SYSTEMATIC REVIEW

Signa Vitae

Artificial intelligence for the triage of COVID-19 patients at the emergency department: a systematic review

Pablo Redruello-Guerrero[1], Carmen Jiménez-Gutiérrez[2],
Antonio Jesús Láinez Ramos-Bossini[3,]*, Paula María Jiménez-Gutiérrez[4],
Mario Rivera-Izquierdo[5,6], José Manuel Benítez Sánchez[7]

[1] Faculty of Medicine, University of Granada, 18016 Granada, Spain
[2] Department of Nursing, University of Granada, 18016 Granada, Spain
[3] Department of Radiology, Virgen de las Nieves University Hospital, 18016 Granada, Spain
[4] Service of Anesthesiology, Resuscitation and Pain Therapeutics, Hospital Universitario Virgen de las Nieves, 18014 Granada, Spain
[5] Service of Preventive Medicine and Public Health, Hospital Universitario Clínico San Cecilio, 18016 Granada, Spain
[6] Department of Preventive Medicine and Public Health, University of Granada, 18016 Granada, Spain
[7] Department of Computer Science and Artificial Intelligence, E.T.S. Ingeniería Informática, University of Granada, 18071 Granada, Spain

*Correspondence
ajbossini@ugr.es
(Antonio Jesús Láinez Ramos-Bossini)

Abstract

The aim of this article is to systematically analyze the available literature on the efficacy and validity of artificial intelligence (AI) applied to medical imaging techniques in the triage of patients with suspected or confirmed coronavirus disease 2019 (COVID-19) in Emergency Departments (EDs). A systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines was conducted. Medline, Web of Science, and Scopus were searched to identify observational studies evaluating the efficacy of AI methods in the diagnosis and prognosis of COVID-19 using medical imaging. The main characteristics of the selected studies were extracted by two independent researchers and were formally assessed in terms of methodological quality using the Newcastle-Ottawa scale. A total of 11 studies, including 14,499 patients, met inclusion criteria. The quality of the studies was medium to high. Overall, the diagnostic yield of the AI techniques compared to a gold standard was high, with sensitivity and specificity values ranging from 79% to 98% and from 70% to 93%, respectively. The methodological approaches and imaging datasets were highly heterogeneous among studies. In conclusion, AI methods significantly boost the diagnostic yield of medical imaging in the triage of COVID-19 patients in the ED. However, there are significant limitations that should be overcome in future studies, particularly regarding the heterogeneity and limited amount of available data to train AI models.

Keywords

COVID-19; Diagnostic imaging; Radiology; Artificial intelligence; Machine learning; Sensitivity; Specificity; Validity; Emergency

1. Introduction

To date, more than 450 million cases and 6 million deaths have been confirmed during the COVID-19 pandemic worldwide [1]. The high contagiousness, multiple reservoirs, and insufficient preventive resources, among other factors, led to confinement regulations in most countries during the first wave of the pandemic [2–4]. From a clinical perspective, the presenting symptoms of the disease are widely varied, including respiratory symptoms, fatigue [5] or neurological symptoms [6, 7], and more uncommon signs like pneumoperitoneum [8, 9], among an endless list of potential manifestations. In many cases, these symptoms overlap with other clinical entities, particularly when respiratory symptoms predominate. Such clinical variability and non-specificity of symptoms make the clinical diagnosis of COVID-19 fairly complicated at the Emergency Department (ED).

On the other hand, early diagnosis is associated with better outcomes. Patients suffering from COVID-19 showed high mortality and hospitalization rates since the beginning of the pandemic [10]. Several risk factors have been reported as predictors of poor outcomes, such as the elderly, prognostic scales [10, 11], different treatment strategies [12–14], or even lifestyle factors [15], but inconsistencies in defining risk profiles of patients complicate decision-making in ED triage [16, 17]. Apart from in-hospital prognosis, Long-COVID perpetuates the symptomatology and generates medium- and long-term sequelae [18, 19], which in turn increase the rate of readmission of patients to EDs, especially those living in long-term care facilities [20]. It should be noted that the pandemic situation has significantly increased stress, anxiety, and depressive symptoms in healthcare workers [21]. Particularly, the EDs were overwhelmed by excessive demand, which saturated EDs since the beginning of the pandemic [10, 20, 22–24]. Considering the importance of early and accurate diagnosis in COVID-19 patient outcomes, fast and accurate identification of cases might have an invaluable impact in healthcare.

To provide solutions to some of the problems detected in clinical practice during the pandemic, several tools have been developed to assess and improve the triaging systems in EDs, increase diagnostic accuracy, early identification of cases and support healthcare workers. Artificial Intelligence (AI)-based techniques, especially deep learning (DL), have demonstrated great utility when combined with other parameters such as laboratory workup tests [25]. Specifically, DL applied to diagnostic imaging has been reported as an efficient alternative diagnostic tool in different medical specialties that take advantage of medical images, such as histological analysis [26] or radiology [27], and in a wide spectrum of diseases, including COVID-19 [28]. In fact, DL techniques have been applied to improve COVID-19 detection on chest radiographs (CXR) [29–31] and computed tomography (CT) [32]. However, the training and implementation of AI models require solving frequent limitations encountered in this setting, including but not limited to unbalanced datasets, non-labeled or unrepresentative images, inter-institutional heterogeneity in imaging equipment and quality of CT and CXR images, and differences in the ground truth used to train and validate AI results.

Some of the most outstanding approaches in this context include the use of fused deep features based classification framework with optimized multi-layer perceptron structures [33] or shrunken features [34]. In addition, data augmentation and transfer learning techniques allow solving common problems encountered in this setting, as is the case with unbalanced dataset images. Despite there is wide variety of AI algorithms, the results are difficult to compare due to high heterogeneity in terms of methodological approaches and study designs [35]. Accordingly, it is necessary to determine the effectiveness and limitations of different AI techniques available in specific clinical scenarios. As mentioned above, one context of special scientific interest is the triage of patients in the ED, specifically concerning imaging techniques, which are performed routinely in this setting and provide data that can be exploited by AI systems. Therefore, this systematic review aims to analyze the efficacy of AI techniques applied to imaging techniques in the triage of patients with suspected COVID-19.

2. Methods

2.1 Study design and search strategy

A systematic review was conducted following international standards. The report of the results followed the recommendations of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement [36]. The protocol was prospectively registered in PROSPERO (code: CRD42021240732) [37]. The development of this study included professionals from medical imaging (radiologists), experts in AI, physicians with experience in the management of critically ill patients, and research methodologists.

Medline, Web of Science Core Collection, and Scopus were used to collect relevant information from inception to 30 January 2022. The following Medical Subject Headings (MeSH) terms were combined in each database as appropriate: "SARS-CoV-2", "COVID 19", "Coronavirus", "pandemic", "emergency department", "emergency room", "artificial intelligence", "machine learning", "deep learning", "triage", "early screening", "diagnosis", "prognosis", "mortality", and "severity". The search was conducted by two researchers independently.

2.2 Selection of studies

Since the aim of this systematic review is to analyze studies exploring AI techniques in the triage of patients at the ED, inclusion criteria covered observational original research (both cross-sectional and longitudinal) studies published in peer-reviewed scientific journals. Language was restricted to English or Spanish. We included all algorithms developed through AI applied to medical imaging in hospital EDs. Medical imaging examinations included chest radiography (anteroposterior and posteroanterior), chest CT and lung ultrasound. Different study designs that were non-relevant to our purpose were excluded from analyses (*e.g.*, case studies, letters to the editor, reviews, editorials, or commentaries). Studies conducted in samples obtained in hospital services different from the ED were also excluded.

The studies that reported data on diagnostic accuracy, severity, or prognosis of COVID-19 comparing AI with ordinary triage for radiological imaging were identified. First, titles and abstracts were read and selected by three independent reviewers. Second, eligible criteria were assessed, and full-text reading was conducted for the selected documents. Inconsistencies or disagreements were solved by consensus with a third experienced researcher.
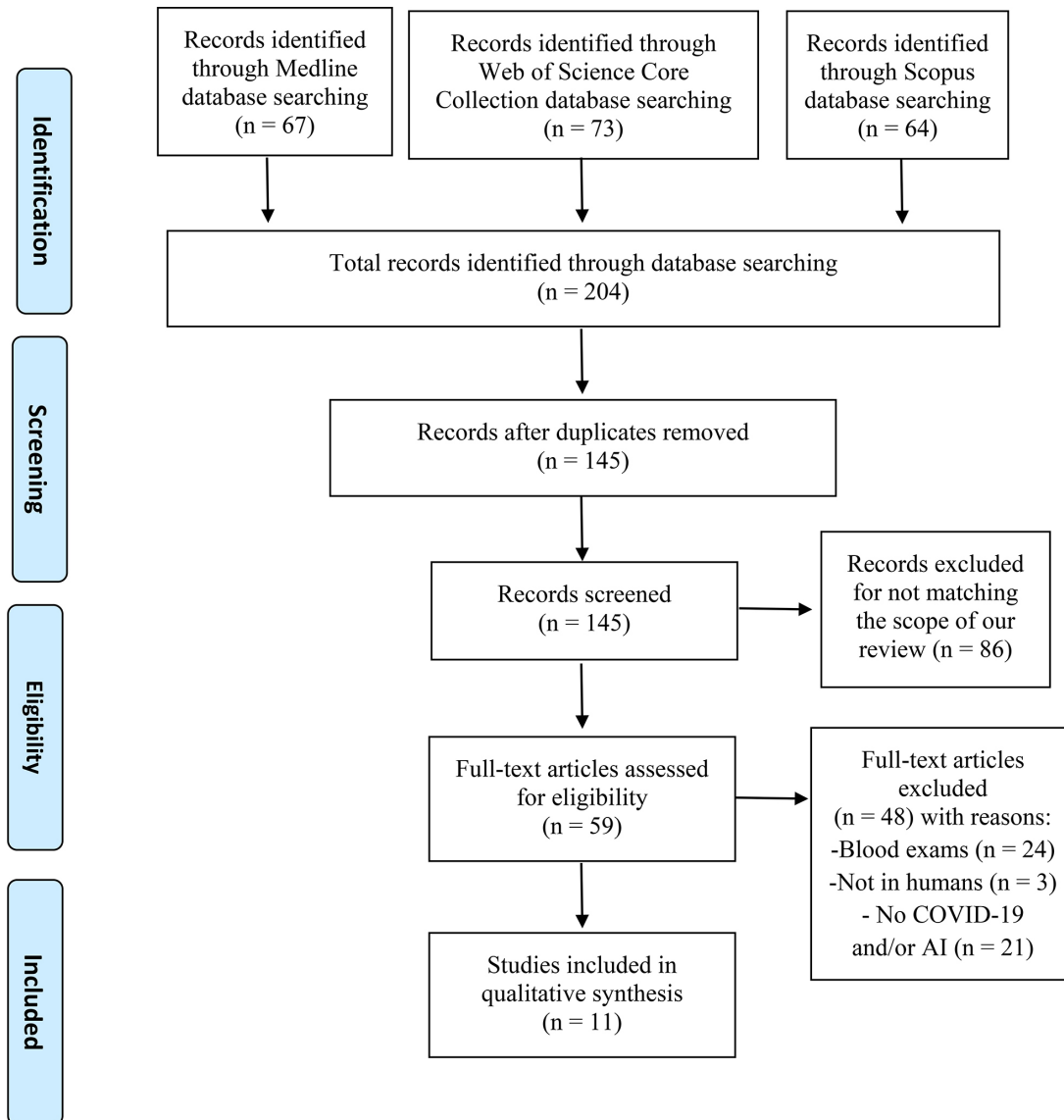
2.3 Quality assessment and data extraction

Data were collected using specific datasheets. The relevant information included authors, year of publication, study design, country, sample size, AI methodology, type of imaging modality, sensitivity, specificity, and the presence of conflicts of interest.

Quality assessment was performed using the nine-star Newcastle-Ottawa scale (NOS) [38] by two independent researchers. The risk of bias was assessed concerning selection, comparability, and outcomes. High quality (low risk of bias) was considered for studies scoring 8 or 9 points; medium quality (medium risk of bias) for studies scoring 6 or 7 points; and low quality (high risk of bias) for studies scoring 5 points or less. Discrepancies were solved after consensus with a third experienced researcher.

3. Results

Fig. 1 summarizes the selection process of the search. A total of 204 studies were screened. After removing duplicates (n = 59) and studies with a design or scope different from inclusion criteria (n = 86), a full-text assessment was conducted. A total of 24 studies were excluded for not using AI at triage for COVID-19 patients, and 3 studies were excluded for using radiological databases instead of diagnostic imaging at the ED. Finally, 21 studies did not meet inclusion criteria since they reported outcomes on AI techniques applied to conditions other than COVID-19. Therefore, 11 studies were included in this systematic review [39–49].

Records identified
through Medline
database searching
(n = 67)

Records identified through
Web of Science Core
Collection database searching
(n = 73)

Records identified
through Scopus
database searching
(n = 64)

Identification

Total records identified through database searching
(n = 204)

Screening

Records after duplicates removed
(n = 145)

Records screened
(n = 145)

Records excluded
for not matching
the scope of our
review (n = 86)

Eligibility

Full-text articles assessed
for eligibility
(n = 59)

Full-text articles
excluded
(n = 48) with reasons:
-Blood exams (n = 24)
-Not in humans (n = 3)
- No COVID-19
and/or AI (n = 21)

Included

Studies included in
qualitative synthesis
(n = 11)

F I G U R E 1. Flow diagram of the selection process for the studies included in this systematic review, according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [36]. COVID-19, coronavirus disease 2019; AI, artificial intelligence.

All selected documents were longitudinal cohort studies, 7 of them prospective [39–41, 44, 46, 48, 49] and 4 retrospective [42, 43, 45, 47]. Only one study was conducted in a multicenter cohort [47].

3.1 Quality assessment

The quality of the studies according to the 9-point NOS is illustrated in Fig. 2. Overall, none of the selected studies showed a high risk of bias. Only three studies were classified as medium quality (6–7 points) [42, 45, 47]. The remaining eight studies (72.7%) were judged to be of low risk of bias (8–9 points). The mean quality score of all the studies we included was 7.45. Overall, patient selection and comparability between assessment groups were the assessment points with the lowest reported scores. All studies classified as medium quality showed potential conflicts of interest.

3.2 Characteristics of the included studies

The 11 included studies gathered a total of 14,999 participants, 45.5% of them in the Americas (including North and South America), and the rest (55.5%) in Europe. The most frequently used diagnostic imaging techniques were chest radiography (63.6%), chest CT (18.2%), and lung ultrasound (9.1%). The gold standard for comparison of all studies was experienced radiologists who agreed between two [42, 43, 46, 48] or more [45, 47]. Some severity scoring scales already described for COVID-19 were employed [40, 41, 44]. Comorbidities were described in only 54.5% of the included patients [39, 40, 42, 43, 46, 47]. Chronic obstructive pulmonary disease was present in four of the studies [40, 42, 43, 46]. The definitive diagnostic test was positivity for reverse transcription polymerase chain reaction (RT-PCR) analysis which is considered to be an established diagnostic criterion for COVID-19 [10]. According to data provided by international morbidity and mortality registry bodies [1] the case fatality rates of COVID-19 infection were
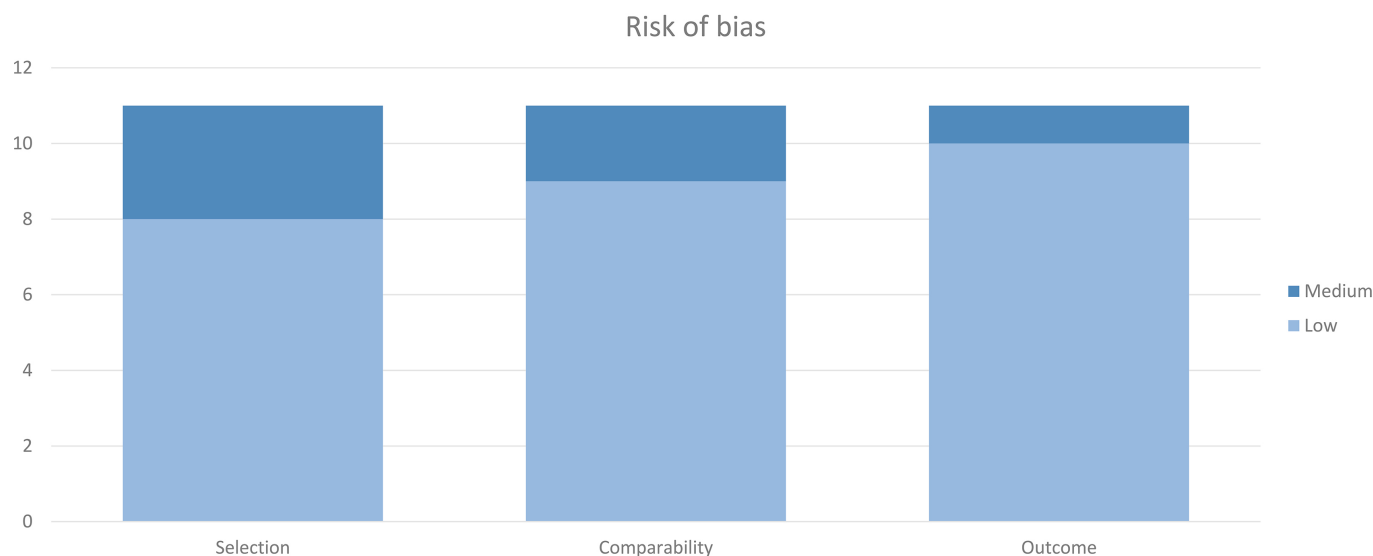
Risk of bias

F I G U R E 2. Quality of the studies included in this review according to the Newcastle-Ottawa scale [38]. For each individual study, the risk of bias was classified into low, medium, or high risk for selection, comparability and outcome.

considerable. They ranged from 8.2% [40] to 26% [47], although the approximate median rate was around 18% in the studies included in this systematic review.

All included studies analyzed the receiver operating characteristic (ROC) curve for the training, validation, and test models in the ED. The only study that did not analyze the ROC area under the curve (AUC) was the work by Li *et al*. [45], in which the analysis was performed using Pearson's correlation between the data obtained by radiologists of each participating hospital and the data obtained through the AI models applied to chest radiographs.

The AUC values of the different models showed large variations, with training model AUC values of 0.97–0.98 as in the Random Forest model reported by Garrafa *et al*. [41] to the mortality prediction AUC value of 0.66 obtained in Qure.ai [46]. However, the models with the highest performance were proposed by La Salvia *et al*. [44] as ResNet18 with a training curve of 99.70 ± 0.01, a validation curve of 99.78 ± 0.20, and a test curve around 97.72 ± 0.63; and the ResNet50 model, with a training curve close to 99.95 ± 0.01, a validation curve of 99.81 ± 0.18 and a test curve close to 99.91 ± 0.07. Butler *et a l*. [40] stratified the AUC according to the severity of patients assessed by CheXNet-Cov19. In those patients admitted to an intensive care unit (ICU) the AUC was lower than in the other strata (0.67) possibly due to the supine position of the CXR. In contrast to what has been described by other authors [46] where their proposed model has a higher AUC in critically ill patients compared to the control group. Therefore, the efficacy in the diagnosis of critically ill patients may be higher with the use of Qure.ai than with CheXNet-Cov19. The combination of clinical parameters optimised the model proposed by Jiao *et al*. [42], where the AUC went from a validation range of 0.753 to 0.792. This was also described in the Swedish study [49] where the combination of CT imaging parameters together with sociodemographic and laboratory variables increased the AUC to 0.91.

As additional results, data on mortality, sensitivity, and

specificity were also analyzed. In general, mortality rates ranged from 8–10% [40, 42] to approximately 20% [41, 46, 48], although Schiaffino *et al*. [47] reported a higher mortality percentage (26%). Regarding the validity parameters of diagnostic tests, the following results were found: the reported sensitivity of the Qure.ai system [46] was high (76.6%), although it was lower than the diagnostic sensitivity of the control by two radiologists using the Radiographic Assessment of Lung Edema (RALE) score (86.6.%); the DL model ResNet18 [44] obtained a sensitivity in training, validation, and test of 96.1%, 97.0%, 97.4%, respectively; the DL model ResNet50 [44] showed a sensitivity in training close to 97.2%, in validation around 97.5% and in test 98.2%; the DL model proposed by Jiao *et al*. [42] showed a specificity of 85.3% in training and 70.1% in test validation, and a sensitivity of 73.8% and 72.8%, respectively; another model developed by Kwon *et al*. [43] showed a sensitivity of 82% in predicting patients requiring mechanical ventilation and 78% for patient mortality; the specificity in training of the model used by Weikert *et al*. [49] was 81.8% and 6 points lower in model validation, with similar values for sensitivity (80.0% and 74.2%, respectively); the BS-EWM machine learning model [41] showed specificity and sensitivity values of 92% and 93% in training, 75% and 82% in validation, and 73% in test; finally, the CheXNet-Cov19 model [40] reported a specificity close to 76% and a sensitivity close to 70%. It is remarkable that eight of the eleven studies included have received funding, two do not indicate whether or not they have received funding, and only one reports not having received funding. This highlights the challenge of this situation that has mobilised the global economy for efficient diagnosis, treatment and prevention worldwide. Table 1 summarises the main characteristics of all studies included in this review.

T A B L E 1. Main characteristics of the studies included.

Study, Year	Country	Sample Size	Design	Diagnostic imaging	Type of Artificial Intelligence	Main results	Conflicts of Interest	Total NOS Score
Bartolucci *et al.*, 2021	Italy	115	Prospective	Chest CT	Machine learning	ROC AUC for Hybrid radiological model: —Training: 0.87 (0.77–0.97) —Validation: 0.82 (0.73–0.97)	No	8
Butler *et al.*, 2021	USA	3.571	Prospective	CXR	CheXNet based deep learning model, namely CheXNet-Cov19	ROC AUC for CheXNet-Cov19 using CXR only: —COVID-19: 0.71 (0.63–0.80) —ARDS: 0.74 (0.66–0.82) —ICU: 0.67 (0.58–0.75) —Death: 0.76 (0.68–0.84)	No	8
Garrafa *et al.*, 2021	Italy	2.782	Prospective	CXR (Brescia chest X-ray)	Machine learning model (BS-EWM)	ROC AUC for Random Forest: —Training: 0.97 (0.97–0.98) —Validation: 0.83 (0.80–0.87) —Test: 0.78 (0.73–0.84) ROC AUC for GBM: —Training: 0.88 (0.86–0.89) —Validation: 0.84 (0.80–0.88) —Test: 0.78 (0.73–0.83) ROC AUC for Logistic regression: —Training: 0.84 (0.82–0.86) —Validation: 0.83 (0.79–0.87) —Test: 0.52 (0.44–0.60)	No	8
Jiao *et al.*, 2021	USA	1.834	Retrospective	CXR	Deep learning	ROC AUC for image-based model: —Training: 0.803 (0.773–0.817) —Validation: 0.753 (0.746–0.772) ROC AUC for image and clinical data combined model: —Training: 0.846 (0.815–0.852) —Validation: 0.792 (0.780–0.803)	Yes	6
Kwon *et al.*, 2020	USA	499	Retrospective	CXR	Deep learning	ROC AUC for: —CXR severity score: 0.80 (0.73–0.88) —Admission: 0.76 (0.68–0.84) —Intubation: 0.66 (0.56–0.75) —Death: 0.59 (0.49–0.69)	No	8

TABLE 1. Continued.

Study, Year	Country	Sample Size	Design	Diagnostic imaging	Type of Artificial Intelligence	Main results	Conflicts of Interest	Total NOS Score
La Salvia *et al.*, 2021	Italy	450	Prospective	Lung ultrasound	Deep learning (ResNet*18* and ResNet*50* architectures)	ROC AUC ResNet18: —Training: 99.70 ± 0.01 —Validation: 99.78 ± 0.20 —Test: 97.72 ± 0.63 ROC AUC ResNet50: —Training: 99.95 ± 0.01 —Validation: 99.81 ± 0.18 —Test: 99.91 ± 0.07	No	8
Li *et al.*, 2020	USA, Brazil	591	Retrospective	CXR	Convolutional Siamese neural network-based model, PXS Score Base Model	r = 0.89. *p* < 0.001	Yes	6
Mushtaq *et al.*, 2021	Italy	697	Prospective	CXR	Deep learning (qXR v2.1 c2, Qure.ai Technologies)	ROC AUC for mortality: —Qure.ai: 0.66 —RALE: 0.67 ROC AUC COVID-19 for critical care: —Qure.ai: 0.76 —RALE: 0.75	No	8
Schiaffino *et al.*, 2021	Italy	897	Retrospective (multicenter)	Chest CT	Machine Learning	ROC AUC: —Training/Validation: 0.871 —Test: 0.844	Yes	6
Shamout *et al.*, 2021	USA	2.943	Prospective	CXR	Deep convolutional neural network model	ROC AUC: 0.808 (95% CI, 0.746–0.866)	No	8
Weikert *et al.*, 2021	Switzerl	120	Prospective	Chest CT	Deep learning	ROC AUC for CT metrics only: —Training: 0.88 (0.79–0.97) —Validation: 0.75 (0.47–1.00) —Test: Unknown ROC AUC for Pulmonary CT metrics and demographics: —Training: 0.84 (0.75–0.94) —Validation: 0.71 (0.39–1.00) —Test: 0.77 (0.66–0.88) ROC AUC for combined parameters: —Training: 0.91 (0.85–0.98) —Validation: 0.75 (0.48–1.00) —Test: Unknown	Yes	8

AUC, Area Under the Curve; CXR, Chest Radiographies; CT, Computerized Tomography; GBM, Gradient Boosting Machine; r, Pearson correlation coefficient; RALE, Radiographic Assessment of Lung Edema; ROC, Receiver Operating Characteristic; USA, United States of America. NOS, Newcastle-Ottawa Scale; PXS, Pulmonary X-ray Severity; qXR, Quantitative X-ray Score; COVID, Coronavirus Disease 2019; ARDS, Adult Respiratory Distress Syndrome; ICU, Intensive Care Unit.

4. Discussion

In this systematic review gathering observational evidence regarding the utility of AI techniques on diagnostic imaging of COVID-19 in EDs, we included 11 studies with almost 15,000 patients. The ROC AUC value was the most common measurement reported across studies, showing diagnostic and prognostic yield values over 99% in most cases. The validity of the evaluated techniques compared to a gold standard showed a sensitivity ranging from 79% to 98% and a specificity ranging from 70% to 93%, depending on the technique. Overall, our study shows the potential usefulness of AI techniques in terms of both diagnostic accuracy and prognostic evaluation of COVID-19 cases. Nevertheless, the variability in terms of techniques and observed results creates difficulties in selecting which AI model could provide better results if implemented in real-world practice. Therefore, it is worth discussing some details, particularly regarding radiological image processing and the AI models applied.

With regard to radiological image acquisition and processing, three different radiological techniques were used in the studies analyzed in this review, including chest radiographs, chest CT studies, and lung ultrasound images. In addition, various image datasets for each imaging modality were selected, increasing heterogeneity. In general, image pre-processing of chest radiographs was performed by normalizing the pixel values of each image and rescaling the image from the center to fit the AI model used. It should be noted that, in radiological practice, technical acquisition problems may arise due to factors such as inadequate patient positioning, breathing difficulties, or interposition of tissues over the lungs. Conversely to other imaging modalities, in chest radiography, these limitations are a clear constraint for the objective evaluation of the lung parenchyma in the images obtained, which often leads to both false positive and negative conclusions. In other words, the intrinsic sensitivity and specificity of chest radiography are limited, and having an optimal image acquisition is of the essence to appropriately train DL models. To obtain the 'ground truth' for each imaging exam, experienced radiologists scored the chest radiographs using different scales such as RALE [50], modified RALE (mRALE) [45, 50], the Brescia score [41, 51], or methods proposed by other authors such as Toussie *et al.* [52]. These scoring systems are based on the degree of lung opacity (1 point for opaque regions and 0 points for healthy regions) and the distribution of these areas (perihilar, upper, middle or lower lobe), which add up, enabling an overall score for each radiograph. Of note, some studies combined chest radiographs with clinical or laboratory data covering a wide range of symptoms and signs to improve diagnostic efficacy [40–42, 46, 49]. This could overcome some of the limitations previously mentioned, not only because of the use of more sources of information but also because of the objective validity of these sources, particularly in the case of laboratory parameters.

In the case of CT images, a more in-depth analysis of the lung parenchyma and mediastinal nodes by radiologists was possible due to the advantages of this technique compared to CXR (*e.g.*, multiplanar capability, higher spatial resolution). Similar to the previous case, an overall score was given to each imaging exam based on different scales that considered the extension and distribution of lung opacities [53]. Regarding acquisition protocols, CT images were reconstructed with slice thicknesses ranging from 0.6 to 1 mm that allowed a "soft tissue" reconstruction for the subsequent application of tracking algorithms [39, 47, 49]. For the segmentation of lung areas, "Region Growing" algorithms were used, allowing the definition of intervals with air-like (*i.e.*, lung) or slightly lower densities, including thresholds from -700 Hounsfield Units (HU)to -250 UH for ground-glass opacity patterns. Due to the higher resolution of CT equipment compared to conventional radiography, these intervals allow a more objective detection of lung opacities. Interestingly, quantitative analysis of imaging textures using radiomics techniques such as RadAR was also performed in some cases, with interesting results [54].

Lung ultrasound images were obtained using convex probes with frequencies ranging between 5 and 12 MHz, allowing an imaging depth of up to 10 cm to reach the pleural lines and the lung under study. The score assigned to each image is based on a modification of the scores by Mongoni *et al.* [55] and Soldati *et al.* [56] proposed by La Salvia *et al.* [44]. The frames that best represented the lung pattern being assessed were extracted from each ultrasound image and were subsequently used to train AI models. It should be noted that ultrasound is a highly operator-dependent imaging technique, thus image quality is generally limited compared to other imaging modalities such as CT, and the heterogeneity among studies is higher.

Concerning the specific AI models used in the studies analyzed in this systematic review, several methodological aspects need to be highlighted considering their variability. For instance, to develop the Qure.ai model [46], DL systems based on convolutional neural networks were used to define individual detection blocks for abnormalities in the processed images after reviewing the areas of greatest interest such as the cardiac silhouette, costodiaphragmatic angles or hilar areas, and detection of nodules, cavitation, fibrosis or opacities. The COVID-Globally-Aware Multiple Instance Classifier (GMIC) model [48] was designed to assess the clinical deterioration of patients in the days following admission by using machine learning algorithms and some clinical parameters such as oxygen requirement, respiratory rate, oxygen saturation, or body temperature. To do this, an initial anatomical mapping was performed to generate a global network to which a local network was subsequently applied to obtain the details that would allow these zones to be labeled in combination with the clinical data previously attached. Other authors used the Pulmonary X-ray Severity (PXS) scoring system [45] and trained a Siamese neural network model with pairs of chest radiographs. This pairing allowed the calculation of the Euclidean distance between the layers of the radiological images, which was used as a continuous measure of severity similarity. The ResNet18 and ResNet50 [44] architectures are two DL networks composed of 18 and 50 layers, respectively, configured as proposed by He *et al.* [57]. To evaluate the results obtained, tools such as Class Activation Mapping (CAM) or Gradient CAM (Grad-CAM) were used to allow the interpretation of the decision-making by the model. The EfficientNet [42] was designed as a DL network using visual geometry architectures such as Visual Geometry Group (VGG)-11 and U-Net. These enabled area

encoding and allowed the model to learn the transformation of the different inputs. As was the case in the previous model, learning of EfficientNet was enhanced using the Adams algorithm [58]. The Hierarchical Data Format (HDF5) dataset used the DenseNet-121 prediction algorithm that yielded a sigmoid function assigning a probability of severity, which was subsequently refined using the binary cross-entropy loss function and the Adam optimizer.

On the other hand, Schiaffimo *et al.* [47] developed machine learning models based on support vector machines and multilayer perceptrons. These systems, after calculating the absolute value of the Least Absolute Shrinkage and Selection Operator (LASSO) regression coefficients, made it possible to rank the importance of each feature according to the severity of the patient [59]. The BS-EWM predictive model [41] based its analyses on an oversampling through the SMOTE procedure that allowed the predictions made through random forests to be extracted. Prediction with the convolutional neural model CheXNet-Cov19 [40] based on the CheXNet DL model [60] replaced the result layer on 3.5 nodes. This resulted in final prediction models using the Light Gradient Boosting Machine (LGBM) algorithm.

A common drawback to most of the AI systems developed was the need for validated datasets. This is particularly important in the case of chest radiographs since their intrinsic capability to detect abnormalities specific to COVID-19 is low compared to chest CT. In some cases, international medical image repositories, such as ImageNet or DenseNet121, were used to train the models, allowing millions of chest radiographs to be processed prior to validation and testing in hospital services [61, 62].

The main strengths of our study lie in the use of international standards in designing and reporting the results of this systematic review, sensitive search equation exploring the most relevant databases in health sciences, and screening and quality assessment by two independent researchers. In addition, the multidisciplinary nature of the research team allowed us to offer diverse and complementary perspectives to address our research hypothesis. Nevertheless, we restricted languages to English and Spanish, and we did not search unpublished documents. Therefore, it is possible that a selection bias affected our identified sample of studies. Moreover, the heterogeneity of the AI techniques included in our study prevented us from conducting precise comparisons, and, consequently, our results should be considered cautiously. In this same vein, future studies should be able to give quantitative information on how much AI algorithms help in the triage of patients in the ED. Another perceived limitation of our review is that discrepancies in the quality assessment were solved by consultation to a third researcher with experience in the field. Although this approach is common in systematic reviews and meta-analyses [63], other approaches such as a weighted average of researchers' decisions could have been used. Finally, most of the included studies were conducted on patients affected by SARS-CoV-2 variants different from the current circulating variants of concern, so the diagnostic accuracy of AI techniques nowadays might not exactly adjust to previously reported results.

Our results reinforce the need for investigating the potential applications of AI in the diagnosis and decision-making regarding COVID-19. Given the excess of clinical demand during epidemiological waves, the potential existence of future variant of concern and the progress made in the use of this technology for this pathology to date, it is recommendable to continue optimizing these models and implementing them into the EDs. Future studies should evaluate the impact of these techniques on saving time, improving healthcare costs, enhancing the quality of care and prognosis of patients with COVID-19.

5. Conclusions

Our study showed that the application of AI algorithms in radiological imaging makes triage of COVID-19 patients in the ED more efficient. In particular, DL methods based on the combination of chest radiography and scoring systems were associated with higher ROC AUC values compared to the control performed by experienced radiologists. However, further studies are needed to evaluate other parameters of efficiency in Emergency Services and to consider the application of these algorithms in the management of other pathologies.

AUTHOR CONTRIBUTIONS

PRG, AJLRB, and JMBS—conceptualized and designed the systematic review. PRG and MRI—wrote the first draft of the manuscript. PRG, CJG and PMJG—conducted the literature search, study selection, and data abstraction. MRI and AJLRB—analyzed the extracted data. All authors have read and agreed to the final version of the manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest. Antonio Jesús Láinez Ramos-Bossini is serving as one of the Guest editors of this journal. We declare that Antonio Jesús Láinez Ramos-

Bossini had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to OK.

REFERENCES

[1] WHO. WHO coronavirus (COVID-19) dashboard. 2022. Available at: https://covid19.who.int. (Accessed: 1 February 2022).

[2] Sharma A, Ahmad Farouk I, Lal SK. COVID-19: A Review on the Novel Coronavirus Disease Evolution, Transmission, Detection, Control and Prevention. Viruses. 2021; 13: 202.

[3] Jerez-Lillo N, Álvarez BL, Gutiérrez JM, Figueroa-Zúñiga JI, Leiva V. A statistical analysis for the epidemiological surveillance of COVID-19 in Chile. Signa Vitae. 2022; 18: 19–30.

[4] Ospina R, Leite A, Ferraz C, Magalhães A, Leiva V. Data-driven tools for assessing and combating COVID-19 outbreaks in Brazil based on analytics and statistical methods. Signa Vitae. 2022; 18: 18–32.

[5] Rivera-Izquierdo M, Valero-Ubierna MDC, Martínez-Diz S, Fernández-García MÁ, Martín-Romero DT, Maldonado-Rodríguez F, et al. Clinical factors, preventive behaviours and temporal outcomes associated with COVID-19 infection in health professionals at a Spanish hospital. International Journal of Environmental Research and Public Health. 2020; 17: 4305.

[6] Sánchez-Merino A, Huerta-Martínez MÁ, Zabava A-O. New-onset seizures associated with COVID-19: a systematic review. Archivos de Medicina Universitaria. 2021; 3: 66–79.

[7] Guiomar-Córdova A, Bustos-Romano M, Sánchez-Correa D. Mechanisms of SARS-CoV-2 entry into neuronal cells and its possible neurological sequelae. Archivos de Medicina Universitaria. 2021; 3: 112–123.

[8] Láinez Ramos-Bossini AJ, Pérez García MDC, Pérez Rosillo MÁ, Gálvez López R. Spontaneous pneumoperitoneum and pneumatosis intestinalis as sole manifestations of a COVID-19 infection. An extremely rare complication. The Spanish Journal of Gastroenterology. 2021; 113: 141–142.

[9] Láinez Ramos-Bossini AJ, Gálvez López R, Pérez Rosillo MÁ. Pneumoperitoneum and COVID-19. A causal association? The Spanish Journal of Gastroenterology. 2021; 113: 689–690.

[10] Rivera-Izquierdo M, Valero-Ubierna M del C, R-delAmo JL, Fernández-García MÁ, Martínez-Diz S, Tahery-Mahmoud A, et al. Sociodemographic, clinical and laboratory factors on admission associated with COVID-19 mortality in hospitalized patients: a retrospective observational study. PLOS ONE. 2020; 15: e0235107.

[11] Fernández-Martínez NF, Ortiz-González-Serna R, Serrano-Ortiz Á, Rivera-Izquierdo M, Ruiz-Montero R, Pérez-Contreras M, et al. Sex differences and predictors of in-hospital mortality among patients with COVID-19: results from the ANCOHVID multicentre study. International Journal of Environmental Research and Public Health. 2021; 18: 9018.

[12] Ramírez-García G, García-Molina P, Flor-Cremades MC, Muñoz-Rojas B, Moleón Moya J. Hydroxychloroquine and tocilizumab in the treatment of COVID-19: a longitudinal observational study. Archivos de Medicina Universitaria. 2021; 3: 32–39.

[13] Rivera-Izquierdo M, Valero-Ubierna MDC, R-delAmo JL, Fernández-García MÁ, Martínez-Diz S, Tahery-Mahmoud A, et al. Therapeutic agents tested in 238 COVID-19 hospitalized patients and their relationship with mortality. Medicina ClíNica. 2020; 155: 375–381.

[14] Aomar-Millán IF, Salvatierra J, Torres-ParejoÚ, Faro-Miguez N, Callejas-Rubio JL, Ceballos-Torres Á, et al. Anakinra after treatment with corticosteroids alone or with tocilizumab in patients with severe COVID-19 pneumonia and moderate hyperinflammation. A retrospective cohort study. Internal and Emergency Medicine. 2021; 16: 843–852.

[15] Dai M, Tao L, Chen Z, Tian Z, Guo X, Allen-Gipson DS, et al. Influence of cigarettes and alcohol on the severity and death of COVID-19: a multicenter retrospective study in Wuhan, China. Frontiers in Physiology. 2020; 11: 588553.

[16] Kucewicz-Czech E, Damps M. Triage during the COVID-19 pandemic. Anaesthesiology Intensive Therapy. 2020; 52: 312–315.

[17] Simons D, Shahab L, Brown J, Perski O. The association of smoking status with SARS-CoV-2 infection, hospitalization and mortality from COVID-19: a living rapid evidence review with Bayesian meta-analyses (version 7). Addiction. 2021; 116: 1319–1368.

[18] Romero-Duarte Á, Rivera-Izquierdo M, Guerrero-Fernández de Alba I, Pérez-Contreras M, Fernández-Martínez NF, Ruiz-Montero R, et al. Sequelae, persistent symptomatology and outcomes after COVID-19 hospitalization: the ANCOHVID multicentre 6-month follow-up study. BMC Medicine. 2021; 19: 129.

[19] De Pablos-Florido V, Córdoba-Peláez P, Jiménez-Gutiérrez PM. Chronic pain as a COVID-19 sequela: a systematic review. Archivos de Medicina Universitaria. 2021; 3: 86–91.

[20] Romero-Duarte Á, Rivera-Izquierdo M, Láinez-Ramos-Bossini AJ, Redruello-Guerrero P, Cárdenas-Cruz A. Factors associated with readmission to the emergency department in a cohort of COVID-19 hospitalized patients. Signa Vitae. 2022; 18: 47–54.

[21] Amador-Castilla LM, Camacho-Tapia JJ, Romero-Barrionuevo ME. Impact of COVID-19 on the mental health of doctors and nurses in Europe: an updated literature review. Archivos de Medicina Universitaria. 2021; 3: 164–171.

[22] Chou EH, Bhakta T, Tzeng C-FT, Shedd A, Hall M, Wolfsohl J, et al. The impact of coronavirus disease 2019 (COVID-19) on alternatives to opioids protocol for opioid reduction in emergency department. Signa Vitae. 2022; 3: 40–46.

[23] Jose S, Dhandapani M, Cyriac MC. Burnout and Resilience among Frontline Nurses during COVID-19 Pandemic: A Cross-sectional Study in the Emergency Department of a Tertiary Care Center, North India. Indian Journal of Critical Care Medicine. 2020; 24: 1081–1088.

[24] Ramos-Bossini AJL, Millán BL, Santiago FR, Rivera-Izquierdo M. Impact of the COVID-19 pandemic on osteoporotic vertebral fracture incidence and follow-up at the emergency department. A retrospective study of a tertiary hospital in southern Spain. Signa Vitae. 2022; 1–9.

[25] Adamidi ES, Mitsis K, Nikita KS. Artificial intelligence in clinical care amidst COVID-19 pandemic: a systematic review. Computational and Structural Biotechnology Journal. 2021; 19: 2833–2850.

[26] Öztürk Ş, Bayram A. Comparison of HOG, MSER, SIFT, FAST, LBP and CANNY features for cell detection in histopathological images. HELIX—The Scientific Explorer. 2018; 8: 3321–3325.

[27] Öztürk Ş. Image inpainting based compact hash code learning using modified U-Net. IEEE Xplore. 2020; 1–5.

[28] Bustos N, Tello M, Droppelmann G, García N, Feijoo F, Leiva V. Machine learning techniques as an efficient alternative diagnostic tool for COVID-19 cases. Signa Vitae. 2022; 18: 23–33.

[29] Oh Y, Park S, Ye JC. Deep learning COVID-19 features on CXR using limited training data sets. IEEE Transactions on Medical Imaging. 2020; 39: 2688–2700.

[30] Chiu WHK, Vardhanabhuti V, Poplavskiy D, Yu PLH, Du R, Yap AYH, et al. Detection of COVID-19 using deep learning algorithms on chest radiographs. Journal of Thoracic Imaging. 2020; 35: 369–376.

[31] Wang M, Xia C, Huang L, Xu S, Qin C, Liu J, et al. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. The Lancet Digital Health. 2020; 2: e506–e515.

[32] Goncharov M, Pisov M, Shevtsov A, Shirokikh B, Kurmukov A, Blokhin I, et al. CT-Based COVID-19 triage: deep multitask learning improves joint identification and severity quantification. Medical Image Analysis. 2021; 71: 102054.

[33] ÖZTÜRK Ş, YİĞİT E, ÖZKAYA U. Fused deep features based classification framework for COVID-19 classification with optimized MLP. Konya Journal of Engineering Sciences. 2020; 8: 15–27.

[34] Öztürk Ş, Özkaya U, Barstuğan M. Classification of coronavirus (COVID-19) from x-ray and CT images using shrunken features. International Journal of Imaging Systems and Technology. 2021; 31: 5–15.

[35] Ortíz-Barrios MA, Coba-Blanco DM, Alfaro-Saíz J-J, Stand-González D. Process improvement approaches for increasing the response of emergency departments against the COVID-19 pandemic: a systematic review. International Journal of Environmental Research and Public Health. 2021; 18: 8814.

[36] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for

reporting systematic reviews. The British Medical Journal. 2021; 372: n71.

[37] Redruello-Guerrero P, Jiménez-Gutiérrez C, Jiménez-Gutiérrez P, Rivera-Izquierdo M, Láinez-Ramos-Bossini AJ, Benítez Sánchez JM. Artificial Intelligence methods for the diagnosis and prognosis of COVID-19 in Hospital and Emergency Services: a systematic review. 2021. Available at: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021240732 (Accessed: 1 February 2022).

[38] Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2021. Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (Accessed: 1 February 2022).

[39] Bartolucci M, Benelli M, Betti M, Bicchi S, Fedeli L, Giannelli F, et al. The incremental value of computed tomography of COVID-19 pneumonia in predicting ICU admission. Scientific Reports. 2021; 11: 15619.

[40] Butler L, Karabayir I, Samie Tootooni M, Afshar M, Goldberg A, Akbilgic O. Image and structured data analysis for prognostication of health outcomes in patients presenting to the ED during the COVID-19 pandemic. International Journal of Medical Informatics. 2021; 158: 104662.

[41] Garrafa E, Vezzoli M, Ravanelli M, Farina D, Borghesi A, Calza S, et al. Early prediction of in-hospital death of COVID-19 patients: a machine-learning model based on age, blood analyses, and chest x-ray score. eLife. 2021; 10: e70640.

[42] Jiao Z, Choi JW, Halsey K, Tran TML, Hsieh B, Wang D, et al. Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. The Lancet Digital Health. 2021; 3: e286–e294.

[43] Kwon YJF, Toussie D, Finkelstein M, Cedillo MA, Maron SZ, Manna S, et al. Combining initial radiographs and clinical variables improves deep learning prognostication in patients with COVID-19 from the emergency department. Radiology: Artificial Intelligence. 2021; 3: e200098.

[44] La Salvia M, Secco G, Torti E, Florimbi G, Guido L, Lago P, et al. Deep learning and lung ultrasound for Covid-19 pneumonia detection and severity classification. Computers in Biology and Medicine. 2021; 136: 104742.

[45] Li MD, Arun NT, Aggarwal M, Gupta S, Singh P, Little BP, et al. Improvement and multi-population generalizability of a deep learning-based chest radiograph severity score for COVID-19. MedRxiv. 2020; 2020.09.15.20195453.

[46] Mushtaq J, Pennella R, Lavalle S, Colarieti A, Steidler S, Martinenghi CMA, et al. Initial chest radiographs and artificial intelligence (AI) predict clinical outcomes in COVID-19 patients: analysis of 697 Italian patients. European Radiology. 2021; 31: 1770–1779.

[47] Schiaffino S, Codari M, Cozzi A, Albano D, Alì M, Arioli R, et al. Machine learning to predict in-hospital mortality in COVID-19 patients using computed tomography-derived pulmonary and vascular features. Journal of Personalized Medicine. 2021; 11: 501.

[48] Shamout FE, Shen Y, Wu N, Kaku A, Park J, Makino T, et al. An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department. NPJ digital medicine. 2021; 4: 80.

[49] Weikert T, Rapaka S, Grbic S, Re T, Chaganti S, Winkel DJ, et al. Prediction of patient management in COVID-19 using deep learning-based fully automated extraction of cardiothoracic CT metrics and laboratory findings. Korean Journal of Radiology. 2021; 22: 994–1004.

[50] Warren MA, Zhao Z, Koyama T, Bastarache JA, Shaver CM, Semler MW, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. Thorax. 2018; 73: 840–846.

[51] Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. La Radiologia Medica. 2020; 125: 509–513.

[52] Toussie D, Voutsinas N, Finkelstein M, Cedillo MA, Manna S, Maron SZ, et al. Clinical and chest radiography features determine patient outcomes in young and middle-aged adults with COVID-19. Radiology. 2020; 297: E197–E206.

[53] Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). Radiology. 2020; 295: 202–207.

[54] Benelli M, Barucci A, Zoppetti N, Calusi S, Redapi L, Della Gala G, et al. Comprehensive analysis of radiomic datasets by RadAR. Cancer Research. 2020; 80: 3170–3174.

[55] Mongodi S, Bouhemad B, Orlando A, Stella A, Tavazzi G, Via G, et al. Modified lung ultrasound score for assessing and monitoring pulmonary aeration. European Journal of Ultrasound. 2017; 38: 530–537.

[56] Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti DF, et al. Proposal for international standardization of the use of lung ultrasound for patients with COVID-19: a simple, quantitative, reproducible method. Journal of Ultrasound in Medicine. 2020; 39: 1413–1419.

[57] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; 770–778.

[58] Kingma DP, Ba J. 'Adam: a method for stochastic optimization', the 3rd International Conference for Learning Representations. San Diego, 2015. ICLR: San Diego. 2015.

[59] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996; 58: 267–288.

[60] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; 2097–2106.

[61] Putha P, Tadepalli M, Reddy B, Raj T, Chiramal JA, Govil S, et al. Can artificial intelligence reliably report chest x-rays? Radiologist validation of an algorithm trained on 2.3 million x-rays. Cornell University. 2019; 1–13.

[62] Deng J, Dong W, Socher R, Li L, Kai Li, Li Fei-Fei. Imagenet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009; 248–255.

[63] Láinez Ramos-Bossini AJ, López Zúñiga D, Ruiz Santiago F. Percutaneous vertebroplasty versus conservative treatment and placebo in osteoporotic vertebral fractures: meta-analysis and critical review of the literature. European Rdiology. 2021; 31: 8542–8553.