74264_Auto_Edited.docx

Applications of artificial intelligence in lung ultrasound: review of deep learning methods for COVID-19 fighting

Laura De Rosa, Serena L'Abbate, Claudia Kusmic, Francesco Faita

Abstract

BACKGROUND

The pandemic outbreak of the novel coronavirus disease (COVID-19) has highlighted the need to combine rapid, non-invasive and widely accessible techniques with the least risk of patient's cross-infection to achieve a successful early detection and surveillance of the disease. In this regard, the lung ultrasound (LUS) technique has been proved invaluable in both the differential diagnosis and the follow-up of COVID-19 patients, and its potential may be destined to evolve. Recently, indeed, LUS has been empowered through the development of automated image processing techniques.

AIM

Here, we provide a systematic review of the application of artificial intelligence (AI) technology in medical LUS analysis of COVID-19 patients using the preferred reporting items of systematic reviews and meta-analysis (PRISMA) guidelines.

METHODS

A literature search was performed for relevant studies published from March 2020 - outbreak of the pandemic - to 30 September 2021. Seventeen articles were included in the result synthesis of this paper.

RESULTS

As part of the review, we presented the main characteristics related to AI techniques, in particular deep learning (DL), adopted in the selected articles. A survey was carried out on the type of architectures used, availability of the source code, network weights and open access datasets, use of data augmentation, use of the transfer learning strategy, type of input data and training/test datasets, and explainability.

CONCLUSION

Finally, this review highlighted the existing challenges, including the lack of large datasets of reliable COVID-19-based LUS images to test the effectiveness of DL methods and the ethical/regulatory issues associated with the adoption of automated systems in real clinical scenarios.

INTRODUCTION

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) is a life-threatening infectious virus and its related disease (COVID-19) represents a still ongoing challenge for humans. At time of writing, over 418 million infections have been recorded worldwide including more than 5.8 million attributable deaths[1]. Despite the large number of vaccination programs introduced from the end of 2020 has represented an opportunity to minimise the risk of severe COVID-19 and death, the spread of new genetic viral variants with a higher probability of contagion has raised a renewed strong concern for either not vaccinated and vaccinated people. Thus, since the

outbreak of the pandemic, research has continuously looked for a quick and reliable way to diagnose the disease, treat and monitor people affected by coronavirus.

To date, molecular test based on real time quantitative reverse transcription polymerase chain reaction (RT-qPCR) assay by nasopharyngeal swabs along with the serological antibody-detecting and antigen-detecting tests are the current accepted diagnostic tools for the conclusive diagnosis of COVID-19[2]. RT-qPCR may take up to 24 h to provide information and requires multiple tests for definitive results and, in addition, it is not relevant to assess the disease severity. Furthermore, the accuracy of molecular and serological tests remains highly dependent on timing of sample collection relative to infection, improper sampling of respiratory specimens, inadequate preservation of samples and technical errors, particularly contamination during RT-qPCR process and cross-reactivity in the immunoassay[3,4].

To complement conventional *in vitro* analytical techniques of COVID-19, biomedical imaging techniques have demonstrated great potential in clinical diagnostic evaluation by providing rapid patient assessment in the presence of high pre-test probability. Furthermore, imaging techniques are currently important in the follow-up of subjects with COVID-19[5,6]. Among the imaging techniques, chest computed tomography (CT) is considered the primary diagnostic modality and an important indicator for assessing severity and progression of COVID-19 pneumonia[7,8], although it has been reported to have limited specificity[9,10,11]. Indeed, the CT imaging features can overlap between COVID-19 and other viral pneumonia. Moreover, CT scanning is expensive, not easy to perform in the COVID-19 context, and multiple risks are associated with it, such as radiation exposure and cross-infection risk associated with repeated use of a CT suite[12], along with unavailability of CT in many parts of the world.

In the last few years, lung ultrasound (LUS) technique has become increasingly popular and a good option for real-time point-of-care testing, with several advantages making it a valuable tool in the fight against COVID-19[13], although it has specificity limits comparable to those of chest CT.

Ultrasound (US) is a low-cost, non-radioactive medical imaging method, particularly indicated for evaluation in pregnant women and children, which is portable to the bedside or patient's home and is easy to sterilise. Moreover, the risk of COVID-19 cross-infection can be limited by making use of disposable ultrasound gel with a portable probe[14]. In addition, some studies indicate that LUS shows excellent performances in speed of execution and accuracy of diagnosis in case of respiratory failure[15]. Furthermore, compared with chest X-ray, LUS demonstrated higher sensitivity in detecting pneumonia[16] and similar specificity in the diagnosis of pneumothorax[15]. On the other hand, the distinctive LUS features (B-lines, consolidations, pleural thickening and brokening) observed in patients with varying severity of COVID pneumonia are similar to the features seen in patients with pneumonia of different aetiologies. Indeed, a recent review[17] on ultrasound findings of LUS in COVID-19 demonstrated that LUS has high sensitivity and reliability in ruling out lung involvement, but at the expense of low specificity. Therefore, especially in the case of low prevalence of the disease, at present LUS cannot be considered a valid gold standard in clinical practice.

Ultrasound image processing techniques have assumed great importance in recent years, with the growing experience that accurate image processing can significantly help in extracting quantitative characteristics to assess and classify the severity of diseases. Accordingly, sophisticated techniques of automated image processing, that include the use of artificial intelligence (AI) methods, have been developed and applied to assist LUS imaging in the detection of COVID-19 and make such assessment more objective and accurate. AI methods - from machine learning (ML) to deep learning (DL), indeed, aim to imitate cognitive functions and stand out in automatically recognizing complex patterns in imaging data, providing quantitative rather than qualitative assessments. The primary purpose of applying AI methods in medical imaging is to improve the visual recognition of certain features in images to produce lower-than-human error rates. Furthermore, an enhancement in LUS performance can reduce the use of more invasive and time-consuming techniques, facilitating both faster diagnosis and recognition of earlier stages of the disease[18]. To allow a quick development of highly performant AI models, a large amount of accessible and validated data to train and test AI models is a critical

requirement that can be achieved, for instance, with the development of shared big data archives. Indeed, one of the most common problems associated with using limited training samples is the over-fitting of DL models. To address this issue, two main approaches can be selected: model optimization and transfer learning. These strategies significantly improve the performance of DL models. Likewise, data pre-processing and data augmentation/enhancement can be useful additional strategies[19,20].

The most common applications of DL methods in clinical imaging, and hence in medical ultrasound imaging as well, are object detection, object segmentation, and object classification[21]. The main architectures applied in current analysis are convolutional neural networks (CNNs) and recurrent neural networks (RNNs)[22]. CNNs are architectures able to work with 2D and 3D input images and RNNs recognize the image's sequential characteristics and use patterns to predict the next likely scenario[23].

Since the outbreak of the pandemic, many proposals have been made based on AI methods applied to LUS scans of COVID-19 patients. Here we propose a comprehensive systematic review of the literature on the use of AI technology, DL in particular, to aid in the fight against COVID-19.

MATERIALS AND METHODS

Study selection

A literature search to identify all relevant articles on the use of DL tools applied to LUS imaging in patients affected by COVID-19 virus was conducted.

This systematic review was carried out using the PubMed/Medline electronic database and according to the Preferred Reporting for Systematic Reviews and Meta-Analysis (PRISMA) guidelines[24,25]. We performed a systematic search covering the period from March 2020 (from the outbreak of the pandemic) to 30 September 2021. The search strategy was restricted to English-language publications.

We performed an advanced research concatenating terms with Boolean operators. In particular, search words and key terms used in the search included ("lung ultrasound" OR "lus") AND ("Covid-19" OR "coronavirus" OR "SARS-CoV2") AND ("artificial intelligence" OR " deep learning" OR "neural networks" OR "CNN").

Eligibility criteria

The inclusion criteria were:

Studies that include COVID-19 patients with LUS acquisitions and developed or tested DL-based algorithms on LUS images or on features extracted from the images;

No restriction on the ground truth adopted to analyse the presence/absence of COVID-19 and/or the severity of lung disease (e.g., PCR, visual evaluation of video/images and score assignment by expert clinicians);

No restriction on the type of DL architecture used in the studies;

Studies on paediatric population were excluded;

Studies were restricted to peer reviewed articles and conference proceedings. However, the following publication types were excluded: reviews and conference abstracts.

Data extraction and analysis

Two investigators (LDR and FF) screened the articles independently. Disagreement between reviewers was resolved by consensus *via* discussion. The reasons for the exclusion of some trials are described in the Results section. Publications by the same research group or by different groups using the same dataset were included in the analysis. After the selection of the articles, we collected the following

characteristics: first author's surname, date of publication, sample size, general characteristics of the study populations, AI techniques used, validation methods and main results obtained.

The study selection process is presented in Figure 1.

RESULTS

Search results

Twenty-four articles resulted after querying the database and screened for eligibility (Figure 1). Of the 24 articles, we discarded four references as review papers. After examining the titles and abstracts, we excluded five articles: one manuscript did not include DL methods applied on US imaging, three papers were not based on AI and DL approaches, and one article was focused on the paediatric population. Moreover, two additional papers, retrieved from the checking of references of the eligible articles, were included.

Finally, 17 articles[26-42] were selected for full-text screening and included in our analysis (Table 1 and Table 2).

The following part of the section provides a concise overview of the studies' main features.

Dataset and source code availability

Authors of seven[27-30,33,39,40] of the seventeen selected articles (41.2%) extrapolated their datasets from the free access LUS database acquired by point-of-care ultrasound (POCUS) imaging and made available firstly by Born *et al*[30]. Instead, an Italian group firstly introduced the Italian COVID-19 Lung Ultrasound DataBase (ICLUS-DB)[38], which is accessible upon mandatory request to the authors, and that was used in two other studies[32,37]. Noteworthy, Roy *et al*[38] have created a platform through which physicians can access algorithms, upload their data and see the algorithm's evaluation of the data.

Besides dataset open access, access to the code for the neural network is also important to reproduce results and compare performances. Seven articles[26-30,32,38] (41.2%) made the source code implementing the proposed DL architecture available for download from the Git-hub repository.

Single-frame/multi-frames or video based architecture

In the majority of the selected papers, DL architectures work with single frame images as input and only three publications[29,34,41] (17.6%) report DL architectures based on image sequences (i.e., video). However, six studies[28,30,32,37-39] (35.3%), despite adopting a DL architecture designed to perform single-frame classification, also propose additional methods to fulfil video-based classification. In particular, Roy *et al*[38] proposed an aggregation layer system of frame-level scores to produce predictions on LUS videos and Mento *et al*[37] proposed an alternative video-based classification using a threshold-based system on the frame-level scores obtained from DL architecture.

Other authors[32] adopted a Long Short-Term Memory (LSTM) system, which has been used to exploit temporal relationships between multiple frames by taking long time series as input, over performing their results obtained by CNN without LSTM.

Finally, Xue *et al*[42] applied AI models for patient-level assessment of severity using a final module across the entire architecture that works with ML rather than DL systems.

Test strategy of DL models

The proposed DL models have been tested on a database entirely independent from the training database in seven articles[26,35-,39,42] (41.2%); five-fold and ten-fold cross-validation techniques were applied in nine[27-34,40] (52.9%) and one[41] (5.9%) studies, respectively. Among the papers that tested DL models on an independent database, the percentage of data used for the testing ranged from 33%[35]

to 20%[38] and 10%[26,36] of the overall data. Born *et al*[29], alongside the five-fold cross-validation technique in the training/test phase of the DL model, also used an independent validation dataset made-up of 31 videos (28 convex and 3 Linear probes) from six patients. Indeed, Roy *et al*[38], for instance, used 80 videos/10709 frames out of the total 277 videos/58924 frames to test their DL model.

In all studies, the splitting of data between training set and test set was performed either at the patient-level or at the video-level. Thus, all the frames of a single video clip belonged either to the training or to the test set.

Data augmentation

Twelve (70.6%) research groups extended their LUS database by augmentation. The main strategies for data augmentation applied to LUS images were:

horizontal/vertical flipping[26,27,29,30,32,33,36,38-40,42], bidirectional arbitrary rotation[26,27,29,30,32,33,35,38-40,42], horizontal and vertical shift[30,32,38,39,42];

filtering, colour transformation, adding salt and pepper noise, Gaussian noise[36,38,42], normalisation of grey levels' intensity[38].

Although proposed by all the authors, only seven papers[26,29,30,32,33,38,40] provided details on the amplitude of image rotation. In particular, Dastider *et al*[32] applied rotations in the range of 0 ± 360 degrees, while other authors have limited image rotations to 10 degrees[26,29,30,33], ± 15 degrees[38] and ± 20 degrees[40], respectively.

The remaining five papers[28,31,34,37,41] (29.4%) did not perform data augmentation.

Explainability

Among the selected articles, tools for interpreting the network output were provided in twelve studies (70.6%), whereas in the remaining five (29.4%) the DL algorithms' outcomes were proposed as black box systems. The majority of papers[26-29,32,35,36,38,40]

reported the Gradient-weighted Class Activation Mapping (Grad-CAM) as the preferred explainability tool. Grad-CAM uses gradients to create a location map to highlight the region of interest of the images[43]. Instead, Sadik *et al*[39] used a colormap jet to visualise a heat map overlay to US images; Ebadi *et al*[34] adopted an activation map system to detect and segment features in LUS scans. Furthermore, one study[42] showed LUS images with overlaid colormaps to indicate the segmentation zone of ultrasound according to the different severity. Roy *et al*[38], differently, provided an ultrasound colormap overlay on the LUS frame/video and used four colours to distinguish the different classes of disease severity recognized by DL architecture.

Clinical use

Most of the selected papers applied the AI system to diagnose COVID-19 and/or discriminate between COVID-19 and other lung diseases (such as bacterial pneumonia)[26-30,33,34,39,40]. The first approach using DL architecture for automatic differential diagnosis of COVID-19 from LUS data was POCOVID-Net[30].

However, a fair number of studies have focused on assessing the severity of COVID-19[31,32,35-38,42]. In particular, a disease severity score is assigned to the single image according to some characteristics visible in the image pattern. Most of the articles used four severity classes by assigning a score to the single frame from 0 to 3[31,32,35-38], as defined by Soldati e*t al*[44]. Xue *et al*[42] proposed a classification in five classes of pneumonia severity (score from 0 to 4) along with a binary severe/non-severe classification. Furthermore, these authors used the DL technology exclusively to implement a segmentation phase based on a VGG network, while the classification phase still employed a more traditional, features-based machine learning approach. Finally, La Salvia *et al*[36] proposed a classification based on three severity classes and a modified version considering a seven-classes scenario.

Furthermore, Arntfield et al[26] showed that their network was able to recognize pathological pattern in LUS images with higher sensitivity than sonographers; whilst an InceptionV3 network proposed by Diaz-Escobar et al[33] was able to discriminate COVID-19 pneumonia from healthy lung and other bacterial pneumonia with an accuracy of 89.1% and an area under the ROC curve of 97.1%. Curiously, one of the eligible papers[41] did not include confirmed cases of COVID-19 patients. The authors' aim was to design an algorithm capable of identifying the presence of pleural effusion. However, we have included this work in our systematic review, because small pleural effusions are rarely reported in COVID-19 patients. Therefore, the detection of pneumonia with pleural effusion can help rule out the hypothesis of COVID-19 disease.

Transfer learning and DL architecture

From our analysis, it emerged that most of the studies have proposed convolutional neural networks (CNNs) as DL models to generate screening systems for COVID-19. In particular, all publications with the exception of one[31] used the CNN network. Conversely, Chen et al[31] developed a multi-layer fully connected neural network for scoring LUS images in assessing the severity of COVID-19 pneumonia.

Among the DL systems included in this review, most of them were generated starting from DL architectures already proposed for other tasks[26-30,32-36,39,42], suitably modified and trained for new tasks. Furthermore, many works compared the results of their architectures with those obtained using existing and well-known architectures[27-30,32,33,35,38-40].

In particular, the following DL architectures were adapted to fulfil the requirements of LUS analysis to assist in COVID-19 detection and/or assessment of the severity of the lung disease, or just to compare their performances:

VGG-19[28,33,39] and VGG-50[28-30,33];

Xception[26,28,39];

ResNet 50[27,33,36,40];

NasNetMobile[27,29,39];

DenseNet[32,39].

More in detail, Awasthi *et al*[27] proposed Mini-COVIDNet, a modified MobileNet model belonging to the CNN's networks family and originally developed for detecting objects in mobile applications[45]. Barros *et al*[28], along with their proposed DL model, also investigated the impact of using different pre-trained CNN architectures in extracting spatial features that were successively classified by a LSTM model. Finally, Born *et al*[29] derived their DL video-based models from a model that was pre-trained on lung CT scans[46]. All aforementioned architectures are pre-trained on ImageNet[47].

Sample size

Partly due to the recent outbreak of the pandemic and to the difficulty of having standardised high quality archives of US images, only few of the selected studies relied on a large dataset in terms of enrolled patients. Six papers (35.3%) reported a sample size greater than 200 subjects (namely, 243, 216, 216, 300, 450 and 313 in ref. [26,29,33,34,36,42] respectively).

However, despite the relatively low number of subjects, the total number of LUS videos reaches up to 5400 in one study[36], with an average equal to 1589 videos[26,29,33,34,36].

Among the studies carried out on a low sample size, Dastider *et al*[32] included 29 patients and 60 videos, whilst 35 patients/45 videos and 35 patients/277 videos were analysed in references [31] and [38], respectively. However, it should be noted that Roy *et al*[38] published their work at the beginning of the COVID-19 pandemic, when the total number of COVID-19 patients was still relatively limited.

In the paper by Xue *et al*[42], the number of frames/video was not reported.

DISCUSSION

The paper reviews the different DL techniques able to work with LUS images in assisting the diagnosis and/or prognosis of the COVID-19 disease published since the outbreak of the pandemic. In the selected documents, the use of DL systems aimed to achieve an accuracy comparable to or better than clinical standards to provide a faster diagnosis and/or follow-up in COVID-19 patients.

Most of the papers present pre-trained DL architectures[26-30,32-36,39,42] that were modified and adapted to new data. This approach is also known as transfer learning (TL) technique - i.e., a training strategy for new DL models with reduced datasets. The network is pre-trained on a very large dataset, such as ImageNet, with millions of images intentionally created to facilitate the training of DL models, focusing on image classification and object location/detection tasks[48]. Indeed, deeper models are difficult to train and provide inconsistent performances when trained on a limited amount of data[49]. Therefore, most of the studies based on DL systems to classify COVID-19 images appropriately use the TL strategy as large datasets of US images from COVID-19 patients are not yet easily available, partly because the coronavirus disease is a relatively recent concern.

Furthermore, most of the proposed systems shared the same design, i.e., CNN's architectures. CNNs have several applications in medical imaging – among others, image segmentation and object detection[50]. However, CNNs are particularly suited for image classification problems[51] and, consequently, represent an optimal solution for the classification of the disease severity from US images.

To date, one of the main challenges faced by DL architectures applied to LUS images of COVID-19 patients are the limited datasets in the available databases. This problem could benefit from creating open access databases that collect large amounts of data from multiple centres. In some of the selected studies, a first attempt to overcome this issue is evident, with particular emphasis on the

work by Born *et al*[30], the authors who first collected a free access dataset of lung images from healthy controls and patients affected by COVID-19 or other pneumonia.

The development of public and multicentre platforms would guarantee the collection of a continuously growing amount of data, large and highly heterogeneous, suited for the training and testing of new DL applications in medical imaging, both in the COVID-19 and LUS field. Furthermore, this would allow an easier comparison of performances among DL models proposed in different studies. However, alternative approaches are often used in the testing phase that do not require the use of independent data sets to evaluate the performance of the model in the event of a limited number of images available. Among these, the k-fold cross-validation is a statistical method used to evaluate the ability of ML models to generalise to previously unseen data. Despite being widely used in ML models, the k-fold cross validation approach is less reliable than tests performed using an external dataset; the latter is always preferable to test model's ability to adapt properly to new, previously unseen data.

Data augmentation techniques are an alternative strategy to overcome the issue of the limited amounts of data, largely adopted in practice. These techniques generate different versions of a real dataset artificially to both increase its size and the power of model's generalisation. Despite the great advantage in increasing data to feed DL architectures, data augmentation techniques should be used with awareness, as some geometric transformations could be unrealistic when applied to LUS images (e.g., angles of rotations greater than 30°). In the field of DL applied to medical imaging, the use of architectures designed to work with 3D images is another interesting challenge. Indeed, a DL system that operates with 3D data input usually requires a larger amount of data for training, as a 3D network contains a parameters' number that is orders of magnitude greater than a 2D network. This could significantly increase the risk of overfitting, especially in the case of limited dataset availability. In addition, the training on large amounts of data requires high computational costs associated with memory and performance requirements of the tools used. LUS images are usually recorded in the form of videoclips (2D + time) and can be assimilated to 3D data. Exploitation of dynamic information naturally embedded in

image sequences has proven very important in the analysis of lung echoes. In particular, changes induced by COVID-19 viral pneumonia are better detectable in LUS through the analysis of multi-frames acquisition due to its ability in capturing dynamic features, *e.g.*, pleural sliding movements and generation of B-line artefacts[44].

Regardless of the data format (i.e., 3D, 2D or 2D+time images), the labelling of ground truth data is required in supervised DL applications and should be provided by skilled medical professionals. However, it is a time-consuming activity, in particular in the 2D approach that is characterised by a high number of samples.

Indeed, some authors demonstrated that the performance in pleural effusion classification on LUS images obtained with the video-based approach was comparable to that obtained with frame-based analysis, despite a significant reduction in labelling effort[41]. Furthermore, Kinetics-I3D network was able to classify LUS video sequences with great accuracy and efficiency[34]. On the other hand, the video-based approach has also revealed a reduced accuracy in patients classification with respect to the single frame analysis; however, this could be explained by the relatively reduced number of available LUS clips[29].

Extending the use of DL architectures beyond multi-frame analysis with respect to single 2D images is highly desirable. In particular, these methods could be effectively used to assign a patient-level disease severity score. In fact, this information plays a key role in the selection of treatment, monitoring of disease progression and management of medical resources (e.g., mechanical ventilator needed).

Code availability is another very critical issue in applications of AI in medical imaging. Indeed, the lack of ability to reproduce the training of the proposed DL models or to test these models on new US images is a rather widespread problem. Often, authors do not provide access to either the source code used to train NNs or the final weight of the trained network. On the other hand, the availability of this information would greatly facilitate the diffusion of new AI systems in the clinical setting.

DL systems are often presented as black boxes - i.e., they produce a result without providing a clear understanding in "human terms"of how it was obtained. The black-box nature of the algorithms has restricted their clinical use until now. Consistently, the

explainability - i.e., making clear and understandable the features that influence the decisions of a DL model - is a critical point to guarantee a safe, ethical, and reliable use of AI. Especially in medical imaging applications, explainability is very important as it gives the opportunity to highlight regions of the image containing the visual features that are critical for the diagnosis. Gradient-weighted Class Activation Mapping (Grad-CAM) is a promising technique for producing "visual explanations" of decisions taken from a large class of CNN-based models, making their internal behaviour more understandable, thus partially overcoming the black-box problem. The basic idea is to produce a rough localization map that highlights the key regions in the image that have a major effect on customization of network parameters, thus maximally contributing to the prediction of outcomes[43].

These maps visualised areas using a blue-to-red scale, with the highest/Lowest contribution to the class prediction operated by the model. The clinical use of DL systems is a crucial issue. One of the major current limitations of LUS imaging in COVID patients is the specificity. Focusing the design of DL systems to overcome this limit could really represent a benefit in the clinical setting.

Along this line, some of the included studies tested the agreement between physicians' ability to classify COVID-19 patients and that proposed by neural networks. Furthermore, this finding suggests that the automated system can capture some features (biomarkers) in US images that are not clearly visible to the human eye.

Finally, another important issue to mention is the use of the quantitative evaluation indicators and the analysis of the benchmarking techniques adopted to evaluate the effectiveness of the proposed methods. Unfortunately, the tools examined in the selected manuscripts had very heterogeneous targets (Table 1, Main results column), ranging from diagnostic to prognostic purposes or assessment of disease severity. This dispersion of intent and the few articles published in the literature at present make any comparison or analysis very difficult.

CONCLUSION

The studies analysed in this article have shown that DL systems applied to LUS images for the diagnosis/prognosis of COVID-19 disease have the potential to provide significant support to the medical community. However, there are a number of challenges to overcome before AI systems can be regularly employed in the clinical setting. On the one hand, the critical issues related to the availability of high-quality databases with large sample size of lung images/videos of COVID-19 patients and free access to datasets must be addressed. On the other hand, existing concerns about the methodological transparency (e.g., explainability and reproducibility) of DL systems and the regulatory/ethical and cultural issues that the clinical use of AI methods raise must be resolved. Finally, a closer collaboration between the communities of informatics/engineers and medical professionals is desirable to facilitate the outcome of adequate guidelines for the use of DL in US pulmonary imaging and, more generally, in medical imaging.

ARTICLE HIGHLIGHTS

Research background

The current COVID-19 pandemic crisis has highlighted the need for biomedical imaging techniques in rapid clinical diagnostic evaluation of patients. Furthermore, imaging techniques are currently important in the follow-up of subjects with COVID-19. The lung ultrasound technique has become increasingly popular and is considered a good option for real-time point-of-care testing, although it has specificity limits comparable to those of chest CT.

Research motivation

The application of artificial intelligence, and of deep learning in particular, in medical pulmonary ultrasound can offer an improvement in diagnostic performance and classification accuracy to a non-invasive and low-cost technique, thus implementing its diagnostic and prognostic importance to COVID-10 pandemic.

Research objectives

This review presents the state of the art of the use of artificial intelligence and deep learning techniques applied to lung ultrasound in COVID-19 patients

Research methods

We performed a literature search, according to PRISMA guidelines, for relevant studies published from March 2020 - to 30 September 2021 on the use of deep learning tools applied to lung ultrasound imaging in COVID-19 patients. Only English-language publications were selected.

Research results

We surveyed the type of architectures used, availability of the source code, network weights and open access datasets, use of data augmentation, use of the transfer learning strategy, type of input data and training/test datasets, and explainability.

Research conclusions

Application of deep learning systems to lung ultrasound images for the diagnosis/prognosis of COVID-19 disease has the potential to provide significant support to the medical community. However, there are critical issues related to the availability of high-quality databases with large sample size and free access to datasets.

Research perspectives

Close collaboration between the communities of computer scientists / engineers and medical professionals could facilitate the outcome of adequate guidelines for the use of deep learning in ultrasound lung imaging.

74264_Auto_Edited.docx

ORIGINALITY REPORT

3%

SIMILARITY INDEX

PRIMARY SOURCES

1 www.esp.org
Internet
45 words — 1%

2 Bruno Barros, Paulo Lacerda, Célio Albuquerque, Aura Conci.
"Pulmonary COVID-19: Learning Spatiotemporal Features Combining
CNN and LSTM Networks for Lung Ultrasound Video Classification", Sensors, 2021
Crossref
24 words — < 1%

3 Yan Wang, Zizhou Wang, Yangqin Feng, Lei Zhang. "WDCCNet: Weighted
Double-Classifier Constraint Neural Network for Mammographic Image
Classification", IEEE Transactions on Medical Imaging, 2022
Crossref
22 words — < 1%

4 pubmed.ncbi.nlm.nih.gov
Internet
15 words — < 1%

5 huggingface.co
Internet
14 words — < 1%