): 99.60, *F*-score(%): 98.59, MCC (%): 97.89
63	Accuracy (%): 99.01, AUC (%): 99.72, sensitivity (%): 99.01, precision (%): 99.01, *F*-score (%): 99.01, MCC (%): 98.20
64	Three classes: accuracy (%): 99.21, precision (%): 99.0, recall (%): 99.0, *F*1-score (%): 99.0, training time: 6 mi
	Classes: precision (%): 95.90, *F*1-score (%): 96.10, sensitivity (%): 96.40, specificity (%): 95.80, accuracy (%): 96.15, training time: 10 min
42	Performance accuracy for ensemble of five models (%): 99.0
65	Training time: 700 s
	Type I: sensitivity (%): 96.10, accuracy (%): 97.40, specificity (%): 98.7, precision (%): 98.7, *F*1-score (%): 97.4
	Type II: sensitivity (%): 97.40, accuracy (%): 98.70, specificity (%): 100, *F*1-score (%): 98.70, precision (%): 100
	Type III: sensitivity (%): 97.40, accuracy (%): 88.70, specificity (%): 97.40, *F*1-score (%): 96.10, precision (%): 94.90
66	Accuracy (%): 95.0
	PPV (%): normal: 96.40, pneumonia: 93.10, COVID-19: 94.10
	Sensitivity (%): normal: 95.0, pneumonia: 94.80, COVID-19: 96.0
67	Accuracy (%): 98.91, precision (%): 100, recall (%): 97.82, *F*1-score (%): 98.89
68	Accuracy (%): 98.30, sensitivity (%): 97.78, *F*1-score (%): 97.61, TPR (%): 97.78, specificity (%): 98.48
69	Accuracy (%): 97, precision, *F*1-score and recall (%): 100

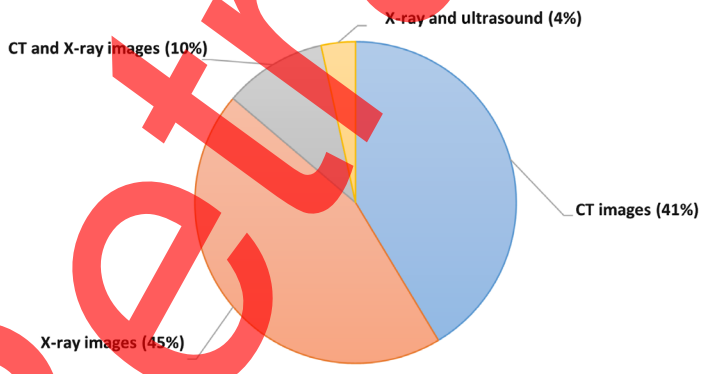

Fig. 5 Modality used by reviewed papers.

4.4 *Time Consideration*

The time issue is highly serious in all sorts of healthcare-relevant issues, especially for a new-found virus such as COVID-19. To understand the status of patients properly, make decisions, and provide on-time clinical support, a time-sensitive DL model is crucial.

4.5 *Transparency and Interpretability*

In almost all DL models, realizing which features of the input images are being applied to predict the output is difficult. Some methods visualize the critical regions in the scans, but there is a need

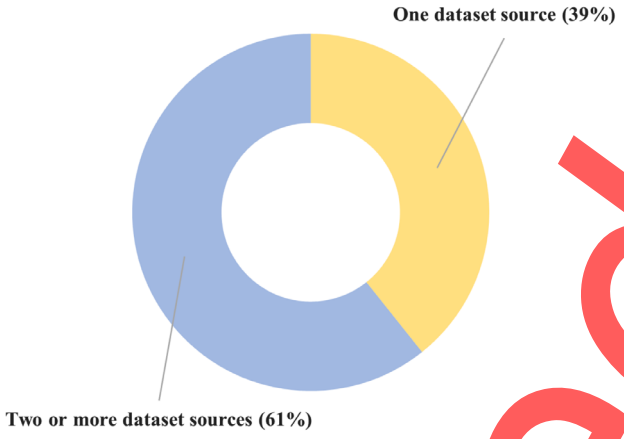

Fig. 6 Number of adopted datasets.

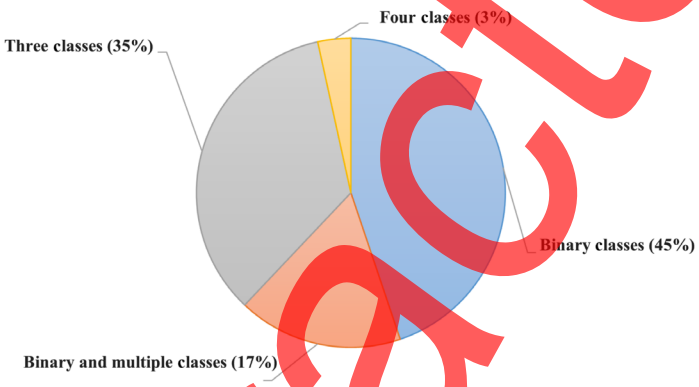

Fig. 7 Number of classes.

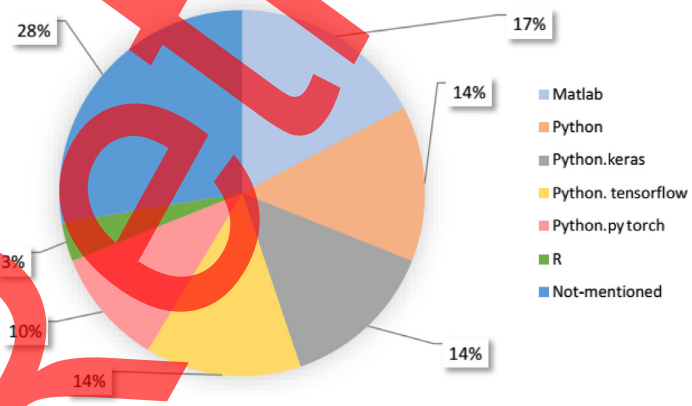

Fig. 8 Simulation tools observation.

for better and more sufficient techniques to visualize which imaging features the model uses to classify images as COVID-19, non-COVID, and so on.

Although AI-based systems, in particular DL methods, are powerful tools for diagnosis and analyzing COVID-19, there are still some other limitations in addition to the above-mentioned cases that need to be considered for future research.

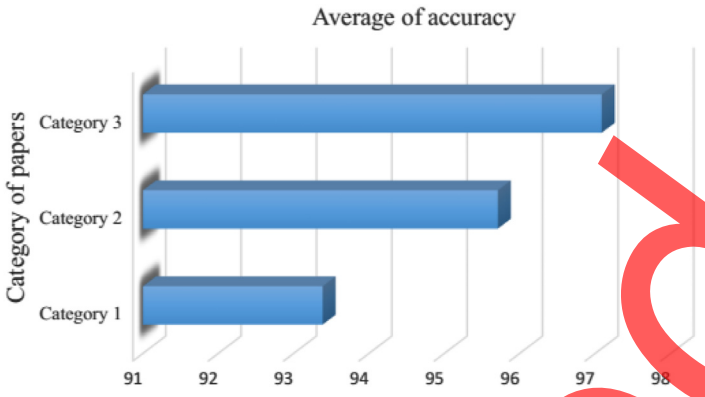

Fig. 9 Average accuracy of the three categories.

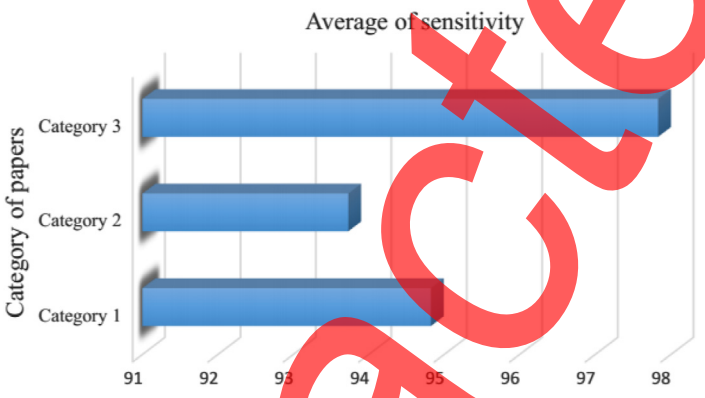

Fig. 10 Average sensitivity of the three categories.

Table 10 Related surveys for detection of COVID-19 using ML methods.

Work	Negative aspect of the work	Paper selection process	Experimental tools	Category	Future research
20	No	Clear	Yes	No	Presented
19	No	Not-clear	No	Yes	Presented
6	No	Not-clear	No	Yes	Presented
21	No	Clear	No	Yes	Not presented
22	No	Not-clear	No	No	Presented
23	No	Clear	No	Yes	Presented
24	No	Clear	No	No	Presented
Our study	Yes	Clear	Yes	Yes	Presented

5 Conclusion

The COVID-19 virus is a newfound disease that exerted immense influence on the world in the shortest possible time. Early detection is necessary for controlling this virus. AI-based systems are known as powerful tools for the fast diagnosis of this disease. We provided a comprehensive review of studies that have proposed ML models for the diagnosis of COVID-19. Surveyed papers are categorized into three classifications: federated ML models, ensemble ML models, and other ML and DL models. The dataset, the number of classes, modality used in the systems,

positive and negative aspects of each study, and their performance metrics are presented and discussed in this paper. X-ray images were the favorite of most researchers for their experiments, especially for ensemble learning models. Also, most of them used multiple data sources. The binary classification was performed in most papers. The preferred experimental tool was Python in almost all papers with Keras and Tensorflow library. Among three categories, the ensemble learning-based models outperformed the two other categories. Although the lack of a large dataset was a big challenge in all surveyed papers, and there is no high certainty for our assertion, the ensemble ML models can be applied as a robust tool for the timely detection of coronavirus disease. Finally, we provided future trends to give more insight to future researchers. We hope to build more accurate and powerful networks with simpler complexity and apply the methods to other infectious diseases such as monkeypox, SARS, and Ebola.

Acknowledgments

The authors did not receive support from any organization for the submitted work. The authors declare that they have no conflicts of interest.

References

1. S. Fouladi et al., "Efficient deep neural networks for classification of COVID-19 based on CT images: virtualization via software defined radio," *Comput. Commun.* **176**, 234–248 (2021).
2. F. Shi et al., "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.* **14**, 4–15 (2020).
3. U. Özkaya et al., "Classification of COVID-19 in chest CT images using convolutional support vector machines," arXiv:2011.05746 (2020).
4. S. H. Khan et al., "Classification and region analysis of COVID-19 infection using lung CT images and deep convolutional neural networks," arXiv:2009.08864 (2020).
5. A. Pannu, "Artificial intelligence and its application in different areas," *Artif. Intell.* **4**(10), 79–84 (2015).
6. T. Alafif et al., "Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions," *Int. J. Environ. Res. Public Health* **18**(3), 1117 (2021).
7. M. Rostami et al., "Gene selection for microarray data classification via multi-objective graph theoretic-based method," *Artif. Intell. Med.* **123**, 102228 (2022).
8. S. Azadifar et al., "Graph-based relevancy-redundancy gene selection method for cancer diagnosis," *Comput. Biol. Med.* **147**, 105766 (2022).
9. M. Rostami et al., "Review of swarm intelligence-based feature selection methods," *Eng. Appl. Artif. Intell.* **100**, 104210 (2021).
10. F. Saberi-Movahed et al., "Decoding clinical biomarker space of COVID-19: exploring matrix factorization-based feature selection methods," *Comput. Biol. Med.* **146**, 105426 (2022).
11. M. Rostami et al., "Integration of multi-objective PSO based feature selection and node centrality for medical datasets," *Genomics* **112**(6), 4370–4384 (2020).
12. A. Ulhaq et al., "COVID-19 control by computer vision approaches: a survey," *IEEE Access* **8**, 179437–179456 (2020).
13. R. P. Singh et al., "Internet of things (IoT) applications to fight against COVID-19 pandemic," *Diabetes Metab. Syndrome: Clin. Res. Rev.* **14**(4), 521–524 (2020).
14. S. Swayamsiddha and C. Mohanty, "Application of cognitive internet of medical things for COVID-19 pandemic," *Diabetes Metab. Syndrome: Clin. Res. Rev.* **14**(5), 911–915 (2020).
15. P. Shah and C. R. Patel, "Prevention is better than cure: an application of big data and geospatial technology in mitigating pandemic," *Trans. Indian Natl. Acad. Eng.* **5**, 187–192 (2020).

16. A. Haleem et al., "Significant applications of big data in COVID-19 pandemic," *Indian J. Orthop.* **54**, 526–528 (2020).
17. K. Iyengar et al., "COVID-19 and applications of smartphone technology in the current pandemic," *Diabetes Metab. Syndrome: Clin. Res. Rev.* **14**(5), 733–737 (2020).
18. S. Banskota, M. Healy, and E. M. Goldberg, "15 smartphone apps for older adults to use while in isolation during the COVID-19 pandemic," *Western J. Emerg. Med.* **21**(3), 514 (2020).
19. M. M. Islam et al., "A review on deep learning techniques for the diagnosis of novel coronavirus (COVID-19)," *IEEE Access* **9**, 30551–30572 (2021).
20. A. Shoeibi et al., "Automated detection and forecasting of COVID-19 using deep learning techniques: a review," https://arxiv.org/abs/2007.10785 (2020).
21. I. Ozsahin et al., "Review on diagnosis of COVID-19 from chest CT images using artificial intelligence," *Comput. Math. Methods Med.* **2020**, 9756518 (2020).
22. M. Ilyas, H. Rehman, and A. Naït-Ali, "Detection of COVID-19 from chest x-ray images using artificial intelligence: an early review," arXiv:2004.05436 (2020).
23. H. S. Alghamdi et al., "Deep learning approaches for detecting COVID-19 from chest x-ray images: a survey," *IEEE Access* **9**, 20235–20254 (2021).
24. M. Ghaderzadeh and F. Asadi, "Deep learning in the detection and diagnosis of COVID-19 using radiology modalities: a systematic review," *J. Healthcare Eng.* **2021**, 6677314 (2021).
25. N. B. Yahia, M. D. Kandara, and N. B. B. Saoud, "Deep ensemble learning method to forecast COVID-19 outbreak," ResearchSquare, Durham, North Carolina (2020).
26. S.-H. Wang et al., "COVID-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network," *Inf. Fusion* **67**, 208–229 (2021).
27. K. Phil, *Matlab Deep Learning with Machine Learning, Neural Networks and Artificial Intelligence*, Vol. **10**, Apress, New York (2017).
28. S. Fouladi et al., "Efficient deep neural networks for classification of alzheimer's disease and mild cognitive impairment from scalp EEG recordings," *Cognit. Comput.* **14**, 1247–1268 (2022).
29. K. Nazari, M. J. Ebadi, and K. Berahmand, "Diagnosis of alternaria disease and leafminer pest on tomato leaves using image processing techniques," *J. Sci. Food Agric.* (2022).
30. M. J. Ebadi and S. Shiri Shahraki, "Determination of scale elasticity in the existence of non-discretionary factors in performance analysis," *Knowl.-Based Syst.* **23**(5), 434–439 (2010).
31. M. J. Ebadi, A. Hosseini, and M. M. Hosseini, "A projection type steepest descent neural network for solving a class of nonsmooth optimization problems," *Neurocomputing* **235**, 164–181 (2017).
32. N. Jamali et al., "Estimating the depth of anesthesia during the induction by a novel adaptive neuro-fuzzy inference system: a case study," *Neural Process. Lett.* **53**, 131–175 (2021).
33. F. Heydarpoor et al., "Solving multi-objective functions for cancer treatment by using metaheuristic algorithms," *Int. J. Combinatorial Optim. Prob. Inf.* **11**(3), 61–75 (2020).
34. F. Heydarpour et al., "Solving an optimal control problem of cancer treatment by artificial neural networks," *Int. J. Interact. Multimedia Artif. Intell.* **6**(4), 18–25 (2020).
35. M. J. Ebadi, A. Hosseini, and H. Jafari, "An efficient one-layer recurrent neural network for solving a class of nonsmooth optimization problems," *J. New Res. Math.* **6**(24), 97–110 (2020).
36. M. J. Ebadi, M. M. Hosseini, and S. M. Karbassi, "An efficient one-layer recurrent neural network for solving a class of nonsmooth pseudoconvex optimization problems," *J. Theor. Appl. Inf. Technol.* **96**(7), 1999–2014 (2018).
37. M. J. Ebadi, H. Farahani, and H. Jafari, "A neurodynamic model for solving nonsmooth constrained optimization problems with affine and bound constraints," *Fuzzy Syst. Appl.* **4**(2), 69–112 (2022).
38. B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in *Int. Conf. Artif. Intell. and Stat.*, PMLR, pp. 1273–1282 (2017).

39. I. Feki et al., "Federated learning for COVID-19 screening from chest x-ray images," *Appl. Soft Comput.* **106**, 107330 (2021).
40. D. Yang et al., "Federated semi-supervised learning for COVID region segmentation in chest ct using multi-national data from China, Italy, Japan," *Med. Image Anal.* **70**, 101992 (2021).
41. P. P. Kulkarni, H. Kasyap, and S. Tripathy, "Dnet: an efficient privacy-preserving distributed learning framework for healthcare systems," *Lect. Notes Comput. Sci.* **12582**, 145–159 (2021).
42. B. N. Wiysobunri, H. S. Erden, and B. U. Toreyin, "An ensemble deep learning system for the automatic detection of COVID-19 in x-ray images," Academic Press (2020).
43. Z. Qiao et al., "Flannel (focal loss based neural network ensemble) for COVID-19 detection," *J. Am. Med. Inf. Assoc.* **28**(3), 444–452 (2021).
44. T. Zhou et al., "The ensemble deep learning model for novel COVID-19 on CT images," *Appl. Soft Comput.* **98**, 106885 (2021).
45. D. Singh et al., "Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks," *Eur. J. Clin. Microbiol. Infectious Diseases* **39**(7), 1379–1389 (2020).
46. D. M. Powers, "Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation," arXiv:2010.16061 (2020).
47. A. Jaiswal et al., "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning," *J. Biomol. Struct. Dyn.* **39**(15), 5682–5689 (2021).
48. O. Attallah, D. A. Ragab, and M. Sharkas, "Multi-deep: a novel cad system for coronavirus (COVID-19) diagnosis from CT images using multiple convolution neural networks," *PeerJ* **8**, e10086 (2020).
49. V. Shah et al., "Diagnosis of COVID-19 using CT scan images and deep learning techniques," *Emerg. Radiol.* **28**(3), 497–505 (2021).
50. C. Ouchicha, O. Ammor, and M. Meknassi, "CVDNet: a novel deep learning architecture for detection of coronavirus (COVID-19) from chest x-ray images," *Chaos, Solitons Fractals* **140**, 110245 (2020).
51. E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in x-ray images," arXiv:2003.11055 (2020).
52. A. R. Martinez, "Classification of COVID-19 in CT scans using multi-source transfer learning," arXiv:2009.10474 (2020).
53. L. Li et al., "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy," *Radiology* **296**(2), E65–E71 (2020).
54. A. Amyar et al., "Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation," *Comput. Biol. Med.* **126**, 104037 (2020).
55. J. Li et al., "Multi-task contrastive learning for automatic CT and x-ray diagnosis of COVID-19," *Pattern Recognit.* **114**, 107848 (2021).
56. A. Qayyum et al., "Collaborative federated learning for healthcare: multi-modal COVID-19 diagnosis at the edge," https://arxiv.org/abs/2101.07511 (2021).
57. Q. Yang et al., "Flop: federated learning on medical datasets using partial networks," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, pp. 3845–3853 (2021).
58. R. Kumar et al., "Blockchain-federated-learning and deep learning models for COVID-19 detection using ct imaging," *IEEE Sens. J.* **21**(14), 16301–16314 (2021).
59. B. Yan et al., "Experiments of federated learning for COVID-19 chest x-ray images," in *Int. Conf. Artif. Intell. and Secur.*, Springer, pp. 41–53 (2021).
60. W. Zhang et al., "Dynamic-fusion-based federated learning for COVID-19 detection," *IEEE Internet Things J.* **8**(21), 15884–15891 (2021).
61. Y. Xu et al., "A collaborative online ai engine for CT-based COVID-19 diagnosis," medRxiv (2020).
62. A. K. Das et al., "Automatic COVID-19 detection from x-ray images using ensemble learning with convolutional neural network," *Pattern Anal. Appl.* **24**(3), 1111–1124 (2021).
63. S. Rajaraman et al., "Iteratively pruned deep learning ensembles for COVID-19 detection in chest x-rays," *IEEE Access* **8**, 115041–115050 (2020).

64. N. Gianchandani et al., "Rapid COVID-19 diagnosis using ensemble deep transfer learning models from chest radiographic images," *J. Ambient Intell. Hum. Comput.* **4**, 1–13 (2020).
65. K. Upadhyay, M. Agrawal, and D. Deepak, "Ensemble learning-based COVID-19 detection by feature boosting in chest x-ray images," *IET Image Process.* **14**(16), 4059–4066 (2020).
66. S. Tang et al., "EDL-COVID: ensemble deep learning for COVID-19 case detection from chest x-ray images," *IEEE Trans. Ind. Inf.* **17**(9), 6539–6549 (2021).
67. P. Saha, M. S. Sadi, and M. M. Islam, "EMCNet: automated COVID-19 diagnosis from x-ray images using convolutional neural network and ensemble of machine learning classifiers," *Inf. Med. Unlocked* **22**, 100505 (2021).
68. C. Öksüz, O. Urhan, and M. K. Güllü, "Ensemble-CVDNet: a deep learning based end-toend classification framework for COVID-19 detection using ensembles of networks," arXiv:2012.09132 (2020).
69. N. K. Chowdhury et al., "ECOVNet: an ensemble of deep convolutional neural networks based on Efficientnet to detect COVID-19 from chest x-rays," arXiv:2009.11850 (2020).
70. E. Soares et al., "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification," medRxiv, www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset (2020).
71. J. Ma et al., "Towards efficient COVID-19 CT annotation: a benchmark for lung and infection segmentation," arXiv:2004.12537 (2020).
72. J. Zhao et al., "COVID-CT-dataset: a CT scan dataset about COVID-19," arXiv:2003.13865 490 (2020).
73. M. E. H. Chowdhury et al., "Can AI help in screening viral and COVID-19 pneumonia?," *IEEE Access* **8**, 132665–132676 (2020).
74. J. P. Cohen et al., "COVID-19 image data collection: prospective predictions are the future," arXiv:2006.11988 (2020).
75. https://www.pyimagesearch.com/category/medical/.
76. J. Deng et al., "Imagenet: a large-scale hierarchical image database," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 248–255 (2009).
77. A. Sorokin and D. Forsyth, "Utility data annotation with Amazon mechanical turk," in *IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit. Workshops*, IEEE, pp. 1–8 (2008).
78. "SPIE-AAPM lung CT challenge - The Cancer Imaging Archive (TCIA) public access - cancer imaging archive Wiki," Cancer Imaging Archive (2014).
79. J. Ma et al., "COVID-19 CT lung and infection segmentation dataset," http://medicalsegmentation.com/covid19/.
80. J. P. Cohen et al., "COVID-19 image data collection: prospective predictions are the future," arXiv:2006.11988 (2020).
81. J. Born et al., "Pocovid-net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (pocus)," arXiv:2004.12084 (2020).
82. S. Jaeger et al., "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantum Imaging Med. Surg.* **4**(6), 475 (2014).
83. R. Kumar et al., "Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging," *IEEE Sens. J.*, https://github.com/abdkhanstd/COVID-19.
84. M. Rahimzadeh, A. Attar, and S. M. Sakhaei, "A fully automated deep learning-based network for detecting COVID-19 from a new and large lung ct scan dataset," *Biomed. Signal Process. Control* **68**, 102588 (2021).
85. J. Zhao et al., "COVID-CT-Dataset: a CT scan dataset about COVID-19," arXiv:2003.13865, https://github.com/UCSD-AI4H/COVID-CT.
86. A. Chung et al., "Figure1-COVID-chestxray-dataset," https://github.com/agchung/Figure1-COVID-chestxray-dataset.
87. "A thread of COVID-19 CXR (all SARS-CoV-2 PCR+)," Spain, https://twitter.com/ChestImaging/status/1243928581983670272.
88. "COVID-19 DATASET," Società Italiana di Radiologia Medica e Interventistica, Italy, https://www.sirm.org/category/senza-categoria/covid-19/.

89. J. Irvin et al., "Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison," *Proc. AAAI Conf. Artif. Intell.* **33**(01), 590–597 (2019).
90. D. Kermany et al., "Labeled optical coherence tomography and chest X-ray images for classification," https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.
91. W. Kong and P. P. Agarwal, "Chest imaging appearance of COVID-19 infection," *Radiol.: Cardiothorac. Imaging* **2**(1), e200028 (2020).
92. D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell* **172**(5), 1122–1131.e9 (2018).
93. G. Shih et al., "Augmenting the National Institutes of Health Chest Radiograph dataset with expert annotations of possible pneumonia," *Radiol. Artif. Intell.* **1**(1), 1–5 (2019).
94. J. P. Cohen, "COVID-19 image data collection," https://arxiv.org/abs/2003.11597.
95. A. Asraf, "COVID19_Pneumonia_Normal_Chest_Xray(PA)_Dataset," 2020, https://www.kaggle.com/amanullahasraf/covid19-pneumonia-normal-chest-xraypa-dataset (accessed 29 June 2020).
96. M. E. Chowdhury et al., "Can ai help in screening viral and COVID-19 pneumonia?" *IEEE Access* **8**, 132665–132676 (2020).
97. D. Deshpande, "COVID-19 detection X-ray dataset," https://www.kaggle.com/datasets/darshan1504/covid19-detection-xray-dataset.
98. A. Narin, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," arXiv:2003.10849, https://github.com/drcerenkaya/COVID-19-Detection (accessed 2020-10-01).
99. L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.* **10**(1), 19549 (2020).
100. A. G. Chung, "Figure1-COVID-chestxray-dataset, and Actualmed-COVID-chestxray-dataset," https://github.com/agchung.
101. The Cancer Imaging Archive, https://www.cancerimagingarchive.net/.
102. "COVID-19 dataset," Società Italiana di Radiologia Medica e Interventistica, Italy, https://www.sirm.org/category/senza-categoria/covid-19/.
103. A. M. Alqudah and S. Qazan, "Augmented COVID-19 x-ray images dataset," Mendeley Data, 4 (2020).
104. http://radiopaedia.org/.
105. X. Wang et al., "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," IEEE CVPR, https://www.kaggle.com/nih-chest-xrays/data (2017).
106. R. Karthik, R. Menaka, and M. Hariharan, "Learning distinctive filters for COVID-19 detection from chest x-ray using shuffled residual cnn," *Appl. Soft Comput.* **99**, 106744 (2021).

Biographies of the authors are not available.