Preprints.org

Review

Ultrasound-Based AI for COVID-19 Detection: A Comprehensive Review of Public and Private Lung Ultrasound Datasets and Studies

Abrar Morshed , Abdulla Al Shihab , Md Abrar Jahin [*] , Md Jaber Al Nahian , Md Murad Hossain Sarker , Md Sharjis Ibne Wadud , Mohammad Istiaq Uddin , Muntequa Imtiaz Siraji , Nafisa Anjum , Sumiya Rajjab Shristy , Tanvin Rahman , Mahmuda Khatun , Md Rubel Dewan , Mosaddeq Hossain , Razia Sultana , Ripel Chakma , Sonet Barua Emon , Towhidul Islam , Mohammad Arafat Hussain [*]

Review

Ultrasound-Based AI for COVID-19 Detection: A Comprehensive Review of Public and Private Lung Ultrasound Datasets and Studies

Abrar Morshed [1,†], Abdulla Al Shihab [2,†], Md Abrar Jahin [3,4,*,†], Md Jaber Al Nahian [5,†], Md Murad Hossain Sarker [6,†], Md Sharjis Ibne Wadud [7,‡], Mohammad Istiaq Uddin [8,†], Muntequa Imtiaz Siraji [9,†], Nafisa Anjum [10,†], Sumiya Rajjab Shristy [11,†], Tanvin Rahman [12,†], Mahmuda Khatun [13,†], Md Rubel Dewan [14,†], Mosaddeq Hossain [15,†], Razia Sultana [16,†], Ripel Chakma [17,†], Sonet Barua Emon [18,†], Towhidul Islam [19,†] and Mohammad Arafat Hussain [20,*,‡]

1   Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, 1208, Bangladesh; abrar03.morshed@gmail.com
2   Department of Electrical and Telecommunication Engineering, Daffodil International University, Dhaka, 1216, Bangladesh; Abdullahshihab54321@gmail.com
3   Department of Industrial Engineering and Management, Khulna University of Engineering & Technology (KUET), Khulna, 9203, Bangladesh
4   Physics and Biology Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Okinawa, 904-0412, Japan
5   Department of Information and Communication Technology, Bangladesh University of Professionals, Dhaka, 1216, Bangladesh; nahianrism@gmail.com
6   Department of Information and Communication Technology, Comilla University, Cumilla, 3506, Bangladesh; mh6367828@gmail.com
7   Department of Biomedical Physics and Technology, University of Dhaka, Dhaka, 1000, Bangladesh; sharjis@du.ac.bd
8   Department of Computer Science and Engineering, International Islamic University Chittagong, Chittagong, 4318, Bangladesh; mistiaq@ieee.org
9   Department of Electrical and Electronic Engineering, Islamic University of Technology, Gazipur, 1704, Bangladesh; muntequaimtiaz@iut-dhaka.edu
10   Department of Electronics and Telecommunication Engineering, Chittagong University of Engineering and Technology, Chittagong, 4349, Bangladesh; u1708003@student.cuet.ac.bd
11   Department of Computer Science and Engineering, Notre Dame University Bangladesh, Dhaka, 1000, Bangladesh; sumiyashristy757@gmail.com
12   Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology, Khulna, 9203, Bangladesh; tanvinrahman28@gmail.com
13   Department of Computer Science and Engineering, Institute of Science and Technology, Dhaka, 1209, Bangladesh; mahmudamohima14@gmail.com
14   Department of Electronics and Communication Engineering, Institute of Science and Technology, Dhaka, 1209, Bangladesh; mdrubeldewan99625@gmail.com
15   Department of Mathematics, Jahangirnagar University, Dhaka, 1342, Bangladesh; tanjib.ju.42@gmail.com
16   Department of Electrical and Electronic Engineering, Daffodil International University, Dhaka, 1216, Bangladesh; raziazuyenna@gmail.com
17   Department of Electrical and Electronic Engineering, Dhaka University of Engineering and Technology, Gazipur, 1707, Bangladesh; ripel.chakma13@gmail.com
18   Department of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University, Noakhali, 3814, Bangladesh; sonetcste11@gmail.com
19   Department of Computer Science and Engineering, Northern University Bangladesh, Dhaka, 1230, Bangladesh; towhidul.islam@nub.ac.bd
20   Boston Children's Hospital, Harvard Medical School, Boston, MA, 02115, USA
*   Correspondence: abrar.jahin.2652@gmail.com (M.A.J.); mohammad.hussain@childrens.harvard.edu (M.A.H.)
†   These authors contributed equally to this work.
‡   Co-first Authors.

Abstract: The COVID-19 pandemic has affected millions of people globally, with respiratory organs being strongly affected in individuals with comorbidities. Medical imaging-based diagnosis and prognosis have become increasingly popular in clinical settings for detecting COVID-19 lung infections. Among various medical imaging modalities, ultrasound stands out as a low-cost, mobile, and radiation-safe imaging technology. In this comprehensive review, we focus on AI-driven studies utilizing lung ultrasound (LUS) for COVID-19 detection and analysis. We provide a detailed overview of both publicly available and private LUS datasets and categorize the AI studies according to the dataset they used. Additionally, we systematically analyzed and tabulated the

studies across various dimensions, including data preprocessing methods, AI models, cross-validation techniques, and evaluation metrics. In total, we reviewed 60 articles, 41 of which utilized public datasets, while the remaining employed private data. Our findings suggest that ultrasound-based AI studies for COVID-19 detection have great potential for clinical use, especially for children and pregnant women. Our review also provides a useful summary for future researchers and clinicians who may be interested in the field.

Keywords: review; COVID-19; deep learning; artificial intelligence; lung ultrasound

---

1. Introduction

The World Health Organization (WHO) declared Coronavirus Disease 2019 (COVID-19) a global pandemic in March 2020, and despite preventive measures, the virus has led to over 704 million cases and 7 million deaths worldwide [96]. COVID-19, like other respiratory infections, primarily affects the lungs, especially in individuals with comorbidities such as heart disease and diabetes [38,88]. With the continued rise in cases and the emergence of new variants, medical imaging modalities such as computed tomography (CT), X-ray, and lung ultrasound (LUS) have become increasingly essential for diagnosing and monitoring COVID-19 lung infections [26,76,93].

Medical imaging is undeniably the most important tool for the diagnosis and management of treatments in clinical settings [95]. Despite ultrasound being known to be a noisy imaging modality compared to various other imaging modalities with exceptional image quality (i.e., CT, magnetic resonance imaging, X-ray, etc.) [72], it stands out for being a low-cost, mobile, and, above all, non-ionizing medical imaging technology [99]. Because ultrasound is radiation-safe, it is the preferred imaging modality for children and pregnant women [75] and has been widely used in the detection and severity assessment of COVID-19 for the same patient group [4]. Lung infection due to COVID-19 can be seen and assessed in LUS images.

Typically, three major tasks can be performed on LUS images for COVID-19 patient management: (i) detection of pneumonia infection in the lung (e.g., [13,15,24,33]), (ii) pneumonia type/severity classification (e.g., [18,20,23,32,81]), and (iii) segmentation of infection in the lung (e.g., [98]). There are usually three types of artifacts that can appear in a LUS image, such as A-lines, B-lines, and irregular pleural lines (see Figure 1) [85]. When ultrasound pulses reach the surface of the lung, healthy lungs exhibit horizontal lines parallel to the surface of the transducer, known as A-lines. On the other hand, a lung infected with pneumonia shows irregular pleural lines, as well as brightness in the lung (see Figure 2). On the contrary, COVID-19, a special kind of pneumonia, typically shows discreet vertical reverberation artifacts, known as B lines, which originate from the pleural surface (see Figures 1 and 2) [23,32]. Based on the presence and appearance of these artifacts, pneumonia can be detected and classified as community-acquired pneumonia (CAP) and COVID-19, respectively. Finally, using segmentation, the spread of pneumonia can be estimated, which can be used for the severity scoring criteria for COVID-19 [19,86].
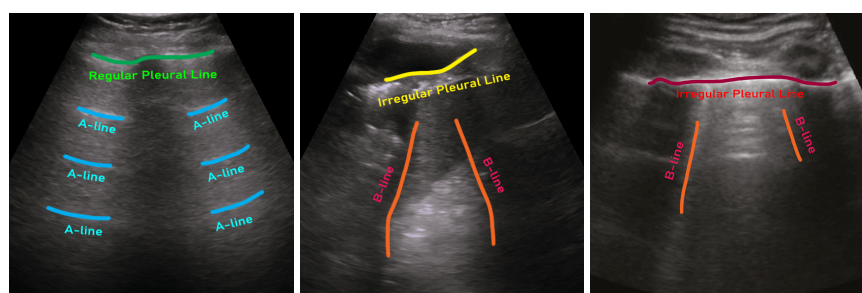

Figure 1. Demonstration of different types of lines that may appear in LUS images. A-lines are marked with blue, B-lines are marked with yellow, and the pleural line is marked with green [13].
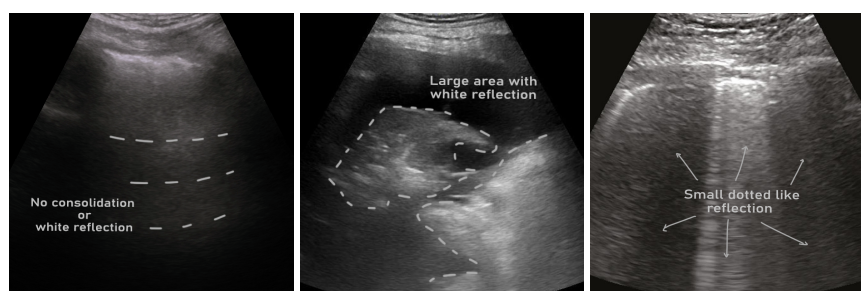
Figure 2. Example ultrasound images of a healthy lung (left), community-acquired pneumonia (CAP)-infected lung (middle), and COVID-19-infected lung (right) [13].

To accelerate the detection and classification of CAP and COVID-19 in clinical settings, artificial intelligence (AI) algorithms [44,66,84,93] have recently been introduced and have shown great promise, which may lessen the burden of expert radiologists/clinicians to detect and assess the severity of pneumonia. Several reviews have examined imaging modalities for COVID-19 detection using AI, but most have focused predominantly on CT and X-ray, with limited coverage of ultrasound-based methods. For instance, Ulhaq et al [90] reviewed 4, Alzubaidi et al [6] 2, and Gudigar et al [34] 3 ultrasound-based papers, with each identifying only 1 unique LUS dataset. Similarly, Huang et al [40] and Siddiqui et al [84] reviewed 1 and 3 papers, respectively, but identified no unique LUS datasets. Reviews by Liu [54] and Mondal et al [66] covered 2 papers each, identifying 1 and 0 unique datasets, respectively. Even the more recent reviews by Wang et al [93], Zhao and Lediju Bell [102], and Gürsoy and Kaya [35] reviewed 35, 9, and 5 papers on ultrasound methods, but identified only 3, 4, and 3 unique datasets, respectively. Lastly, Vafaeezadeh et al [91] reviewed 2 papers that combined LUS with Vision Transformer (ViT) models yet identified no unique LUS datasets. In contrast, our study comprehensively reviews 60 LUS-specific papers, identifying 7 unique public and 21 private COVID-19 LUS datasets. Moreover, our review is organized around these datasets, allowing readers to focus on the methods employed for each specific dataset, which facilitates a more direct comparison of performance across studies. This structure enhances clarity and enables more informed conclusions regarding the effectiveness of different approaches in COVID-19 detection using LUS. In addition, this extensive dataset analysis fills a crucial gap left by previous reviews, which either focused on other modalities or reviewed a limited number of ultrasound papers with insufficient dataset exploration.

Our study aims to address several key research questions (RQ) that have not been thoroughly explored in previous reviews on AI-based COVID-19 detection using LUS. Specifically, we explored the following questions: **RQ1:** What are the most commonly used public and private LUS datasets for COVID-19 detection? **RQ2:** How do the AI-based methods applied to COVID-19 ultrasound detection vary across different datasets, and what are the key performance metrics? **RQ3:** Which ultrasound image preprocessing and augmentation techniques are commonly used to enhance model performance? **RQ4:** What are the existing limitations and challenges in using ultrasound for COVID-19 detection, and how can future research address these challenges?

In this comprehensive review, we address the above RQs to provide a thorough analysis of AI-based COVID-19 detection using LUS. Our contributions are as follows:

1. We identify and catalog the most commonly used public and private LUS datasets for COVID-19 detection.
2. We analyze how different AI-based methods applied to COVID-19 ultrasound detection vary across these datasets and summarize the key performance metrics.
3. We review the ultrasound image preprocessing and augmentation techniques commonly employed to enhance model performance.
4. We identify and discuss the existing limitations and challenges in using ultrasound for COVID-19 detection and offer recommendations for addressing these challenges in future research.

Search Strategy. We searched Google Scholar thoroughly for all scholarly publications: peer-reviewed journal papers, papers published in the proceedings of conferences or workshops, and non-peer-reviewed pre-prints from January 2020 to August 2024. Our search query was `(COVID-19 | corona virus disease) (detect* | predict* | class*) (ultrasound)`. We applied a rigorous selection process to identify relevant articles for our review. The criteria for inclusion were: (1) the full text had to be accessible online or published in reputable journals or conferences indexed in databases such as PubMed, IEEE Xplore, Scopus, or Web of Science; (2) the article must have utilized AI techniques, either conventional machine learning or deep learning, specifically for the detection or analysis of COVID-19 using LUS data; (3) the hypothesis posed by the study had to be supported by robust qualitative and quantitative results; and (4) the article had to meet a minimum quality standard, ensuring no missing abstracts or methodologies, no reference errors, and clear figure legends and axis titles. Similar search strategies and selection criteria have been used in other recent reviews, e.g., [10]. In addition, we took great care to include all relevant studies utilizing LUS and AI for COVID-19 detection and analysis, though a few papers may have been inadvertently overlooked. Our goal, however, was to provide a comprehensive overview of the field. In total, we have reviewed 60 articles in this study.

The remainder of the paper is organized as follows. Details of the datasets, the collection procedure of the ultrasound images, and the image processing techniques before the model building are presented in Section 2. An overview of the architecture of the AI models employed in the studies is presented in Section 3.1. Specific dataset-based studies with their methods and findings are tabulated and discussed in Sections 4, 5, 6 and 6.2. Challenges, limitations, and gaps in the reviewed literature were summarized in Section 9. Discussion and potential future work are described in Section 10. Finally, concluding remarks are presented in Section 11.

2. Lung Ultrasound COVID-19 Datasets

Supervised learning using deep neural networks, a category of AI, has been extensively used for medical imaging applications in recent years [97]. Adequate training of deep models for medical data requires prohibitive amounts of annotated data at the image/pixel/voxel level. Using such deep models on LUS data for COVID-19 detection and analysis is also not an exception. Furthermore, it is also critical to have public access to such datasets, as many research groups lack the clinical setup for data collection. In addition, reproducing a claimed performance by an AI method and possible future improvement greatly relies on access to the exact dataset. However, only a few publicly accessible LUS datasets are available. In this section, we discuss such datasets and their attributes.

2.1. Publicly Accessible LUS COVID-19 Datasets

In Table 1, we list publicly accessible LUS COVID-19 datasets and their associated class labels. We briefly discuss each dataset below:

Point-of-Care Ultrasound (POCUS): Born et al [13,15] published and have been maintaining the POCUS dataset since 2020. This dataset initially contains a total of 261 lung ultrasound recordings by combining 202 videos and 59 still images collected from 216 patients. In this dataset, data from 92, 90, 73, and 6 are associated with COVID-19, healthy control, bacterial pneumonia, and viral pneumonia, respectively. These data were collected using either convex or linear probes. Each film in their dataset also comes with visual pattern-based expert annotation (e.g., B-Lines or consolidations).

Italian COVID-19 Lung Ultrasound Database (ICLUS-DB): Soldati et al [86] published an internationally standardized acquisition protocol and four-level scoring schemes for lung ultrasound (LUS) in March 2020, shortly known as ICLUS-DB. This dataset contains 277 ultrasound videos (consisting of 58,924 frames) of 17 confirmed COVID-19, four suspected COVID-19, and 14 healthy subjects. These data were collected at various clinical centers in Italy using ultrasound scanners using either linear or

convex probes. To evaluate the progress of pathology, this data consortium defined a four-level scoring system ranging from 0 to 3. The presence of continuous pleural-line and horizontal A-lines indicates a healthy lung with a score of 0. Score 1 is tagged for initial abnormality when alterations in the pleural line appear. Score 2 is more severe than one and is associated with small consolidations in the lung. Score 3 is the most severe grade, which is associated with the presence of a larger hyperechogenic area below the pleural surface (i.e., white lung).

COVIDx-US: Ebadi et al [28] published an open-access LUS benchmark dataset gathered from multiple sources in 2021. The dataset was assembled from a variety of sources (e.g., POCUS Atlas, GrepMed, Butterfly Network, and Life in the Fast Lane). This dataset (i.e., version 1.5) contains 242 videos (with 29,651 extracted images) corresponding to 71 COVID-19, 49 CAP, 94 non-pneumonia lung diseases, and 28 healthy classes.

Boston Emergency Department Lung UltraSound (BEDLUS): The BEDLUS dataset consists of 1,419 LUS videos from 113 patients admitted to Brigham and Women's Hospital, MA, between November 2020 and March 2021 with flu-like symptoms [56]. The dataset includes 188,670 video frames, with 50.7% of the videos positively labeled for B-lines, which are indicative of conditions like heart failure and pneumonia. Videos were acquired using a low-frequency transducer across various lung zones, with frames annotated for B-line origins by lung ultrasound experts. The dataset is preprocessed and de-identified, with available annotations and model parameters accessible online.

COVID-19 Simulated and Labeled *In Vivo* Dataset (CSLID): Zhao et al [103] utilized both simulated and in vivo ultrasound data to investigate B-line detection strategies in COVID-19 patients. Simulated data were generated using the MATLAB Ultrasound Toolbox, modeling a convex probe with 192 elements, a 73° field of view, a 4 MHz center frequency, a 10 cm imaging depth, and a 60 MHz sampling frequency. This simulated data underwent standard ultrasound processing to create B-mode images with a 60 dB dynamic range.

Fictional Lumen Dissection Dataset (FLDD): Vinod et al [92] accumulated approximately 7,050 ultrasound images from an ambiguous source(s), evenly distributed among COVID-19 positive cases, normal individuals, and pneumonia patients (2,350 images each). Images had been standardized to a resolution of 512×512 pixels with Red-Green-Blue (RGB) reversion applied. The dataset was processed using gradient mapping. It also included computed Haralick features for both spatial (i.e., Gray-level difference matrix (GLDM), gray-level cooccurrence matrix (GLCM), and Texture) and frequency (i.e., Discrete wavelet transform (DWT), and fast Fourier transform (FFT)) domains. After modification of LUS data, Vinod et al [92] termed their dataset as "Fictional Lumen Dissection Dataset" and made it public, intended for use in image segmentation and analysis tasks.

Lung Ultrasound COVID Phantom Dataset (LUCPD): This dataset [62] consists of 564 phantom ultrasound images, focusing on five key objects and artifacts: Rib, Pleural line, A-line, B-line, and B-line confluence. These images were captured using a commercial lung ultrasound phantom (CAE Healthcare Inc., Blue Phantom COVID-19 Lung Ultrasound Simulator), which is designed to simulate features ranging from healthy to severely damaged lungs. The dataset was created by acquiring B-mode ultrasound videos with clinical systems and then extracting and annotating images using the VGG Image Annotator (VIA) tool. Multiple individuals with varying levels of ultrasound expertise labeled the images for segmentation purposes, ensuring high-quality annotations. This dataset provides a valuable resource for investigating LUS segmentation and learning to identify pathological signs associated with COVID-19.

Table 1. List of publicly accessible LUS COVID-19 datasets.

Sl.	Dataset	Year	Number of Samples	Class Distribution	Note
1	Point-of-Care Ultrasound (POCUS)	2020	(216 patients) 202 videos 59 images	COVID-19 (35%) Bacterial Pneumonia (28%) Viral Pneumonia (2%) Healthy (35%)	Link[1]
2	Italian COVID-19 Lung Ultrasound Database (ICLUS-DB)	2020	(35 patients) 277 videos 58,924 frames	Score 0: Continuous A-line (34%) Score 1: Alteration in A-line (24%) Score 2: Small consolidation (32%) Score 3: Large consolidation (10%)	Link[2]
3	COVIDx-US	2021	242 videos 29,651 images	COVID-19 (29%) CAP (20%) Non-pneumonia diseases (39%) Healthy (12%)	Link[3]
4	Boston Emergency Department Lung UltraSound (BEDLUS)	2022	(113 patients) 1419 videos 188,670 frames	B-lines (50.7%) No B-lines (49.3%)	Link[4]
5	COVID-19 Simulated and Labeled *In Vivo* Dataset (CSLID)	2024	40,000 simulated images 174 external *in vivo* images 958 internal *in vivo* images	A-line, B-line, and Consolidation features (10,000 phantoms per feature)	Link[5]
6	Fictional Lumen Dissection Dataset (FLDD)	2024	7050 ultrasound images	Pneumonia (31%) Normal (34%) COVID (35%)	Link[6]
7	Lung Ultrasound COVID Phantom Dataset (LUCPD)	2024	564 phantom ultrasound images	5 classes of object/artefact: Rib, Pleural line, A-line, B-line, and B-line confluence	Link[7]

2.2. Non-Accessible LUS COVID-19 Private Datasets

In contrast to the publicly accessible datasets described in Section 2.1, some studies used private datasets, and some of these datasets are mentioned as available on request. However, these data sets have variations in terms of patient origin, hospital location, and data collection protocols. We list these datasets in Table 2 with the number of available samples and associated labels/classes. Below, we also briefly summarize the imaging protocols and types of transducers used in those datasets.

Regardless of the variation of ultrasound scanners, scanning areas on skin targeting the lung are typically similar across datasets. Durrani et al [27] considered six distinctive scanning regions in their study. Panicker et al [71] adopted the scan protocol of Soldati et al [86] and also aimed at six acquisition points for data extraction. Quentin Muller et al [77] scanned on ten thoracic sites in their study. Although video of the costophrenic region was excluded in [8], most studies followed a twelve-zone scanning protocol for the data acquisition process [17,21,39,52,94,98]. Furthermore, Mento et al [63] used fourteen scanning areas, following the scan protocol of by Soldati et al [86]. Another study [80] followed the scan protocol by Mento et al [64] and Perrone et al [74].

Variations in transducer types and frequency were also observed in the studies. For example, some studies used low-frequency (1–5 MHz) curved array [17,21,39,71] and phased array [8,27] transducers. On the other hand, Roshankhah et al [80] used both linear and convex transducers in multi-sites with a wide range of center frequencies. Similarly, La Salvia et al [52] used both linear and convex transducers with a frequency of 5 and 12 MHz, respectively, and Mento et al [63] used 3.5 to 6.6 MHz in their study. [51] used a dataset that includes 56 subjects, with 41 COVID-19 patients and 15 controls. Lung point-of-care ultrasound (POCUS) and CT scans were performed, analyzing 397 lung zones in patients and 180 in controls, with some zones excluded due to positioning limitations. The retrospective study conducted by [83] involved data from multiple institutions and public databases, and a total of 52 ultrasound scans were performed using various scanners. Different probes were used depending on the imaging requirements, with findings verified by radiologists experienced in LUS related to COVID-19.

Faita et al [31] collected 2,067 LUS videos from 135 COVID-19-positive patients across two cohorts: 1,564 videos from 104 patients in 2020 (cohort 1) and 503 videos from 31 patients in 2022 (cohort 2). The videos were annotated with a clinically validated severity score by expert sonographers, ranging from 0 to 3 based on symptom severity. Kimura et al [50] used a private dataset that included video images from 69 patients with suspected congestive heart failure (CHF), captured using Lumify devices with a 3 MHz cardiac transducer. Li et al [53] used the dataset, including 1,447 frames, categorized into 113 mild, 21 moderate, eight severe, and 25 critical cases. The dataset typically contains eight images per examination, with imaging settings tailored to each patient.

Table 2. List of private (publicly non-accessible) COVID-19 ultrasound datasets. Acronyms- N: number of samples, Tr: training, Va: validation, and Te: test.

Sl.	Dataset	Year	N	Tr/Va/Te	Classes	Note
1	London Health Sciences Centre's two tertiary hospitals (Canada) [8]	2020	(243 patients); 600 videos; 121,381 frames	~80/20	COVID, Non-COVID, Hydrostatic Pulmonary Edema	-
2	ULTRACOV (Ultrasound in Coronavirus disease) [17]	2022	(28 COVID-19 patients) 3 sec video each	-	A-Lines, B-Lines, consolidations, and pleural effusions	Available upon request
3	Huoshenshan Hospital (Wuhan, China) [21]	2021	(31 patients) 1,527 images	-	Normal, septal syndrome, interstitial-alveolar syndrome, white lung	Source Link[h]
4	Royal Melbourne Hospital (Australia) [27]	2022	(9 patients) 27 videos; 3,827 frames	-	Normal, consolidation/collapse	Available upon request
5	Ultrasound lung data [28]	2021	(300 patients) 1530 videos; 287,549 frames	80/20	A-line artifacts, B-line artifacts, presence of consolidation/pleural effusion	-
6	Huoshenshan Hospital (Wuhan, China) [39]	2022	(31 patients); 2,062 images	-	Normal, septal syndrome, interstitial-alveolar syndrome, white lung	Source Link[i]
7	Fondazione IRCCS Policlinico San Matteo's Emergency Department (Pavia, Italy) [52]	2021	(450 patients) 2,908 frames	75/15/10	A-lines with two B-lines, slightly irregular pleural line, artifacts in 50% of the pleura, damaged pleural line, visible consolidated areas, damaged pleura/irregular tissue	-
8	Third People's Hospital of Shenzhen (China) [55]	2020	(71 COVID-19 patients) 678 videos; 6,836 images	-	A-line, B-line, pleural lesion, pleural effusion	-
9	Fondazione Policlinico Universitario Agostino Gemelli (Rome, Italy), Fondazione Policlinico San Matteo (Pavia, Italy) [63]	2021	(82 patients) 1,488 videos; 314,879 frames	-	4 severity levels [86]	-
10	CHUV (Lausanne, Switzerland) [77]	2020	(193 patients) 1,265 videos; 3,455 images	80/20	True (experts' approval), False (experts' disapproval)	-
11	Various online sources [68]	2022	792 images	-	COVID-19, healthy	-
12	Spain, India [71]	2021	(10 subjects) 400 videos, 5,000 images	-	A-lines, lack of A-lines, the appearance of B-lines, the confluent appearance of B-lines, the appearance of C-lines	Available upon request
13	Private clinics (Lima, Peru) [79]	2021	1,500 images	-	Healthy, COVID-19	Available upon request

---

[h]   https://bio-hsi.ecnu.edu.cn/
[i]   https://bio-hsi.ecnu.edu.cn/

Table 2 (Continued): List of private (non-accessible publicly) COVID-19 ultrasound datasets. Acronyms- N: number of samples, Tr: training, Va: validation, and Te: test.

Sl.	Dataset	Year	N	Tr/Va/Te	Classes	Note
14	BresciaMed (Brescia, Italy), Valle del Serchio General Hospital (Lucca, Italy), Fondazione Poli-clinico Universitario A. Gemelli IRCCS (Rome, Italy), Fondazione Poli-clinico Universitario San Matteo IRCCS (Pavia, Italy), and Tione General Hospital (Tione, Italy) [80]	2021	(32 patients) 203 videos; 1,863 frames	90/10	Healthy, indentation of pleural line, discontinuity of the pleural line, white lung	-
15	Beijing Ditan Hospital (Beijing, China) [94]	2021	(27 COVID-19 patients) 13 moderate, 7 severe, 7 critical	-	Severe, non-severe	-
16	Cancer Center of Union Hospital, West of Union Hospital, Jianghan Cabin Hospital, Jingkai Cabin Hospital, Leishenshan Hospital [98]	2021	(313 COVID-19 patients) 10-second video from each	-	Normal, presence of 3-5 B-lines, ≥6 B-lines or irregular pleura line, fused B-lines or thickening pleura line, consolidation	-
17	Juntendo University Graduate School of Medicine (Tokyo, Japan) [51]	2023	(56 subjects) 577 lung zones	-	Count of B-lines in each zone	-
18	Collection of Data from Various Unspecified Institutions and Public Databases [83]	2023	52 ultrasound examinations	-		-
19	Italian National Research Council and University of Pisa (Italy) [31]	2024	104 patients 1564 LUS videos	-	3 severity levels	-
20	Mayo Clinic, Rochester, Minnesota [50]	2024	110 video clips 69 patients	-	Normal, Mild-moderately abnormal, Severely abnormal	-
21	Beijing Ditan Hospital (Beijing, China) [53]	2024	(152 patients) 1,447 frames	-	COVID severity levels: Mild, Moderate, Severe, Critical	-

Figure 3 presents a pie-chart showing the distribution of datasets of the reviewed articles in this study, with POCUS accounting for 21 (35%), Non-open Access for 19 (32%), ICLUS-DB for 9 (15%), COVIDx-US for 7 (11%), and other sparsely used public datasets for 4 (7%).
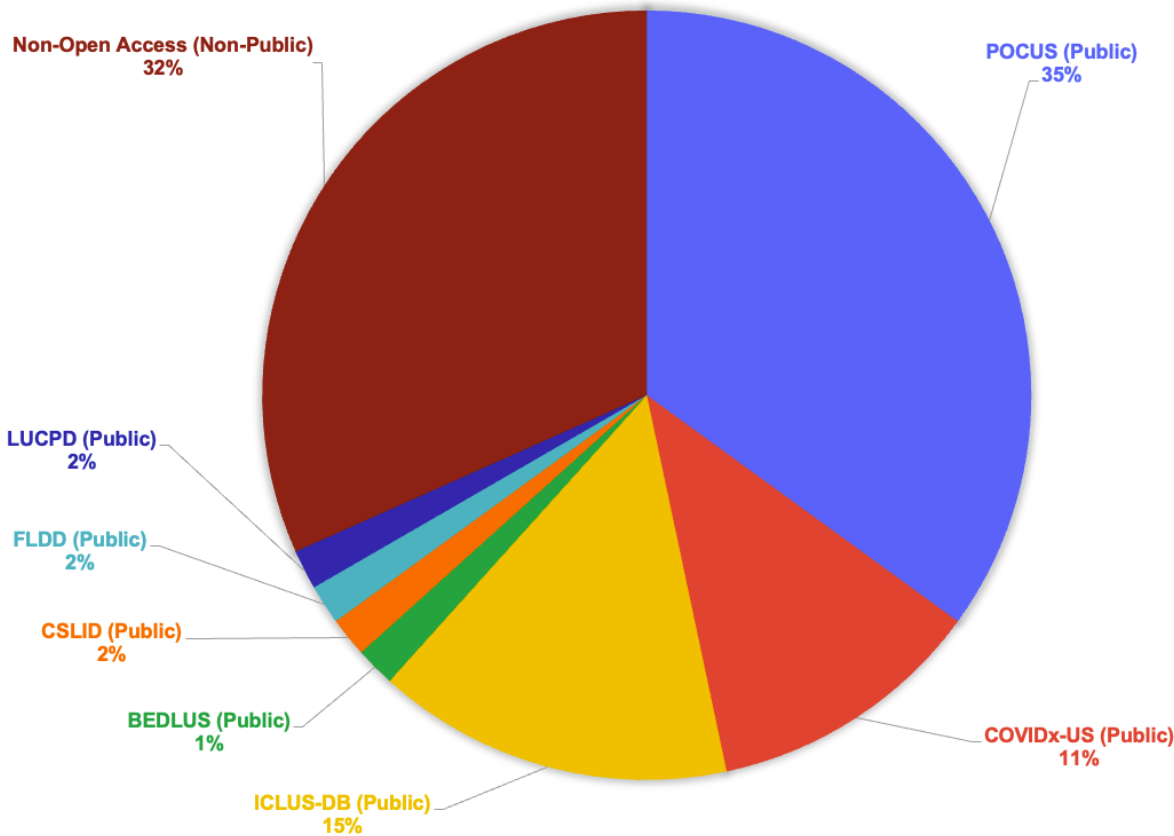

Figure 3. A pie-chart showing the percentage of reviewed articles in this study per LUS datasets.

2.3. Data Pre-Processing and Augmentation

Various image processing techniques are typically used before feeding the data to AI models. Image processing techniques include, but are not limited to, curve-to-linear conversion, image resizing, intensity normalization, standardization, augmentation, etc. In this section, we briefly discuss different image pre-processing techniques used in the reviewed articles.

2.3.1. Curve-to-Linear/Linear-to-Curve Conversion

Acquired ultrasound videos and images using convex transducers are typically fan-shaped (i.e., narrower close to the probe surface while wider at depth). In contrast, ultrasound videos and images that use linear transducers are usually rectangular. Thus, harmonizing images acquired by convex and linear transducers requires the conversion of fan-shaped images to rectangular images and vice versa. Therefore, various automatic built-in conversion techniques in the scanner, as well as external user-defined interpolation techniques [16], are typically used for this conversion task, and ultrasound-based COVID-19 AI studies are not an exception, e.g., [18,53,101].

2.3.2. Image Resizing

Image resizing is the most common image pre-processing technique used for AI model training. Typically, ultrasound images come with various resolutions in terms of pixel count. On the other hand, AI models, especially deep learning models, typically require all input images to be of equal dimension. In addition, the larger input image dimension and the number of channels cause a higher computational overhead in the AI model optimization process. Therefore, AI studies often resize input images to a widely used common dimension across datasets. Most of the reviewed articles in this paper, for example, [29,45,47,59,68,73,77,79,87], etc., also used the common image dimension of 224×224 pixels as well-known computer vision deep learning models are typically designed to intake images of 224×224 pixels. However, other image dimensions are also found for ultrasound COVID-19 studies. For example, Karar et al [46] resized all ultrasound images to 28×28 pixels to avoid a higher computational overhead. In addition, Nehary et al [69], Mateu et al [61], Lucassen et al [56], Durrani et al [27], Vinod et al [92], Muhammad and Hossain [67], and Gare et al [33] resized their ultrasound images to 128×128, 254×254, 384×256, 806×550, 512×512, and 624×464 pixels, respectively.

2.3.3. Intensity Normalization

Intensity normalization is another common image pre-processing technique used in AI studies. This process ensures a common intensity range across images and datasets. In most cases, all image data are converted to a common intensity range of [0, 1], or [0, 255] [73], followed by mean subtraction and division by standard deviation [1,53,56,67,77,80,89,94].

2.3.4. Image Augmentation

Image augmentation is a widely used technique in AI studies, which is used to increase the amount of training data and the variation and diversity in the appearance of an image. one of the most prevalent steps that have been executed in most of the studies. Various conventional (as in Hussain et al [43]) and learning-based data augmentation [65] techniques are present in the literature. Conventional image augmentation techniques such as image cropping, random rotation, horizontal and vertical flipping, histogram equalization, random image shifting, zooming in and out, and/or a combination of these operations, etc., are more prevalent in AI studies, and articles in this review (e.g., Adedigba and Adeshina [1], Arntfield et al [8], Born et al [13], Faita et al [31], Gare et al [33], Howell et al [37], Khan et al [49], La Salvia et al [52], Muhammad and Hossain [67], Nabalamba [68], Rojas-Azabache et al [79], Roy et al [81], Song et al [87], Zhao et al [103]) mostly adopted this type of augmentation.

2.3.5. Other Image Processing Techniques

Apart from the common image pre-processing techniques discussed above, other processes are often used in ultrasound AI studies. Ultrasound images are known to be a noisy modality [70]. Therefore, ultrasound-based studies often use noise reduction filters for pre-processing of images [23], such as circular averaging filter [7], median filter [41], non-linear diffusion filter [42], contrast-limited adaptive histogram equalization (CLAHE) [82], etc.

Ebadi et al [28] performed several pre-processing operations to make resulting ultrasound images in COVIDx-US[10] dataset easily usable to AI models. They cropped video frames into rectangular windows to remove the background or visible text from the image periphery. Any video frame with a moving pointer on it was also ignored when frames were extracted to use as images.

Other image preprocessing techniques used by reviewed articles in this study include image blurring [49,103], elastic warping [49], and variable time-gain compensation [37].

3. AI in LUS COVID-19 Studies

The accuracy of identifying COVID-19 infection and assessing its severity is based primarily on the expertise of clinicians, which is often difficult and time-consuming. To overcome this limitation, AI approaches have been widely used in recent years. AI approaches used in COVID-19 ultrasound studies can be categorized into conventional machine learning (CML) and deep learning (DL) approaches. CML approaches (e.g., support vector machine (SVM), linear regression, etc.) typically require hand-engineering of features, which are often difficult to define optimally [43]. Overcoming this limitation, DL using convolutional neural networks (CNN) has exploded in popularity throughout the last decade. Various CNN architectures have been widely used on natural image and medical image-based classification and segmentation tasks. However, medical imaging data are often very difficult to collect, which results in a small training data cohort. To overcome this limitation, DL on medical imaging often leverages the transfer learning strategy, where the deep model is pre-trained on a much larger natural image dataset and then finetuned on the target smaller medical data. This transfer learning strategy is also used in many articles (e.g., Al-Jumaili et al [2], Barros et al [12], Diaz-Escobar et al [25], Nabalamba [68], Rojas-Azabache et al [79]) we reviewed in this study. In addition, many studies in this review (e.g., Born et al [13], Diaz-Escobar et al [25]) used cross-validation techniques to avoid overfitting.

3.1. AI Models

In Table 3, we list all the articles we reviewed in this study and the corresponding AI methods used by those articles. We also mark in the table whether a study used CML, DL, or both. We see in the table that only three studies used CML approaches (see rows 12, 40, and 44 of Table 3), and five studies combined CML and DL (see rows 2, 14, 53, 55, and 56 of Table 3). Except for these studies, all other studies we reviewed used DL approaches. This tendency to prefer DL approaches over CML approaches is motivated by the fact that DL models are capable of learning optimal feature representation by themselves without requiring manual intervention and the availability of more complex and powerful computation facilities. In Figure 4, we organized all the reviewed articles in terms of the type of AI model and configuration. We also describe different types of AI models used by state-of-the-art LUS COVID-19 studies in the following sections.

---

[10]   https://github.com/nrc-cnrc/COVID-US

Table 3. A list of the articles reviewed in this study and the corresponding AI methods used by those articles. Acronyms- Sl.: serial, CML: conventional machine learning, DL: deep learning, RNN: recurrent neural network, SVM: support vector machine, LSTM: long short-term memory, STN: spatial transformer network, GAN: generative adversarial network, ViT: vision transformer, gMLP: multi-layer perceptron with gating.

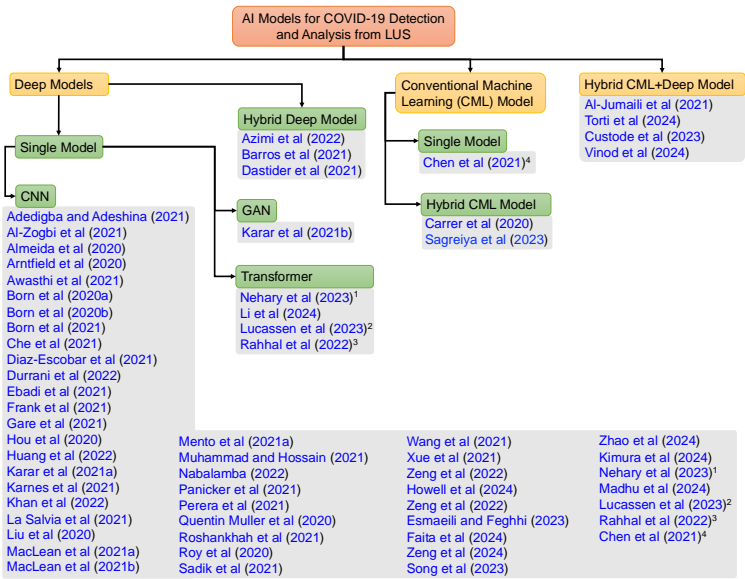| Sl.	Studies	AI Methods	CML	DL
1	Adedigba and Adeshina [1]	SqueezeNet, MobileNetV2	✗	✓
2	Al-Jumaili et al [2]	ResNet-18, RestNet-50, NASNetMobile, GoogleNet, SVM	✓	✓
3	Al-Zogbi et al [3]	DenseNet	✗	✓
4	Almeida et al [5]	MobileNet	✗	✓
5	Arntfield et al [8]	Xception	✗	✓
6	Awasthi et al [9]	MiniCOVIDNet	✗	✓
7	Azimi et al [11]	InceptionV3, RNN	✗	✓
8	Barros et al [12]	Xception-LSTM	✗	✓
9	Born et al [13]	VGG-16	✗	✓
10	Born et al [14]	VGG-16	✗	✓
11	Born et al [15]	VGG-16	✗	✓
12	Carrer et al [18]	Hidden Markov Model, Viterbi Algorithm, SVM	✓	✗
13	Che et al [20]	Multi-scale Residual CNN	✗	✓
14	Chen et al [21]	2-layer NN, SVM, Decision Tree	✓	✓
15	Diaz-Escobar et al [25]	InceptionV3, VGG-19, ResNet-50, Xception	✗	✓
16	Dastider et al [23]	Autoencoder-based Hybrid CNN-LSTM	✗	✓
17	Durrani et al [27]	Reg-STN	✗	✓
18	Ebadi et al [29]	Kinetics-I3D	✗	✓
19	Frank et al [32]	ResNet-18, MobileNetV2, DeepLabV3++	✗	✓
20	Gare et al [33]	Reverse Transfer Learning on U-Net	✗	✓
21	Hou et al [36]	Saab transform-based SSL, CNN	✗	✓
22	Huang et al [39]	Non-local channel attention ResNet	✗	✓
23	Karar et al [45]	MobileNet, ShuffleNet, MENet, MnasNet	✗	✓
24	Karar et al [46]	A semi-supervised GAN, a modified AC-GAN	✗	✓
25	Karnes et al [47]	Few-shot learning using MobileNet	✗	✓
26	Khan et al [48]	CNN	✗	✓
27	La Salvia et al [52]	ResNet-18, ResNet-50	✗	✓
28	Liu et al [55]	Multi-symptom multi-label (MSML) network	✗	✓
29	MacLean et al [57]	COVID-Net US	✗	✓
30	MacLean et al [58]	ResNet	✗	✓
31	Mento et al [63]	STN, U-Net, DeepLabV3+	✗	✓
32	Muhammad and Hossain [67]	CNN	✗	✓
33	Nabalamba [68]	VGG-16, VGG-19, ResNet	✗	✓
34	Panicker et al [71]	LUSNet (a U-Net like network for ultrasound images)	✗	✓
35	Perera et al [73]	ViT	✗	✓
36	Quentin Muller et al [77]	ResNet-18	✗	✓
37	Roshankhah et al [80]	U-Net	✗	✓
38	Roy et al [81]	STN, U-Net, U-Net++, DeepLabV3, Model Genesis	✗	✓
39	Sadik et al [82]	DenseNet-201, ResNet-152V2, Xception, VGG-19, NasNetMobile	✗	✓
40	Wang et al [94]	SVM	✓	✗
41	Xue et al [98]	U-Net	✗	✓
42	Zeng et al [100]	COVID-Net US-X	✗	✓
43	Kuroda et al [51]	AI-POCUS (Model specifics are not disclosed)	-	-
44	Sagreiya et al [83]	Calculated Lung Ultrasound (CLU)	✓	✗
45	Khan et al [49]	ResNet-18, ResNet-50, ResNet-101, DenseNet-121, DenseNet-201, InceptionV3, RegNetX, EfficientNetB7	✗	✓
46	Howell et al [37]	Lightweight U-Net	✗	✓
47	Esmaeili and Feghhi [30]	Uniform Local Binary Pattern on Five Intersecting Planes and CNN (ULBPFP-Net)	✗	✓
48	Faita et al [31]	Inflated 3D Convolutional Network (I3D)	✗	✓
49	Zeng et al [101]	COVID-Net L2C-ULTRA	✗	✓
50	Song et al [87]	COVID-Net USPro	✗	✓
51	Zhao et al [103]	U-Net	✗	✓
52	Kimura et al [50]	CNN	✗	✓
53	Torti et al [89]	ResNet-50+K-means	✓	✓
54	Nehary et al [69]	VGG16, ViT	✗	✓
55	Custode et al [22]	STN, U-Net+DeepLabV3+, Decision Tree	✓	✓
56	Vinod et al [92]	GAN+Random Forest	✓	✓
57	Madhu et al [59]	Xception Convolutional Neural Network (XCovNet)	✗	✓
58	Li et al [53]	Knowledge Fusion with Latent Representation (KFLR) Transformer	✗	✓
59	Lucassen et al [56]	ResNet3D-18, ResNet(2+1)D-18, 3D U-Net, DenseNet-121, EfficientNetB0, ViT, DeepLabV3+	✗	✓
60	Rahhal et al [78]	EfficientNetB2, ViT, gMLP	✗	✓

Figure 4. Organization of reviewed articles in terms of the AI model types and configurations. Hybrid models represent those studies that used two or more types of machine learning strategies together as part of a common predictive model. Four studies used both CML and DL strategies, but they were not part of a common predictive model. Those articles are shown twice under different categories and indicated with common superscripts, 1-4.

3.1.1. Convolutional Neural Networks (CNN)

SqueezeNet, MobileNetV2, VGG-16/19, NasNetMobile, DenseNet-121/201, ResNet-18/50/101/152V2, ResNet(2+1)D-18, ResNet3D-18, InceptionV3, GoogleNet, EfficientNetB0/B2/B7, XCovNet: These CNNs are DL models specifically designed for image processing tasks. They typically consist of convolutional layers that extract features from input images and pooling layers that downsample the feature maps. These models typically include building blocks such as convolutional filters, activation functions (e.g., ReLU), pooling layers (e.g., MaxPooling), and fully connected layers for classification. Convolution layers apply a set of learnable filters to the input data. Each filter performs a convolution operation, which involves sliding across the input and computing dot products with local regions. This process helps extract important visual features, such as edges, textures, and patterns. Pooling layers, on the other hand, downsample the spatial dimensions of the feature maps generated by convolutional layers. They reduce the computational complexity and extract the most important information by summarizing groups of values in the feature maps. The most common type of pooling is max pooling, which selects the maximum value within each pooling region. Furthermore, fully connected layers connect every neuron from the previous layer to the subsequent layer, similar to traditional neural networks. Fully connected layers are responsible for the final classification or regression task, combining the extracted features to make predictions. Most of the articles we reviewed in this study used different types of CNNs, for example, SqueezeNet [1], MobileNetV2 [1], VGG-16 [13,14,68,69], VGG-19 [68,82], ResNet-18 [2,49,52,77], ResNet-50 [2,49,52,89], ResNet-101 [49], ResNet-152V2 [82], DenseNet-121 [49,56], DenseNet-201 [49,82], InceptionV3 [11,49], GoogleNet [2], EfficientNetB0 [56], EfficientNetB2 [78], EfficientNetB7 [49], XCovNet [59] and NasNetMobile [2,82].

ULBPFP-Net: It is a DL model that combines texture analysis with CNNs for image classification. It uses the Uniform Local Binary Pattern on Five Intersecting Planes (ULBPFP), a texture descriptor that captures local patterns from multiple image planes, to extract robust features. These features are then processed by a CNN, which learns complex patterns and representations for accurate classification. The integration of ULBP with CNNs enhances the ability of a model ability to capture both texture and spatial information, improving performance in tasks like medical imaging. Esmaeili and Feghhi [30] used this approach in their study.

Inflated 3D Convolutional Network (I3D): I3D is a deep learning model designed for video analysis, extending 2D convolutional networks by inflating 2D filters into 3D. This allows the model to capture both spatial and temporal features from video data, making it highly effective in processing and classifying videos, particularly for tasks like action recognition. A study by Faita et al [31] utilized the I3D approach in analyzing ultrasound COVID-19 data. *Kinetics-I3D*, on the other hand, builds upon the I3D model by combining it with the Kinetics dataset, a large-scale video dataset specifically designed for action recognition. Kinetics-I3D leverages pre-training on this dataset, enabling the model to learn rich spatiotemporal representations from millions of labeled video clips, effectively capturing motion and temporal dependencies. One study [29] we reviewed employed Kinetics-I3D in their study.

3.1.2. Recurrent Neural Networks (RNN)

RNNs are a type of neural network that can process sequential data by capturing temporal dependencies. They are commonly used for tasks involving sequential inputs or outputs, such as natural language processing and time series analysis. RNNs have recurrent connections that allow information to flow from one time step to the next. This enables the network to maintain a memory of previous inputs and utilize that information to make predictions or analyze the current input. At each time step, an RNN produces an output based on the current input and the hidden state from the previous time step. The hidden state serves as the network's memory, storing information about previous inputs. It is updated and passed along to the next time step, allowing the network to learn and capture long-term dependencies in the sequence. RNNs can be "unfolded" in time, creating a series of interconnected layers that correspond to each time step. This unfolding helps visualize the flow of information through the network and enables the application of backpropagation through time, a training algorithm that adjusts the network's weights based on the sequence of inputs and desired outputs. One of the articles we reviewed in this study used RNN [11].

3.1.3. COVID-Net

COVID-Net US, COVID-Net US-X, COVID-Net USPro, COVID-Net L2C-ULTRA: These architectures are specifically developed for the detection and diagnosis of COVID-19 from medical imaging, particularly chest X-ray images. COVID-Net US is a CNN architecture designed for the classification of chest X-ray images to detect COVID-19 cases. It has been trained on a large dataset of X-ray images and is capable of distinguishing COVID-19 from other respiratory conditions. The architecture of COVID-Net US includes convolutional layers for feature extraction, pooling layers for downsampling, and fully connected layers for classification. COVID-Net US-X is an extended version of COVID-Net US with improved performance and capabilities. It incorporates advancements such as additional layers, refined architecture, or enhanced training techniques to enhance the accuracy and reliability of COVID-19 detection from chest X-ray images. Two articles we reviewed in this study used COVID-Net US [57] and COVID-Net US-X [100]. On the other hand, COVID-Net USPro is a few-shot learning model designed to classify unlabelled data by comparing it to labeled examples. It operates by creating a prototype representation for each class from labeled data and assigns unlabelled data to the class with the closest prototype in an embedding space, using a distance metric to measure similarity. This model is trained in an episodic setting, allowing it to handle limited labeled data effectively. One of the studies [87] in our review used COVID-Net USPro. Another study [101] in our review used COVID-Net L2C-ULTRA, which is a data augmentation learning method designed to address data scarcity and heterogeneity in POCUS images. It enhances the diversity of training data by applying random projective and piecewise affine transformations, helping linear probe images resemble those from convex probes. This technique improves the generalization of deep models by exposing them to more diverse and visually consistent data.

MiniCOVIDNet: It is a compact and efficient neural network architecture designed for COVID-19 detection from chest X-ray images. It is specifically developed to provide a smaller model that can be deployed on resource-constrained devices or in scenarios where computational efficiency

is important. The architecture of MiniCOVIDNet typically includes convolutional layers, pooling layers, and fully connected layers, aiming to accurately classify X-ray images as COVID-19 positive or negative while minimizing computational requirements. One of the articles we reviewed in this study used MiniCOVIDNet [9].

3.1.4. Long Short-Term Memory (LSTM)

LSTM is a type of RNN architecture that addresses the vanishing gradient problem of traditional RNNs and is capable of capturing long-term dependencies in sequential data. LSTMs are widely used in various tasks involving sequential data, such as natural language processing, speech recognition, and time series analysis. The key feature of LSTM networks is their memory cell, which allows them to retain information over long sequences and selectively forget or update that information. LSTMs achieve this through a set of gates, including an input gate, a forget gate, and an output gate. These gates regulate the flow of information and enable the LSTM to remember or forget specific information based on the context. One of the articles we reviewed in this study used LSTM [23].

Xception-LSTM: Xception-LSTM refers to a specific model architecture that combines the Xception CNN with an LSTM layer. Xception is a deep CNN architecture that was proposed as an extension of the Inception architecture. It introduces a novel concept called "depthwise separable convolutions" to reduce the number of parameters and computations required by traditional convolutions. Another article we reviewed in this study used Xception-LSTM [12].

3.1.5. Hidden Markov Model (HMM)

HMM and Viterbi Algorithm are both fundamental concepts in the field of probabilistic modeling and sequential data analysis. One of the articles we reviewed in this study used both HMM and Viterbi Algorithm [18].

Hidden Markov Model: An HMM is a statistical model that represents a system with unobservable (hidden) states and observable outputs. It is a generative model that assumes the underlying system can be modeled as a Markov process, where the current state depends only on the previous state. However, the actual state is not directly observable; instead, it emits observable symbols or outputs. HMMs have been widely used in various applications such as speech recognition, natural language processing, bioinformatics, and pattern recognition.

Viterbi Algorithm: The Viterbi Algorithm, on the other hand, is an efficient dynamic programming algorithm used to find the most likely sequence of hidden states in a Hidden Markov Model. Given a sequence of observations, the Viterbi Algorithm computes the optimal sequence of hidden states that maximizes the probability of the observations. It takes into account both the transition probabilities between states and the emission probabilities of observations from the states. The algorithm iteratively computes the most likely path by considering the accumulated probabilities at each time step, resulting in the most probable sequence of hidden states.

3.1.6. Generative Adversarial Networks (GAN)

GAN is a class of machine learning models that consists of two neural networks, namely the generator and the discriminator, which are trained together in a competitive setting. The generator network takes random noise as input and generates synthetic samples, such as images, based on that noise. The objective of the generator is to generate samples that resemble real data as closely as possible. On the other hand, the discriminator network takes both real samples from the dataset and synthetic samples from the generator as input and aims to classify them correctly as real or fake. The discriminator's objective is to distinguish between real and generated samples accurately. During training, the generator and discriminator are trained in alternating steps. The generator tries to fool the discriminator by generating realistic samples, while the discriminator aims to improve its ability to distinguish real from fake samples. This back-and-forth training process creates a competitive dynamic

where the generator improves its ability to generate realistic samples, and the discriminator becomes more adept at discriminating between real and fake samples.

AC-GAN: Auxiliary Classifier GAN (AC-GAN) is an extension of the GAN framework that includes an auxiliary classifier in addition to the discriminator. The auxiliary classifier is a separate network that is trained to predict additional class labels or attributes associated with the generated samples. This helps in controlling the generated samples to have specific attributes or belong to specific classes. The addition of the auxiliary classifier in AC-GAN allows for more control over the generated samples and enables the generation of samples conditioned on specific attributes or classes. It has been used in various applications, including image synthesis, text-to-image generation, and image-to-image translation, where the generation process can be guided by specific attributes or class labels. One of the articles we reviewed in this study used a semi-supervised GAN and AC-GAN [46].

3.1.7. Transformer

Transformer is a DL architecture designed for sequence modeling tasks like natural language processing. It relies on self-attention mechanisms to capture relationships between all elements in a sequence, allowing it to process data in parallel and learn long-range dependencies more efficiently than recurrent models.

Vision Transformer (ViT): ViT applies the Transformer architecture to image data by dividing an image into patches, treating each patch as a token similar to words in a sentence. It processes these patches through self-attention mechanisms, allowing the model to capture global image features and achieve strong performance in image classification tasks. Some studies [56,69,73,78] we reviewed in this work used ViT for LUS COVID-19 data analysis and classification.

Knowledge Fusion with Latent Representation (KFLR) Transformer: KFLR Transformer is a specialized Transformer model that integrates multiple sources of information by learning a shared latent representation. This fusion of knowledge from different domains or modalities improves the ability of a model to handle complex tasks, as it combines diverse insights while maintaining high-quality latent feature representations. One study in our review used KFLR Transformer to predict the severity of COVID-19 from LUS [92].

3.1.8. Spatial Transformer Network (STN)

STN is a type of neural network module that can be integrated into deep learning architectures to enable the spatial transformation of input data. The purpose of the STN network is to learn spatial transformations, such as rotations, translations, scaling, and cropping, that can be applied to input images or feature maps. The key idea behind the STN network is to introduce a spatial transformer module that can learn to automatically align and transform input data to improve the overall performance of the model. The module consists of three main components- (i) Localization Network: The localization network takes the input data and learns to predict the parameters of the spatial transformation. It typically consists of convolutional and fully connected layers that extract features and output the transformation parameters, such as translation, rotation, and scaling. (ii) Grid Generator: The grid generator takes the predicted transformation parameters from the localization network and generates a set of sampling grid points. These grid points define how the input data should be transformed to align with the desired output. (iii) Sampler: The sampler takes the input data and the generated grid points and performs the spatial transformation. It applies interpolation techniques, such as bilinear interpolation, to sample the input data at the grid points and produce the transformed output. By incorporating the STN network into a larger neural network architecture, the model can learn to automatically adjust and align the input data to improve performance. The STN module can be trained end-to-end with the rest of the network using backpropagation, allowing the model to learn the appropriate spatial transformations for the given task. Several articles we reviewed in this study used Transformer or STN [63,73,81].

Reg-STN: Reg-STN stands for Regression Spatial Transformer Network. It is an extension of the STN that incorporates regression-based localization instead of classification-based localization. One of the articles we reviewed in this study used Reg-STN [27].

3.1.9. U-Net

U-Net is a convolutional neural network architecture that was specifically designed for biomedical image segmentation but has since been applied to various other domains. It consists of an encoder-decoder structure with skip connections. The encoder part gradually reduces the spatial dimensions while capturing hierarchical features, and the decoder part upsamples the feature maps and recovers the spatial resolution. The skip connections help preserve fine-grained details by concatenating feature maps from the encoder to the corresponding decoder layers. U-Net has been widely used for tasks such as medical image segmentation, cell segmentation, and more. Several articles we reviewed in this study used U-Net[33,63,80,81,98].

U-Net++: U-Net++ is an extension of the U-Net architecture that aims to further enhance the segmentation performance. It introduces a nested and densely connected skip pathway structure. In U-Net++, each encoder block is connected to all corresponding decoder blocks through skip connections, creating a more extensive and interconnected network. This architecture allows for better information flow and feature reuse across different scales, leading to improved segmentation accuracy and boundary delineation. One of the articles we reviewed in this study used Reg-UNet++ [81].

LUSNet: LUSNet (Lung Ultrasound Net) is a specific implementation of the U-Net-like network architecture designed for lung ultrasound image segmentation. It incorporates the U-Net framework with modifications tailored for lung ultrasound images. LUSNet leverages the inherent characteristics of lung ultrasound images, such as the presence of artifacts, pleural lines, and specific structures like A-lines and B-lines, to perform accurate segmentation. By adopting the U-Net architecture to the unique properties of lung ultrasound images, LUSNet aims to provide reliable segmentation for various lung-related applications, including disease diagnosis and monitoring. One of the articles we reviewed in this study used LUSNet [71].

3.1.10. Few-Shot Learning

Few-shot learning is a machine learning paradigm that addresses the problem of learning from limited labeled data. In traditional machine learning approaches, a large amount of labeled data is typically required to train a model effectively. However, in real-world scenarios, collecting and annotating large datasets can be time-consuming, expensive, or impractical. Few-shot learning aims to overcome this limitation by enabling models to learn new concepts or tasks with only a few labeled examples. It focuses on the ability of a model to generalize and adapt to new classes or tasks based on a small amount of labeled data, often referred to as the "support set." The key idea in few-shot learning is to leverage prior knowledge or information learned from related tasks or classes to facilitate learning on new tasks or classes with limited examples. This is achieved through various techniques such as meta-learning, where the model learns to quickly adapt to new tasks based on its previous experience, or by using generative models to synthesize additional training examples. One of the articles we reviewed in this study used Few-shot learning [47].

3.1.11. Transfer Learning

Transfer learning is a machine learning technique that involves leveraging knowledge learned from one task or domain to improve performance on another related task or domain. In transfer learning, a pre-trained model that has been trained on a large dataset and a related task is used as a starting point for a new task. The idea is that the pre-trained model has learned general features and representations that can be useful for the new task, even if the specific classes or labels are different. By using transfer learning, the model can benefit from the knowledge and representations learned

from the large pre-training dataset, which can save training time and improve performance, especially when the target dataset is limited or the target task is challenging.

Reverse Transfer Learning: Reverse transfer learning, on the other hand, is a less commonly used term and refers to the process of transferring knowledge or models from a target domain or task back to the source domain or task. It involves utilizing the information or insights gained during the target task and applying them to improve the performance of the original source model or task. In reverse transfer learning, the experience and knowledge gained from the target domain or task are used to enhance the performance or update the source model. This can be particularly useful when the target task provides new insights or data that can benefit the original source task or when there is a feedback loop between the target and source tasks. One of the articles we reviewed in this study used Reverse transfer learning [33].

3.1.12. Support Vector Machine (SVM)

SVM is a popular CML algorithm used for classification and regression tasks. SVM is known for its ability to handle both linear and non-linear data by finding an optimal hyperplane that separates different classes or predicts the continuous target variable. In the case of classification, SVM aims to find the best decision boundary that maximally separates different classes in the input feature space. This decision boundary is determined by a subset of training samples called support vectors. SVM works by mapping the input data into a higher-dimensional feature space using a kernel function, which allows the algorithm to find a hyperplane that effectively separates the classes. The choice of the kernel function, such as linear, polynomial, or radial basis function (RBF), affects the SVM's ability to handle complex patterns and non-linear relationships. In the case of regression, SVM aims to find a hyperplane that best fits the data while minimizing the error between the predicted and actual target values. The SVM regression algorithm aims to find a balance between fitting the data closely and controlling the complexity of the model to avoid overfitting. Several articles we reviewed in this study used SVM [2,18,21,94].

3.1.13. Decision Tree

A decision tree is another supervised CML algorithm that is commonly used for classification and regression tasks. It is a flowchart-like structure where internal nodes represent feature tests, branches represent the outcomes of those tests, and leaf nodes represent the predicted class or value. The decision tree algorithm recursively splits the data based on different features to create a tree-like model that can make predictions. At each internal node, a decision is made based on the values of a particular feature, and the data is split into subsets accordingly. This splitting process continues until a stopping criterion is met, such as reaching a maximum tree depth or a minimum number of samples at a node. During training, the decision tree algorithm determines the optimal splits by evaluating different feature and split point combinations based on certain criteria, such as Gini impurity or information gain. The goal is to create splits that result in homogeneous subsets with respect to the target variable. Once the decision tree is trained, it can be used to make predictions by traversing down the tree based on the feature values of an unseen sample. The path followed through the tree leads to a leaf node, which provides the predicted class for classification tasks or the predicted value for regression tasks. Decision trees are popular due to their interpretability and simplicity. They can handle both categorical and numerical features and can capture non-linear relationships between features and the target variable. However, decision trees are prone to overfitting, especially when the tree becomes too complex. Techniques like pruning and ensemble methods, such as random forests, are often used to mitigate overfitting and improve the performance of decision trees. Two articles we reviewed in this study used Decision Tree [21,22].

3.1.14. K-Means

K-means is an unsupervised clustering algorithm that partitions data into a specified number of clusters (k). It assigns each data point to the nearest cluster by minimizing the distance to the

cluster's centroid, which is iteratively updated until the assignments stabilize. It is commonly used for unsupervised learning tasks like grouping similar data points or segmenting datasets. One of the studies [89] we reviewed used K-means on the generated features by ResNet-50.

3.1.15. Random Forest

Random Forest is a CML ensemble method used for both classification and regression tasks. It builds multiple decision trees (described in Section 3.1.13) during training and combines their predictions to improve accuracy and reduce overfitting. Each tree is trained on a random subset of the data, making Random Forest robust and capable of handling complex data patterns. One of the studies [89] we reviewed used Random Forest on the generated data by GAN.

3.1.16. Other Architectures

Saab transform-based successive subspace learning model: It refers to a specific approach for feature extraction and dimensionality reduction in image processing and computer vision tasks. It is based on a series of transformations called the Successive Subspace Learning (SSL) framework, with the Saab transform being one of the key components. The Saab transform is a non-linear transformation that aims to capture discriminative and compact representations of image features. It operates on local image patches and applies a series of operations, including patch-wise mean removal, PCA (Principal Component Analysis), and non-linear transformation using sigmoid functions. These operations are performed successively to obtain a hierarchical representation of the input image.

Non-local channel attention ResNet: It refers to a variant or modification of the ResNet architecture that incorporates non-local channel attention mechanisms. It aims to enhance the representation power of ResNet models by introducing non-local operations that capture long-range dependencies across channels. In the context of the "Non-local channel attention ResNet," the term "non-local" refers to the inclusion of non-local operations within the ResNet architecture. Non-local operations are designed to model relationships between spatial or temporal positions in an image or sequence, capturing dependencies that extend beyond local regions. The "channel attention" component refers to the specific attention mechanism applied to the channels or feature maps within the ResNet. Attention mechanisms help the network focus on relevant information by assigning importance weights to different channels or feature maps. This allows the model to selectively emphasize or suppress certain channels during feature extraction, enhancing its ability to capture important patterns or details. One of the articles we reviewed in this study used Non-local channel attention ResNet [39].

The Multi-symptom multi-label (MSML) network: MSML network refers to a neural network architecture designed to address the problem of multi-label classification in the context of medical diagnosis. In medical diagnosis, patients may exhibit multiple symptoms simultaneously, and it is important to accurately predict the presence or absence of various medical conditions or diseases based on these symptoms. The MSML network is specifically designed to handle this scenario by taking into account multiple symptoms as inputs and predicting multiple labels as outputs. It is a type of deep learning architecture that combines techniques from multi-label classification and DNNs. The MSML network typically consists of multiple layers of interconnected neurons, including input layers, hidden layers, and output layers. The input layer receives the symptom information as input features, which are then processed through the hidden layers to extract relevant patterns and representations. Finally, the output layer produces predictions for multiple medical conditions or diseases associated with the given symptoms. One of the articles we reviewed in this study used MSML network [55].

Model Genesis: "Model Genesis" refers to a specific deep learning architecture designed for medical image analysis tasks. It is an automated framework that aims to generate high-quality medical image segmentation models using a combination of deep convolutional neural networks (CNNs) and genetic algorithms. The concept behind Model Genesis is to leverage the power of genetic algorithms to automatically search and optimize the architecture and hyperparameters of a CNN for a given medical image segmentation task. The genetic algorithm explores a population of potential CNN architectures,

evaluates their performance on a specific objective function (such as accuracy or Dice coefficient), and evolves the population over successive generations to improve the model's performance. The genetic algorithm in Model Genesis operates by using principles inspired by natural selection and evolution. It applies genetic operators such as mutation, crossover, and selection to create new CNN architectures from existing ones, gradually improving their fitness and convergence toward an optimal solution. By integrating genetic algorithms into the architecture search process, Model Genesis aims to automate the design of CNN models for medical image segmentation tasks, reducing the need for manual architecture design and hyperparameter tuning. This approach can potentially save time and effort for researchers and practitioners in the medical imaging field, allowing them to focus more on the analysis and interpretation of medical images. One of the articles we reviewed in this study used Model Genesis [81].

3.2. Loss Functions

A classification model can be defined as $\hat{y} = f_\theta(x)$, where the AI model f_θ is parameterized by a set of parameters θ and an input image x is assigned to the most probable class $\hat{y}$. Given a training set of ultrasound images x_i and their ground truth class $y_i \{(x_i, y_i); i = 1, ..., N\}$, training a classification model consists of finding the model parameters θ that minimize loss $\mathcal{L}$, such as:

$$\theta^* = \arg\min_\theta \sum_{i=1}^{N} \mathcal{L}(\hat{y}_i \mid y_i) \tag{1}$$

Therefore, the choice of the appropriate loss function $\mathcal{L}$ is important, and we briefly discuss the loss functions used in the articles reviewed in this study.

3.2.1. Cross-Entropy Loss

Training an AI model on a binary decision-making task (e.g., COVID-19 *vs.* CAP, or COVID-19 *vs.* healthy, etc.) usually utilizes binary cross-entropy or simply cross-entropy loss defined as:

$$\mathcal{L}_{CE}(X, Y; \theta) = -\frac{1}{N} \sum_{i=1}^{N} y_i \times log(\hat{y}_i) + (1 - y_i) \times log(1 - \hat{y}_i). \tag{2}$$

The cross-entropy loss appears in the majority of ultrasound COVID-19 AI studies (e.g., Born et al [13,15], Che et al [20], Diaz-Escobar et al [25], Frank et al [32], Gare et al [33], Perera et al [73]).

3.2.2. Categorical Cross-Entropy

Categorical cross-entropy works on multiclass (more than two classes; e.g., COVID-19 *vs.* CAP *vs.* Healthy) classification problems. This loss is typically used in an AI model when the model must select one or more categories among numerous possible categories/classes. This loss can be defined as:

$$\mathcal{L}_{CCE}(X, Y; \theta) = -\frac{1}{N} \sum_{i=1}^{N} y_i \times log(\hat{y}_i). \tag{3}$$

Like cross-entropy loss, categorical cross-entropy loss also appears in many ultrasound COVID-19 AI studies (e.g., Barros et al [12], Karar et al [45], Sadik et al [82]).

3.2.3. L1 Loss

L1 loss, also known as mean absolute loss, is typically used when an AI model is tasked to predict a continuous value (e.g., the distance between two landmarks, optimal location for lung scanning using ultrasound, etc.). It is defined as:

$$\mathcal{L}_1(X, Y; \theta) = \sum_{i=1}^{N} \mid y_{true} - y_{predict} \mid, \tag{4}$$

where y_{true} and $y_{predict}$ are the ground truth and predicted continuous values, respectively. Al-Zogbi et al [3] used this loss function to train their deep model to predict landmarks for optimal ultrasound scanning.

3.2.4. Focal Loss

The focal loss is a dynamically scaled cross-entropy loss and is used when there is a class in the training data. Focal loss incorporates a modulating term in the conventional cross-entropy loss so that it can emphasize learning from difficult data samples that lead to misclassification more often. This loss is defined as:

$$\mathcal{L}_{FL}(X, Y; \theta) = -\frac{1}{N} \sum_{i=1}^{N} (1 - \hat{y}_i)^{\gamma} \times log(\hat{y}_i), \tag{5}$$

where γ controls the weight of different samples and $\gamma = 0$ transforms Eq. 5 into a binary cross-entropy loss. Awasthi et al [9] used focal loss in their ultrasound-based COVID-19 study.

3.2.5. Soft Ordinal (SORD) Loss

When output classes are independent of each other, their relative order in the loss calculation during deep model training does not matter. This scenario allows using one-hot encoding, i.e., setting all wrong classes to be infinitely far from the true class. However, there exists a soft order among classes in an ordinal regression scenario, where certain categories are more correct than others with respect to the true label [24] (i.e., a true class is no longer infinitely far from false classes, resulting in a continuity among classes). For these continuously related classes, Roy et al [81] introduced a modified cross-entropy, called soft ordinal (SORD) loss, defined as:

$$\mathcal{L}_{SORD}(X, Y; \theta) = -\sum_{i=1}^{|\mathcal{N}|} \left(\frac{e^{-\delta(n,i)}}{\sum_{j \in \mathcal{N}} e^{-\delta(j,i)}} \right) \times log \left(\frac{e^{f_\theta(x_i)}}{\sum_{j}^{|\mathcal{N}|} e^{f_\theta(x_j)}} \right), \tag{6}$$

where $\mid \mathcal{N} \mid$ is the set of possible soft-valued classes, n is a possible ground truth soft value, δ is a user-defined distance (e.g., weighted square distance) between scores/levels, f_θ is the deep model, and x_i is the i-th input data.

3.2.6. Dice Loss

Dice Loss is a widely used loss function for image segmentation tasks, especially in medical imaging. It is derived from the Dice Similarity Coefficient (DSC), which measures the overlap between the predicted segmentation and the ground truth. The Dice coefficient ranges from 0 to 1, where 1 represents perfect overlap, and 0 indicates no overlap at all. The Dice Loss is formulated to minimize this overlap difference, making it ideal for handling class imbalance in segmentation problems. Dice Loss is defined as:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \cdot |A \cap B|}{|A| + |B|} \tag{7}$$

Here, A represents the set of predicted pixels, and B represents the set of ground truth pixels. The intersection $|A \cap B|$ measures the common pixels between the prediction and the ground truth,

while $|A|$ and $|B|$ represent the sizes of the predicted and ground truth sets, respectively. Dice loss effectively addresses the issue of class imbalance, which is common in segmentation tasks where the background class may dominate the image. By focusing on maximizing the overlap between predicted and true regions, Dice loss ensures that even small, underrepresented regions are accurately captured by the model. This makes it particularly effective in medical applications, such as tumor or organ segmentation, where precise boundary delineation is essential. Its ability to handle imbalanced data and focus on correct segmentation overlap makes it a preferred choice for segmentation tasks in DL. Zhao et al [103] and Vinod et al [92] used Dice loss in their LUS COVID-19 study.

3.3. Evaluation Criteria

The effectiveness of a predictive model is measured by an evaluation metric. This often entails building a model on a dataset, testing it on a holdout dataset that was not used during training, and comparing the predictions to the holdout dataset's expected values. The metrics compare the actual class label to the predicted class label for the classification problems. The different studies reviewed in this article used different types of evaluation criteria, which we briefly discuss below. We also clarify a few key acronyms that are typically used to define different evaluation criteria here.

- True Positive (TP): A result that is positive as both the actual value and expected value.
- True Negative (TN): A result that is negative as both the actual value and expected value.
- False Positive (FP): A false positive occurs when a projected outcome is indicated as being positive when it is actually negative.
- False Negative (FN): A false negative occurs when a projected outcome is indicated as being negative when it is actually positive.

3.3.1. Precision

The ratio of accurate positive predictions and all positive predictions is known as precision. Precision is the proportion of true positives to all predicted positives, which is defined as:

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{8}$$

3.3.2. Recall

Recall, also known as Sensitivity, estimates the ratio of the number of predicted positive samples and the actual number of positive samples, which is defined as:

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{9}$$

3.3.3. Specificity

Specificity is the complement of Sensitivity, which estimates the ratio of the number of predicted negative samples and the actual number of negative samples. It is defined as:

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{10}$$

3.3.4. Accuracy

The proportion of accurately predicted samples among all predictions is known as accuracy, which is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{11}$$

3.3.5. F1–Score

The weighted average of precision and recall is the F1 score. This metric is generally more beneficial than accuracy, especially if there is an uneven class distribution. F1-score is defined as:

$$\text{F1–score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{12}$$

3.3.6. Intersection over Union (IoU)

IoU is typically used in segmentation accuracy estimation, which is the ratio of overlap between the bounding box around a predicted object and the bounding box around the ground truth object mask. It can be defined as:

$$\text{IoU} = \frac{TP}{TP + FP + FN}. \tag{13}$$

3.3.7. Sørensen–Dice Coefficient

Sørensen–Dice coefficient, or simply Dice, is another common metric used in segmentation accuracy estimation, which is defined as:

$$\text{Sørensen–Dice} = \frac{2 \times TP}{(2 \times TP) + FP + FN}. \tag{14}$$

3.3.8. Mean Squared Error (MSE)

Mean Squared Error (MSE) measures the average squared difference between predicted and actual values, which is defined as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{15}$$

where n is the number of samples, y_i represents the true value, and $\hat{y}_i$ represents the predicted value.

4. Studies Using POCUS Dataset

We discussed the POCUS dataset [13] in Section 2.1, which can be used in making breakthroughs in the diagnosis, monitoring, and reporting of COVID-19 pneumonia in patients. This dataset contains COVID-19 (35%), bacterial pneumonia (28%), viral pneumonia (2%), and healthy (35%) classes.

4.1. Studies

In Table 4, we summarize studies that used the POCUS dataset to develop and evaluate AI methods. Al-Jumaili et al [2] utilized a set of pre-trained CNN models, namely ResNet-18, ResNet-50, GoogleNet, and NASNet-Mobile, to extract the features from the images. These features are then fed to an SVM classifier to classify the images into COVID-19, CAP, and healthy classes. A regression task was performed by Al-Zogbi et al [3], who employed DenseNet to approximate the position of the ultrasound probe in the desired scanning areas of the torso. Almeida et al [5] investigated a lightweight neural network, MobileNets, in the context of computer-aided diagnostics and classified ultrasound videos among abnormal, B-lines, mild B-lines, severe B-lines, consolidations, and pleural thickening classes. Awasthi et al [9] also focused on lightweight networks that can operate on mobile or embedded devices to enable rapid bedside detection without additional infrastructure. Their method classified ultrasound images into COVID-19, CAP, and healthy classes. Barros et al [12] proposed a CNN-LSTM hybrid model for the classification of lung ultrasound videos among COVID-19, bacterial pneumonia, and healthy classes. The extraction of the spatial feature was performed by CNNs, while the time dependency was established using the LSTM module. Born et al. published three consecutive articles [13–15] using POCOVID-Net, VGG-16, and Model-genesis, respectively, to classify lung ultrasound images into COVID-19, CAP, and healthy classes. Several pre-trained neural networks such as VGG-19, InceptionV3, Xception, and RestNet-50 have been fine-tuned on the lung ultrasound

image by Diaz-Escobar et al [25] to detect COVID-19 in the lung ultrasound test data. Gare et al [33] used reverse transfer learning in a U-Net, where weights were pre-trained for segmentation and then transferred for the COVID-19, CAP, and Healthy ultrasound image classification task. In order to address the need for a less complex, power efficient, and less expensive solution to screen lung ultrasound images and monitor lung status, Hou et al [36] introduced a Saab transform-based subspace learning model to find the A-line, B-line, and consolidation in lung ultrasound data. Karar et al [45] introduced a lightweight deep model, COVID-LWNet, to make an efficient CNN-based system for classifying lung ultrasound images into COVID-19, bacterial pneumonia, and healthy classes. In addition, Karar et al [46] proposed a GAN to perform the same task on ultrasound images. Few-shot learning is a machine learning framework where a machine learning model is trained with supervision using a few training samples. Karnes et al [47] used the few-shot learning on the POCUS dataset and classified test images into COVID-19, CAP, and healthy classes. A few other approaches also used state-of-the-art CNNs [67,82] or transformers [73] to classify lung ultrasound images into COVID-19, CAP, and healthy classes. Esmaeili and Feghhi [30] introduces ULBPFP-Net, a model combining handcrafted Uniform Local Binary Pattern on Five intersecting Planes (ULBPFP) and VGG16-generated features from lung ultrasound data. The model is designed to diagnose COVID-19, overcoming the challenges of noisy and low-quality ultrasound images. The use of five proposed planes helps capture distinctive disease features. Madhu et al [59] presents XCovNet, an optimized Xception-based CNN designed for diagnosing COVID-19 using POCUS images. The model employs depth-wise separable convolutions, which reduce computational complexity while maintaining high performance. The study emphasizes high accuracy and efficient feature extraction. Nehary et al [69] proposes a fusion method that combines VGG-16 and ViT-generated features with a Histogram of Oriented Gradients (HoG) to classify COVID-19 and bacterial pneumonia. Dimensionality reduction is achieved using principal component analysis (PCA) on HoG features, improving classification efficiency. Finally, Rahhal et al [78] compared the performance of EfficientNetB2, gMLP, and ViT models for COVID-19 detection. Their approach leverages these advanced architectures to achieve impressive recall and precision metrics, effectively distinguishing between COVID-19, bacterial pneumonia, and healthy cases.

Table 4. A summary of studies that used the POCUS dataset. ✗ indicates either absent or not discussed in the article, and ✓ indicates present but not discussed in the article. Acronyms- ROC: Receiver Operative Characteristic, AUC: Area Under the Curve.

Studies	AI models	Loss	Results	Cross-validation	Augmentation/ Pre-processing	Prediction Classes	Code
Al-Jumaili et al [2]	ResNet-18, RestNet-50, NASNetMobile, GoogleNet, SVM	Categorical cross-entropy	Accuracy: 99%	k=5	✗	COVID-19, CAP, Healthy	✗
Al-Zogbi et al [3]	DenseNet	L1	Mean Euclidean error 14.8±7.0 mm	✗	✗	-	✗
Almeida et al [5]	MobileNet	Categorical cross-entropy	Accuracy: 95-100%	✗	✗	Abnormal, B-lines, Mild B-lines, Severe B-lines, Consolidations, Pleural thickening	✗
Awasthi et al [9]	Modified MobileNet, CNN, and other lightweight models	Focal loss	Accuracy 83.2%	k=5	✗	COVID-19, CAP, Healthy	✗
Barros et al [12]	POCOVID-Net, DenseNet, ResNet, NASNet, Xception-LSTM	Categorical cross-entropy	Accuracy: 93%, Sensitivity: 97%	k=5	✗	COVID-19, Bacterial Pneumonia, Healthy	Available[11]
Born et al [13]	POCOVID-Net	Categorical cross-entropy	AUC: 0.94, Accuracy: 0.89, Sensitivity: 0.96, Specificity: 0.79, F1-score: 0.92	k=5	Rotations of up to 10°; Horizontal and vertical flipping; Shifting up to 10% of the image height or width	COVID-19, CAP, Healthy	✗
Born et al [14]	VGG-16	Categorical cross-entropy	Sensitivity: 0.98±0.04, specificity: 0.91±0.08	k=5	Horizontal and vertical flips, rotations up to 10° and translations of up to 10%	COVID-19, CAP, Healthy	✗
Born et al [15]	Frame based: VGG-16 Video-based: Models Genesis	Categorical cross-entropy	Sensitivity: 0.90±0.08, specificity: 0.96±0.04	k=5	Resizing to 224×224 pixels; Horizontal and vertical flips; Rotation up to 10°; Translations of up to 10%	COVID-19, CAP, Healthy	Available[12]
Diaz-Escobar et al [25]	InceptionV3, ResNet-50, VGG-19, Xception	Cross-entropy	Accuracy: 89.1%, ROC-AUC: 97.1%	k=5	Rotations (10°), horizontal and vertical flips, shifts (10%), and zoom (zoom range of 20%)	COVID-19, non-COVID	✗

Table 4. *Cont.*

Studies	AI models	Loss	Results	Cross-validation	Augmentation/ Pre-processing	Prediction Classes	Code
Gare et al [33]	U-Net (reverse-transfer learning; segmentation to classification)	Cross-entropy	mIoU: 0.957±0.002, Accuracy: 0.849, Precision: 0.885, Recall: 0.925, F1-score: 0.897	k=3	Left-to-right flipping; Scaling grey image pixels;	COVID-19, CAP, Healthy	✗
Hou et al [36]	Saab transform based successive subspace CNN model	Categorical cross-entropy	Accuracy: 0.96	✗	Saab transformation	A-line, B-line, Consolidation	✗
Karar et al [45]	MobileNets, ShuffleNets, MENet, MnasNet	Categorical cross-entropy	Accuracy: 99%	✓	Grayscale conversion	COVID-19, Bacterial Pneumonia, Healthy	✗
Karar et al [46]	A semi-supervised GAN, and a modified AC-GAN with auxiliary classifier	Min-Max loss: special form of cross-entropy	Accuracy: 91.22%	✓	Grayscale conversion	COVID-19, CAP, Healthy	✗
Karnes et al [47]	Few-shot learning (FSL) visual classification algorithm	Mahalanobis distances	ROC-AUC > 85%	k=10	✗	COVID-19, CAP, Healthy	Available upon request
Muhammad and Hossain [67]	CNN	Categorical cross-entropy	Accuracy 91.8%, Precision 92.5%, Recall 93.2%	k=5	Reflection around x- and y-axes; Rotation by [-20°, +20°]; Scaling by a factor [0.8, 1.2]	COVID-19, CAP, Healthy	✗
Sadik et al [82]	DenseNet-201, ResNet-152V2, Xception, VGG-19, NasNetMobile	Categorical cross-entropy	Accuracy: 0.906 (with SpecMEn), F1-score: 0.90	✓	Contrast-Limited Adaptive Histogram Equalization	COVID-19, CAP, Healthy	✗
Perera et al [73]	Transformer	Categorical cross-entropy	Accuracy: 93.9%	✓	✗	COVID-19, CAP, Healthy	✗
Esmaeili and Feghhi [30]	ULBPFP-Net	Categorical cross-entropy	Accuracy (A): 99.76%, Specificity: 99.89%, Sensitivity: 99.87%, F1-score: 99.75%	✓	Resizing to 224×224 pixels; Isotropic transformation; Stratified random selection	COVID-19, Viral pneumonia, Healthy	✗
Madhu et al [59]	XCovNet	Categorical cross-entropy	Accuracy: 98.33% for COVID-19 detection	✓	Use of Uniform Local Binary Pattern on 5 intersecting Planes (ULBPFP) to extract features from ultrasound data	COVID-19, Lung disease, Healthy	✗
Nehary et al [69]	Fusion of VGG-16/ViT-generated and HoG Features	Cross-entropy/Categorical cross-entropy	(For ViT and HoG) Accuracy: 91%	k=10	Resizing to 128×128 pixels; PCA for dimensionality reduction of HOG features	COVID-19, Bacterial Pneumonia	✗
Rahhal et al [78]	EfficinetNetB2, gMLP, ViT	Categorical cross-entropy	COVID-19 Detection Recall: 99.88%, Precision: 95.84%, F1-score: 97.41%	k=5	Cropping to 224×224 pixels	COVID-19, Bacterial Pneumonia, Healthy	✗

4.2. Evaluation

Studies using the POCUS dataset reported impressive results across various metrics and method-ologies. For instance, Al-Jumaili et al [2] achieved accuracy, precision, and F1-score of above 99%. Awasthi et al [9] developed a power and memory-efficient network that attained an impressive highest accuracy of 83.2%. Among pre-trained models, Diaz-Escobar et al [25] found that the InceptionV3 model had the highest accuracy of 89.1% and ROC-AUC of 97.1%. In semantic segmentation, Gare et al [33] reported high scores for various metrics, including mIoU (0.957), accuracy (0.849), precision (0.885), recall (0.925), and F1-score (0.897). Saab transform-based successive subspace learning model was reported to have an accuracy of 0.96 by Hou et al [36]. Additionally, modified AC-GAN (accuracy: 99.45%) outperformed semi-supervised GAN (accuracy: 99%) in a study by Karar et al [46], while MnasNet achieved the best accuracy of 99% among six pre-trained networks. Muhammad and Hossain [67] obtained high scores for accuracy, precision, and recall (91.8%, 92.5%, and 93.2%, respectively) with a fusion-based less complex CNN architecture. Real-time mass COVID-19 testing by Perera et al [73] resulted in over 90% accuracy, while spectral mask enhancement (SpecMEn) improved the accuracy score of DenseNet-201 from 89.5% to 90.4% in a study by [82]. The ULBPFP-Net approach [30] achieved an exceptional COVID-19 detection accuracy of 99.76%, demonstrating robustness to noise in LUS images. Specificity and sensitivity were equally impressive, reaching 99.89% and 99.87%, respectively, with a high F1-score of 99.75%. Similarly, XCovNet [59] outperformed state-of-the-art deep learning models in COVID-19 classification, delivering a strong accuracy of 98.33%. Nehary et al [69]'s fusion of ViT-generated features with HoG attained a solid accuracy of 91%, effectively classifying COVID-19 and bacterial pneumonia. Lastly, Rahhal et al [78] achieved a remarkable recall of 99.88%, with a precision of 95.84%, and an F1-score of 97.41%, underlining the robustness of their model in distinguishing between COVID-19, bacterial pneumonia, and healthy subjects.

5. Studies using ICLUS-DB Dataset

We discussed the ICLUS-DB dataset in Section 2.1, which can also be used in making break-throughs in the diagnosis, monitoring, and reporting of COVID-19 pneumonia in patients. This resource may enable AI in the identification of the disease's progress, rate, and response to treatment, facilitating more effective and personalized patient care. This dataset contains lung ultrasound data with different COVID-19 severity scores, defined as score 0: Continuous A-line (34% of the total data), score 1: alteration in A-line (24% of the total data), score 2: small consolidation (32% of the total data), and score 3: large consolidation (10% of the total data). The following table (Table 5) summarizes the literature on the detection of COVID-19 through the use of the ICLUS-DB dataset.

Table 5. A summary of studies that used the ICLUS-DB dataset. ✗ indicates either absent or not discussed in the article, and ✓ indicates present but not discussed in the article.

Studies	AI models	Loss	Results	Cross-validation	Augmentation/ pre-processing	Prediction Classes	Code
Carrer et al [18]	HMM, VA, SVM	✗	Accuracy: 88% (convex probe) 94% (linear probe)	k=10	✗	Severity Score (0, 1, 2, 3)	✗
Che et al [20]	Multi-scale residual CNN	Cross-entropy	Accuracy: 95.11%, F1-score: 96.70%	k=5	Generation of local phase filtered and radial symmetry transformed images	COVID-19, non-COVID	✗
Dastider et al [23]	Autoencoder-based Hybrid CNN-LSTM	Categorical cross-entropy	Accuracy: 67.7% (convex probe) 79.1% (linear probe)	k=5	Rotation, horizontal and vertical shift, scaling, horizontal and vertical flips	Severity Score (0, 1, 2, 3)	Available[13]
Frank et al [32]	ResNet-18, ResNet-101, VGG-16, MobileNetV2, MobileNetV3, DeepLabV3++	SORD, cross-entropy	Accuracy: 93%, F1-Score: 68.8%	✗	Affine transformations, rotation, scaling, horizontal flipping, random jittering	Severity Score (0, 1, 2, 3)	✗
Roy et al [81]	Spatial Transformer Network (STN), U-Net, U-Net++, DeepLabV3, Model Genesis	SORD, cross entropy	Accuracy: 96%, F1-score: 61±12%, Precision: 70±19%, Recall: 60±7%	k=5	✓	Severity Score (0, 1, 2, 3)	✗
Khan et al [48]	Pre-trained CNN from [81]	SORD, cross-entropy	Agreement-based scoring (82.3%)	✗	✗	Severity Score (0, 1, 2, 3)	✗
Khan et al [49]	ResNet-18, ResNet-50, ResNet-101, RegNetX, DenseNet-121, DenseNet-201, EfficientNetB7, and InceptionV3	Cross-entropy	(ResNet-18) F1-score: 65.9% (frame-level). Agreement-based scoring: 59.51% (video level), 63.29% (exam level), and 84.90% (prognostic level)	✗	Image cropping, removal of redundant information, data transformation (e.g., elastic warping, scaling, blurring, rotation)	Severity Score (0, 1, 2, 3)	✗
Torti et al [89]	ResNet-50+K-means	✗	Accuracy: 86.2% (binary), 84.3% (three-way classification), and 74.3% (four-way classification)	✗	Frames extracted and normalized using OpenCV; color parameters converted to grayscale; ResNet-50 used for feature extraction	Severity Score (0, 1, 2, 3)	✗
Custode et al [22]	STN, U-Net+DeepLabV3+, Decision Tree	MSE	Comparable or better performance than state-of-the-art DL methods.	k=5	✗	Severity Score (0, 1, 2, 3)	Available[14]

5.1. Studies

In Table 5, we summarize studies that used the ICLUS-DB dataset [86] to develop and evaluate AI methods. Carrer et al [18] proposed an automatic and unsupervised method to locate the pleural line using the HMM and VA. Afterward, the localized pleural line is used in a supervised SVM to classify the lung ultrasound image into COVID-19 severity scores 0-3. Che et al [20] extracted local phase and radial symmetry features from lung ultrasound images, which were then fed to a multi-scale residual CNN to classify the image between COVID-19 and non-COVID classes. Dastider et al [23] incorporated an LSTM module in DenseNet-201 to predict the COVID severity between 0 and 3 in lung ultrasound images. Frank et al [32] incorporated domain-based knowledge such as anatomical features and pleural and vertical artifacts in conventional CNNs (i.e., ResNet-18, ResNet-101, VGG-16, MobileNetV2, MobileNetV3, and DeepLabV3++) to detect the severity of COVID-19 in lung ultrasound images. Roy et al [81] trained several benchmark CNN models such as U-Net, U-Net++, DeepLabV3, and model genesis, incorporating STNs to simultaneously predict COVID-19 severity scores as well as localize pathological artifacts in a weakly-supervised way in the lung ultrasound images. In a unique study, [48] evaluated the performance of AI deep models in COVID-19 severity scoring by varying the image resolution and gray-level intensity of lung ultrasound images. Khan

et al [49] presented a comparative study utilizing multiple deep learning architectures, including ResNet and DenseNet variants, to assess the accuracy of frame-, video-, exam-, and prognostic-level COVID-19 severity scoring from ultrasound images. This study highlights the importance of data augmentation techniques like elastic warping and scaling to improve model performance. Torti et al [89] applied ResNet-50 combined with K-means clustering for COVID-19 severity classification, achieving competitive accuracy across binary, three-way, and four-way classification tasks, with the frames pre-processed through grayscale conversion and normalization. Lastly, Custode et al [22] proposed a hybrid neuro-symbolic model integrating STN, U-Net+DeepLabV3+, and decision trees for COVID-19 severity assessment, showing comparable or superior results to state-of-the-art deep learning approaches.

5.2. Evaluation

Studies that used the ICLUS-DB, as summarized in Table 5, reported impressive results across various metrics. Carrer et al [18] reported an accuracy of 88% and 94% for lung ultrasound images acquired with the convex and linear probes, respectively, when they used SVM in detecting pleural line alterations due to COVID-19. Che et al [20] reported an accuracy of 95.11% and an F1-score of 96.70% in predicting the COVID-19 severity scores in lung ultrasound. Other studies mostly predicted the COVID-19 severity scores [0, 3] using the ICLUS-DB lung ultrasound dataset as summarized in Table 5. For example, accuracy in severity scoring is reported to be 67.7-79.1%, 93%, 96%, and 82.3% by Dastider et al [23], Frank et al [32], Roy et al [81], and Khan et al [48]. Khan et al [49] expanded the evaluation of severity scoring by benchmarking various deep learning models, achieving a frame-level F1-score of 65.9% with ResNet-18 and agreement-based scoring from 59.51% at the video level to 84.90% at the prognostic level. Torti et al [89] reported competitive accuracies in binary, three-way, and four-way classification tasks, achieving 86.2%, 84.3%, and 74.3%, respectively, highlighting the model's robustness across different classification scales. Lastly, Custode et al [22] demonstrated comparable or superior performance to state-of-the-art models in COVID-19 severity assessment, integrating deep learning and decision trees for an interpretable and effective analysis.

6. Studies Using COVIDx-US Dataset

The COVIDx-US is another large public dataset (discussed in Section 2.1) that has been thoroughly reviewed, analyzed, and validated to develop and assess AI models and algorithms [28]. Table 6 summarizes existing deep learning approaches that used this dataset for COVID-19 identification and characterization in lung ultrasound images.

Table 6. A summary of studies that used the COVIDx-US dataset. ✗ indicates either absent or not discussed in the article. Acronyms- ROC: Receiver Operative Characteristic, AUC: Area Under the Curve.

Studies	AI models	Loss	Results	Cross-validation	Augmentation/ pre-processing	Prediction Classes	Code
Adedigba and Adeshina [1]	SqueezeNet, MobileNetV2	Categorical cross-entropy	Accuracy: 99.74%, Precision: 99.58%, Recall: 99.39%	✗	Rotation, Gaussian blurring, random zoom, random lighting, random warp	COVID-19, CAP, Normal, Other	✗
Azimi et al [11]	InceptionV3, RNN	Cross-entropy	Accuracy: 94.44%	✗	Padding	Positive (COVID-19), Negative (non-COVID-19)	Available[15]
MacLean et al [57]	COVID-Net US	Cross-entropy	ROC-AUC: 0.98	✗	✗	Positive (COVID-19) Negative (non-COVID-19)	Available[16]
MacLean et al [58]	ResNet	Categorical cross-entropy	Accuracy: 0.692	✗	✗	Lung ultrasound severity score (0, 1, 2, 3)	✗
Zeng et al [100]	COVID-Net US-X	Cross-entropy	Accuracy: 88.4%, AUC: 93.6%	✗	Random projective augmentation	Positive (COVID-19) Negative (non-COVID-19)	✗
Zeng et al [101]	COVID-Net L2C-ULTRA	✗	Accuracy: 90.6%, AUC: 97.1%, Recall: 97.1%, Precision: 90.8%	✗	Random projection, piecewise affine transformation, linear-convex transformation	Positive (COVID-19) Negative (non-COVID-19)	✗
Song et al [87]	COVID-Net USPro	Cross-entropy	Accuracy: 99.55%, Recall: 99.93%, Precision: 99.83%	✗	Resizing to 224 × 224 pixels; rotation of 90°, 180°, and 270°	Positive (COVID-19) Negative (non-COVID-19)	Available[17]

6.1. Studies

We summarize studies that used the COVIDx-US dataset to develop and evaluate AI methods in Table 6. Adedigba and Adeshina [1] used computation and memory efficient SqueezeNet and MobileNetV2 to classify lung ultrasound images in COVID-19, CAP, normal, and other classes. Using a hybrid network consisting of the InceptionV3 model to extract spatial information and recurrent neural network (RNN) for extracting temporal features, Azimi et al [11] did binary classification of lung ultrasound images into COVID-19 and non-COVID classes. MacLean et al [57] proposed a DNN, COVID-Net US, leveraging a generative synthesis process that finds an optimal macro-architecture design in classifying lung ultrasound images into COVID-19 and non-COVID classes. Furthermore, MacLean et al [58] used ResNet to classify lung ultrasound images into one of the four lung ultrasound severity scores (i.e., 0, 1, 2, 3). Zeng et al [100] proposed an improved COVID-Net US network, called COVID-Net US-X, that leveraged a projective transformation-based augmentation to transform linear probe data to better resemble convex probe data. This approach performed binary classification of lung ultrasound images into COVID-19 and non-COVID classes. Zeng et al [101] introduced COVID-Net L2C-ULTRA, an advanced model employing multiple transformation techniques such as random projection, piecewise affine, and linear-convex transformation. The network achieves high-performance metrics in classifying lung ultrasound images into COVID-19 and non-COVID categories. Song et al [87] developed COVID-Net USPro, which classifies lung ultrasound images into COVID-19 and non-COVID categories. This model incorporates image preprocessing techniques like resizing and rotation and demonstrates exceptional performance.

6.2. Evaluation

The COVIDx-US dataset was used to implement various models, whose performance is illustrated by various evaluation metrics in Table 6. The models implemented by Adedigba and Adeshina [1] achieved high levels of accuracy (99.74%), precision rate (99.58%), and recall (99.39%). Meanwhile, Azimi et al [11]'s hybrid network attained an overall accuracy of 94.44% and learned to categorize COVID-19 as a binary classification problem. MacLean et al [57]'s deep model achieved an area-under-the-curve (AUC) of over 0.98 while reducing architectural and computational complexity and inference times significantly. The ResNet implemented by MacLean et al [58] achieved a total accuracy of 69.2% with varying sensitivity values for different classes. Among all the models, the MobileNet and SqueezeNet variations of CNN performed the best on this dataset, with Zeng et al [100] achieving a gain of 5.1% in test accuracy and 13.6% in AUC. COVID-Net L2C-ULTRA by Zeng et al [101] achieved remarkable performance with an accuracy of 90.6%, AUC of 97.1%, recall of 97.1%, and precision of 90.8%, marking a significant improvement in classification efficiency for LUS. Song et al [87] also reported exceptional results with their COVID-Net USPro model, achieving an accuracy of 99.55%, recall of 99.93%, and precision of 99.83% in COVID vs. non-COVID classification, demonstrating near-perfect classification performance on LUS.

7. Other Publicly Accessible LUS COVID-19 Datasets

This section discusses three additional public datasets: Boston Emergency Department Lung UltraSound (BEDLUS), COVID-19 Simulated and Labeled *In Vivo* Dataset (CSLID), Fictional Lumen Dissection Dataset (FLDD), and Lung Ultrasound COVID Phantom Dataset (LUCPD). Despite public access, to our knowledge, these datasets are used by one study each to date. Table 7 summarizes the AI studies that used these datasets.

Table 7. A summary of studies used the BEDLUS, CSLID, FLDD, and LUCPD datasets. ✗ indicates either absent or not discussed in the article. Acronyms- AUC: Area Under the Curve.

Studies	Dataset	AI models	Loss	Results	Cross-validation	Augmentation pre-processing	Prediction Classes	Code
Lucassen et al [56]	BEDLUS	3D U-Net, ResNet3D-18, ResNet(2+1)D-18, DenseNet-121, EfficientNetB0, ViT, DeepLabV3+	Cross-entropy	AUC: 95.5%; F1-score for single-point B-line localization: 65%	k=5	Random translation, rotation, scaling, flipping, occlusions, random brightness, contrast change, Gaussian noise and blurring, resizing to 384×256 pixels	B-lines, No B-lines	Available[18]
Zhao et al [103]	CSLID	U-Net	Dice Loss	Maximum Test Dice: 0.741±0.185	k=3	CSLID Data itself is used as augmentated data	B-line pixels	Available[19]
Vinod et al [92]	FLDD	GAN+Random Forest	-	(COVID-19 Detection) Recall: 98%, Precision: 97%, F-score: 98%	k=5, 10	Resizing to 512×512 pixels, RGB conversion	COVID-19, CAP, Healthy	✗
Howell et al [37]	LUCPD	Lightweight U-Net	Combo Loss = (2×Dice Loss + Cross-entropy)	Mean Test Dice: 0.74±0.02	✗	Random flipping and rotation, varying gain, and time-gain compensation, random cropping and padding, resizing to 256×256	Pixels for background, rib, pleural line, A-line, B-line, and B-line confluence	Available[20]

7.1. Studies

Using the BEDLUS dataset, Lucassen et al [56] proposed an approach leveraging several deep learning architectures such as 3D U-Net, ResNet3D-18, ResNet(2+1)D-18, DenseNet-121, EfficientNetB0, ViT, and DeepLabV3+ for B-line detection in lung ultrasound images. The models were trained using a combination of cross-entropy loss and extensive data augmentation techniques, including random translation, rotation, scaling, flipping, and various noise and brightness adjustments. Using the CSLID, Zhao et al [103] used a U-Net-based architecture designed specifically for B-line segmentation in ultrasound images. Dice loss was utilized to optimize the model, and the CSLID dataset itself served as the augmented data for training in conjunction with small-cohort real LUS data. Using the FLDD dataset, Vinod et al [92] employed a hybrid model combining a GAN with a Random Forest classifier to detect COVID-19 from chest imaging data, with preprocessing steps involving image resizing and RGB conversion. Additionally, Howell et al [37] used the LUCPD dataset to develop a lightweight U-Net model for lung ultrasound segmentation. The architecture was optimized using a combination of Dice loss and cross-entropy, while augmentation techniques such as random flipping, rotation, and cropping were applied to improve the robustness of the model in detecting key lung features like A-lines, B-lines, and pleural lines.

7.2. Evaluation

Lucassen et al [56] reported an AUC of 95.5% in B-line detection, with a moderate F1-score of 65% for single-point B-line localization, indicating strong detection performance but challenges in precise localization. Zhao et al [103] achieved a maximum test Dice score of 0.741±0.185 for B-line segmentation using the CSLID dataset, highlighting the variability in segmentation performance. Meanwhile, Vinod et al [92] demonstrated robust results in COVID-19 detection, reporting recall and F1-scores of 98% and precision of 97%, showcasing the effectiveness of their approach in distinguishing COVID-19 cases from other conditions such as CAP and healthy individuals. Finally, Howell et al [37] achieved a mean test Dice score of 0.74±0.02 for segmentation of lung features such as ribs, pleural lines, and B-lines, demonstrating the proficiency of the model in identifying various structures relevant to lung ultrasound interpretation, though with some variability in performance.

8. Studies Using Non-Accessible LUS COVID-19 Private Datasets

Several studies utilized privately owned datasets, which are not publicly accessible, as mentioned in Section 2.2. However, some of these datasets' primary sources, such as hospitals, clinics, and online repositories, have overlapped with those of public data. Although some private dataset links could

not be traced due to lack of availability in the articles, some can be accessed by sending a request for use, for example, Camacho et al [17], Durrani et al [27], Rojas-Azabache et al [79].

Table 8. A summary of studies that used private datasets. ✗ indicates either absent or not discussed in the article. Acronyms- ROC: Receiver Operative Characteristic, AUC: Area Under the Curve.

Studies	AI models	Loss	Results	Cross-validation	Augmentation/ pre-processing	Prediction Classes	Code
Arntfield et al [8]	Xception	Binary Cross Entropy	ROC-AUC: 0.978	✗	Random zooming in/out by ≤10%, horizontal flipping, horizontal stretching/contracting by ≤20%, vertical stretching/contracting (≤5%), and bi-directional rotation by ≤ 10°	Hydrostatic pulmonary edema (HPE), onn-COVID acute respiratory distress syndrome (ARDS), COVID-19 ARDS	Available[21]
Chen et al [21]	2-layer NN, SVM, Decision Tree	✗	Accuracy: 87%	k=5	Curve-to-linear conversion	Score 0: Normal, Score 1: Septal syndrome, Score 2: Interstitial-alveolar syndrome, Score 3: White lung syndrome	✗
Durrani et al [27]	CNN, Regularized STN (Reg-STN)	SORD	Accuracy: 89%, PR-AUC: 73%	k=10	Replacing overlays, resizing to 806×550 pixels	Consolidation present, consolidation absent	✗
Ebadi et al [29]	Kinetics-I3D	Focal loss	Accuracy: 90% Precision: 95%	k=5	✗	A-line (normal), B-line, Consolidation and/or pleural effusion	✗
Huang et al [39]	Non-local Channel Attention ResNet	Cross-entropy	Accuracy: 92.34%, F1-score: 92.05%, Precision: 91.96% Recall: 90.43%	✗	Resizing to 300×300 pixels	Score 0: normal, Score 1: septal syndrome, Score 2: interstitial-alveolar syndrome, Score 3: white lung syndrome	Available[22]
La Salvia et al [52]	ResNet-18, ResNet-50	Cross-entropy	F1-score: 98%	✗	Geometric, filtering, random centre cropping, and colour transformations	Severity score: 0, 0*, 1, 1*, 2, 2*, 3	✗
Liu et al [55]	Multi-symptom multi-label (MSML) network	Cross-entropy	Accuracy: 100% (with 14.7% data)	✗	Random rotation (up to 10 degrees) and horizontal flips	A-line, B-line, Pleural lesion, Pleural effusion	✗
Mento et al [63]	STN, U-Net, DeepLabV3+	✗	Agreement between AI scoring and expert scoring 85.96%	✗	✗	Expert scores: 0, 1, 2, 3	✗
Quentin Muller et al [77]	ResNet-18	Cross-entropy	Accuracy (Val): 100%	✗	Resizing to 349×256	Ultrasound frames with (positive) and without (negative) clinical predictive value	✗
Nabalamba [68]	VGG-16, VGG-19, ResNet	Binary cross-entropy	Accuracy: 98%, Recall: 1, Precision: 96%, F1-score: 97.82%, ROC-AUC: 99.9%	✗	Width and height shifting, random zoom within 20%, brightness variations within [0.4, 1.3], rotation up to 10 degrees	COVID-19, Healthy	✗
Panicker et al [71]	LUSNet (U-Net based CNN)	Categorical cross-entropy	Accuracy: 97%, Sensitivity: 93%, Specificity: 98%	k=5	Generation of local phase and shadow back scatter product images	Classes: 1, 2, 3, 4, 5	Available[23]
Roshankhah et al [80]	U-Net	Categorical cross-entropy	Accuracy: 95%	✗	Randomly cropping and rotating the frames	Severity score: 0, 1, 2, 3	✗
Wang et al [94]	SVM	✗	ROC-AUC: 0.93, Sensitivity: 0.93, Specificity: 0.85	✗	✗	Non-severe, severe	✗
Xue et al [98]	U-Net (with modality alignment contrastive learning of representation (MA-CLR))	Dice, cross-entropy	Accuracy: 75% (4-level) 87.5% (binary)	✗	Affine transformations (translation, rotation, scaling, shearing), reflection, contrast change, Gaussian noise, and Gaussian filtering	Severity score: 0, 1, 2, 3	✗
Kuroda et al [51]	AI-POCUS (model specifics are not disclosed)	✗	(COVID-19 Detection) Accuracy: 94.5% for 12-zone, 83.9% for 8-zone Sensitivity: 92.3% for 12-zone, 77.5% for 8-zone Specificity: 100% for both 12- and 8-zone	✗	✗	Count of B-lines in each zone (≥ 3 means abnormal)	✗
Sagreiya et al [83]	Calculated Lung Ultrasound (CLU)	✗	100% concordance between CLU and radiologist findings in COVID-19 severity	✗	✗	Presence of A-lines, tiny B-lines, confluent B-lines, pleural effusion, thick B-lines, and B-lines with consolidations.	✗
Faita et al [31]	I3D	✗	Accuracy dropped from 70.4% (in 2020) to 48.3% (in 2022), with increased MAE	✗	Random cropping, horizontal flipping, rotation	Severity score: 0, 1, 2, 3	✗
Kimura et al [50]	CNN	✗	Accuracy: 79% (layperson), Accuracy: 80% (physician); Sensitivity: 84% (laypersons), Sensitivity: 94% (physicians)	✗	✗	Normal (no B-lines), Mild-moderately abnormal (1-2 B-lines), Severely abnormal (≥ 3 or coalesced B-lines)	✗
Li et al [53]	KFLR Transformer	Categorical cross-entropy	Binary-level accuracy: 96.4%, Four-level accuracy: 87.4%	✗	Image normalization, pixel size calibration, intensity normalization, and transformation into polar coordinates	Binary-level severity, Four-level severity (0, 1, 2, 3)	✗

8.1. Studies

Arntfield et al [8] highlighted the need for collaborative research involving multi-center for the discrepancy in results between the model and people, which shows the presence of hidden biomarkers

within ultrasound images. In addition, they trained the Xception neural network to classify LUS images into hydrostatic pulmonary edema (HPE), non-COVID acute respiratory distress syndrome (ARDS), and COVID-19 ARDS. Chen et al [21] employed a 2-layer neural network to extract image features, which were subsequently used in an SVM and decision tree algorithm for predicting LUS scores between 0 to 3 (i.e., score 0: normal, score 1: septal syndrome, score 2: interstitial-alveolar syndrome, and score 3: white lung syndrome). Durrani et al [27] used an autonomous deep learning-based technique to detect consolidation/collapses in LUS images. A CNN and Reg-STN-based model has been used with a SORD cross-entropy loss function. A fast and dependable interpretation of LUS images without preprocessing or frame-by-frame analysis was presented by Ebadi et al [29]. They proposed a two-stream inflated 3D CNN, referred to as Kinetics-I3D, to detect A-line (normal), B-line, consolidation, and/or pleural effusion in LUS images. Huang et al [39] proposed a non-local channel attention ResNet to facilitate extraction of the dependencies between distant pixels and stressing specific key channels. Their method classified LUS images into four scores (i.e., score 0: normal, score 1: septal syndrome, score 2: interstitial-alveolar syndrome, and score 3: white lung syndrome). La Salvia et al [52] used ResNet-18 and ResNet-50 to perform a seven-way classification of LUS images. Classes include score 0: A-lines, score 0*: A-lines not defined, score 1: an irregular or damaged pleural line along with visible vertical artifacts, score 1*: pleural line not defined, score 2: broken pleural line with either small or broad consolidated areas with wide vertical artifacts below (white lung), score 2*: broken pleural line not defined, and score 3: dense and broadly visible white lung with or without larger consolidations. Liu et al [55] proposed a novel multi-symptom multi-label (MSML) network incorporating a semi-supervised two-stream active learning strategy, which detected A-line, B-line, pleural lesion, and pleural effusion in LUS images. In a different type of study, Mento et al [63] estimated the agreement of the COVID-19 severity scores predicted by deep models (i.e., STN, U-Net, and DeepLabV3+) to the expert scores. Quentin Muller et al [77] used a pre-trained ResNet-18 to automate the selection of clinically meaningful and predictive image frames from LUS videos that have high clinical predictive value. Nabalamba [68] used three pre-trained deep learning models (i.e., VGG-16, VGG-19, and ResNet) to detect COVID-19 from LUS images. Panicker et al [71] designed a U-Net for lung ultrasound image analysis, called LUSNet, which is trained to classify ultrasound images into five classes of increasing severity of regions. They have followed the rectification of the ultrasound images to make them agnostic to the type of probe employed and to restrict unwanted edge effects, particularly in the case of convex and sector probes. In a typical abnormal lung ultrasound image, B-line artifacts appear, which often evolve into white lung patterns in the more severe cases. Exploiting these anatomical changes, Roshankhah et al [80] used the U-Net-based segmentation approach to automatically stage the progression of COVID-19. While most AI approaches for COVID-19 detection and analysis adopted deep learning techniques, Wang et al [94] extracted hand-engineered features such as thickness and roughness of the pleural line, and the accumulated with an acoustic coefficient of B lines, which were subsequently used in an SVM to classify lung ultrasound images into severe and non-severe cases. Xue et al [98] performed a comprehensive study using the features from LUS data and clinical information in supervised attention-based multiple instance learning (DSA-MIL) modules to classify LUS images into four severity grades. Kuroda et al [51] investigated an AI-based POCUS (AI-POCUS) for detecting COVID-19 pneumonia. They used a commercially available app developed by Philips and compared the AI-POCUS findings with CT scans. The study involved 56 subjects and focused on quantifying B-lines in lung zones. The AI-POCUS system demonstrated high accuracy, even with minimal user experience, showcasing its potential as a rapid screening tool in resource-limited settings. Faita et al [31] utilized 3D-based deep learning models (Inflated 3D ConvNet) on lung ultrasound videos to predict COVID-19 severity. The model was trained on a 2020 dataset and tested on a 2022 cohort to evaluate its robustness over time. The results indicated a significant drop in performance as the disease characteristics evolved, highlighting the importance of continuous model retraining. Kimura et al [50] explored an AI-driven audio output method to assist laypersons in recognizing pulmonary edema or COVID-19 lung infection on ultrasound images. A CNN model was

trained to output audio cues, making it easier for untrained users to interpret the images. The study found that laypersons using audio cues performed comparably to physicians, suggesting its potential for self-monitoring and remote healthcare applications. Li et al [53] introduced a novel Knowledge Fusion Latent Representation (KFLR) framework for assessing COVID-19 pneumonia severity using LUS images. Integrating clinician-guided knowledge into deep learning models improved both accuracy and interpretability. The model achieved high accuracy for both binary and multi-level severity assessments, demonstrating its effectiveness in clinical decision-making.

8.2. Evaluation

Various metrics have been used to evaluate the performance of methods that used private datasets. Arntfield et al [8] were able to distinguish between COVID-19 (AUC = 1.0), non-COVID (AUC = 0.934), and HPE (AUC = 1.0) with high AUCs, whereas Physicians' performance for COVID-19, non-COVID, and HPE detection had AUCs of 0.697, 0.704, and 0.967, respectively. Camacho et al [17] achieved high agreement between the expert and algorithm for detecting B-Lines (88.0%), consolidations (93.4%), and pleural effusion (99.7%), and moderate agreement for the individual video score (72.8%). Chen et al [21] performed a comparison of performance by CNN, SVM, and decision tree models, where the CNN performed the best, achieving 87% accuracy over traditional machine learning models. In the study of Durrani et al [27], the video-based supervised learning method outperformed a fully supervised frame-based method in terms of PR-AUC, with scores of 73.34 and 60.08, respectively. Using a classification model originally developed for recognizing human action, Ebadi et al [29] achieved high accuracy (90%) and average precision (95%). Using a non-local channel attention ResNet, [39] achieved superior performance compared to conventional ResNet, VGG, and other networks, with an accuracy of 92.34% and F1-score of 92.05%. Liu et al [55] reported 100% accuracy for regional classification by training only 14.7% of the data, with comparable performance in sensitivity (92.38%) and specificity (100%). Nabalamba [68] also achieved an accuracy of 98%, along with other high metrics (precision of 95.74, recall of 1.00, F1-score of 97.82%, and ROC-AUC of 99.99%) for the classification of patients at high risk of clinical deterioration and patients at low risk. Similarly, Mento et al [63] showed a high percentage of agreement (85.96%) for the classification of patients at high risk of clinical deterioration and patients at low risk with that by expert radiologists. Quentin Muller et al [77] employed a transfer learning-based approach that achieved high validation accuracy (99.74%) for data with varying brightness levels. Using deep learning approaches, higher accuracy of 97% and 95% are also reported in COVID-19 detection in ultrasound by Panicker et al [71] and Roshankhah et al [80], respectively. Wang et al [94], on the other hand, used an SVM classifier that achieved a good binary classification performance between severe and non-severe cases (sensitivity = 0.93, specificity = 0.85, ROC-AUC = 0.93). By combining lung ultrasound data and clinical information in a multiple instance learning framework, Xue et al [98] were able to categorize patients' clinical severity into four groups with 75% accuracy and into severe/non-severe groups with 87.5% accuracy. Kuroda et al [51] demonstrated that AI-POCUS had a high accuracy of 94.5% for detecting COVID-19 pneumonia when using a 12-zone scan and 83.9% for an 8-zone scan, validated against CT findings. The study showed the potential of AI-POCUS as an effective screening tool. Still, it was limited by its small sample size, single-center design, and lack of comparison with expert-performed traditional ultrasound. Faita et al [31] reported a significant decline in model accuracy, from 70.4% in the 2020 cohort to 48.3% in the 2022 cohort, when using 3D-based deep learning models on lung ultrasound videos. The study highlighted the need for continuous retraining due to evolving disease characteristics, with increased Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) over time. Kimura et al [50] found that AI-generated audio outputs allowed laypersons to identify B-lines in lung ultrasound images with an accuracy of 79%, comparable to physician accuracy of 80%. The sensitivity of laypersons was 84%, slightly lower than that of physicians, at 94%. While promising, the study's small sample size and lack of video interpretation in the layperson group limited its generalizability. Li et al [53] achieved high accuracy with their KFLR model, reporting 96.4% accuracy for binary-level severity assessment and

87.4% for four-level severity assessment using lung ultrasound images. The study demonstrated the method's effectiveness in severity evaluation but was limited by its single-center dataset and reliance on static images, warranting further validation on larger datasets.

9. Challenges, Limitations, and Gaps in Research

Despite significant advancements in the application of AI for COVID-19 detection using LUS data, several challenges and gaps continue to hinder progress and clinical adoption in this field.

Limited Availability and Diversity of Datasets. A critical challenge is the limited availability of large, diverse, and high-quality ultrasound datasets. Most publicly available LUS datasets are small, region-specific, and lack sufficient representation across different patient demographics, comorbidities, and disease stages. This narrow scope restricts the generalizability of AI models, making it difficult for them to perform consistently across diverse populations and clinical settings. Additionally, the lack of longitudinal datasets further limits the ability to evaluate the long-term reliability and clinical utility of AI models in detecting COVID-19.

Lack of Standardized Methodological Reporting. Many studies fail to provide comprehensive documentation of their methodologies, particularly regarding image preprocessing, augmentation, and model implementation. Inconsistent or incomplete reporting makes it challenging to replicate experiments and benchmark new models. The absence of standardized protocols for data handling, model training, and evaluation impedes reproducibility and limits the transparency required for scientific validation. To advance the field, the research community must prioritize the development of clear, standardized guidelines for methodological reporting.

Limited Public Availability of AI Models and Codebases. Another significant limitation is the scarcity of publicly available AI model codebases. While datasets are crucial, open access to code is equally important for fostering collaboration and accelerating advancements. Unfortunately, few studies share their models and codebases openly, e.g., [11–13,23,81], which hinders researchers from validating, improving, or building upon existing models. To promote transparency and reproducibility, the sharing of code and detailed implementation guidelines should become a standard practice in this field.

Lack of Explainability and Interpretability in AI Models. While many AI models achieve impressive accuracy in detecting COVID-19 from ultrasound data, a critical gap exists in their explainability and interpretability. In medical applications, AI models must be transparent and interpretable to gain the trust of healthcare professionals and facilitate clinical adoption. Currently, few studies address the need for explainable AI (XAI) frameworks, leaving clinicians in the dark regarding how AI models make decisions. Without interpretability, it becomes difficult to integrate these models into real-world clinical workflows, where trust and understanding of AI outputs are essential.

Insufficient Focus on Clinical Integration and Longitudinal Validation. Although various AI models have shown high diagnostic performance, there is a lack of longitudinal studies assessing their clinical utility over time. Most models focus solely on diagnostic accuracy without considering their practical integration into existing healthcare systems. Questions remain about how these models perform in dynamic clinical environments, how they can be integrated into decision-making processes, and their ability to adapt to new challenges, such as emerging COVID-19 variants. Addressing these issues requires large-scale clinical trials and the collaboration of healthcare professionals in evaluating AI-driven ultrasound diagnostics.

10. Discussion and Future Works

We began this survey with 874 initial search-yielded articles on the topic of COVID-19 detection using AI on LUS from Google Scholar. After several filtering phases, as discussed in Section 1, we reviewed a total of 60 LUS studies that focused on COVID-19 detection or analysis using AI. Some of the key observations that can be noted from this review are as follows:

COVID-19 Severity Assessment. LUS can be helpful in COVID-19 severity assessment in patients as supported by the studies in the survey [18,20,23,32,81]. COVID-19 primarily affects the respiratory system, causing pneumonia and acute respiratory distress syndrome (ARDS). LUS can detect these lung abnormalities earlier than chest X-rays and provide detailed information on the extent and severity of lung involvement [60]. It can also help differentiate COVID-19 pneumonia from bacterial or viral pneumonia (i.e., CAP). Overall, ultrasound is a safe and non-invasive imaging modality that can provide valuable information for the assessment and management of COVID-19 in patients, especially pregnant women and children. It can help detect early lung involvement, monitor disease progression, and guide clinical decision-making.

Data Partition for Benchmarking. Although numerous publicly available datasets are available, studies have reported varying degrees of quantitative accuracy in detecting, segmenting, and assessing the severity of COVID-19 independently. Without replicating the results of a particular study that used a publicly available ultrasound dataset, it is impossible to make a fair comparison of methodological performance. However, this issue of complexity can be resolved by partitioning a specific portion of a publicly available dataset for quantitative validation across studies. This benchmark dataset can then be used for model validation and quantitative accuracy comparison.

Addressing the Research Questions. Our review paper comprehensively addressed the key research questions (RQs) related to AI-based COVID-19 detection using LUS data. RQ1 focused on identifying the most commonly used public and private LUS datasets for COVID-19 detection. We systematically reviewed and cataloged 7 unique public datasets and 21 private datasets, providing a thorough overview of dataset accessibility and availability. RQ2 examined the variation in AI-based methods applied to COVID-19 detection across different datasets, highlighting key performance metrics such as accuracy, sensitivity, and specificity. Our analysis revealed notable variations in performance metrics across datasets, which underscores the importance of dataset-specific optimization and benchmarking. RQ3 investigated common ultrasound image preprocessing and augmentation techniques used to enhance model performance. We identified and summarized prevalent methods, such as normalization, data augmentation, and denoising, and discussed their impact on model efficacy. RQ4 addressed the existing limitations and challenges in the use of ultrasound for COVID-19 detection. We highlighted issues such as the limited availability of diverse datasets, the lack of standardized methodologies, and the need for explainable AI frameworks. Our review provides actionable insights and recommendations for overcoming these challenges, emphasizing the importance of robust, transparent, and reproducible research practices. Through this comprehensive analysis, our paper fills a critical gap in the literature by offering a detailed and organized perspective on current practices, challenges, and future directions in the field of AI-based COVID-19 ultrasound diagnostics.

Potential Future Works. Based on the observation in this review, we foresee several research directions that can be pursued in the future:

- *Developing a standardized protocol for ultrasound-based severity assessment of COVID-19*: The studies in the survey highlight the potential of LUS in assessing the severity of COVID-19. However, there is a need to develop a standardized protocol for LUS-based severity assessment to ensure consistency across studies and to facilitate comparisons between different AI models. This protocol should include standardized imaging techniques, imaging parameters, and diagnostic criteria.
- *Integration of LUS with other imaging modalities*: While LUS is a useful tool for COVID-19 assessment, it has limitations, such as limited penetration depth and difficulty in imaging certain structures. Future work can focus on combining LUS with other imaging modalities, such as CT or X-ray (if available), to provide a more comprehensive assessment of COVID-19.
- *Integrating AI models for early detection and monitoring of COVID-19*: LUS can detect early lung involvement and monitor disease progression in COVID-19 patients. Future work can focus not only on developing but also on integrating AI models in clinical settings that can accurately

detect COVID-19 at an early stage and monitor disease progression over time, enabling timely intervention and better patient outcomes.

- *Comparison of AI models using benchmark datasets*: As highlighted in the discussion, there is a need for benchmark datasets for quantitative accuracy comparison of different AI models. Future work can focus on developing benchmark datasets and using them to compare the performance of different AI models for COVID-19 detection and analysis.
- *Integration of AI models into clinical practice*: The potential of AI models for COVID-19 detection and analysis is vast, but their integration into clinical practice is still limited. Future work can focus on developing user-friendly and interpretable AI models that can be easily integrated into clinical workflows, improving the accuracy and speed of COVID-19 diagnosis and treatment.
- *Exploration of novel pre-processing and augmentation techniques*: The quality of input data is crucial for the accuracy of AI models. Future work can focus on exploring novel pre-processing and augmentation techniques for ultrasound images to improve the quality of input data and the performance of AI models. These techniques can include advanced filtering, contrast enhancement, or more sophisticated augmentation methods.
- *Integration of clinical and imaging data*: AI models for COVID-19 detection and analysis can benefit from the integration of clinical and imaging data. Future work can focus on developing AI models that can integrate clinical and imaging data to provide a more comprehensive assessment of COVID-19 and its impact on patients.
- *Development of explainable AI (XAI) for ultrasound-based diagnosis*: Given the importance of transparency and trust in medical AI applications, future research can focus on incorporating explainable AI techniques more into LUS-based COVID-19 detection models. This could help healthcare professionals understand the underlying decision-making process of the models and improve their adoption in clinical settings.
- *Federated learning for privacy-preserving model training*: Privacy concerns around patient data sharing remain a critical issue. Future work could explore the use of federated learning, where models are trained across decentralized healthcare institutions without sharing sensitive patient data. This could facilitate large-scale AI model training on diverse datasets without compromising privacy.
- *Personalized AI models for COVID-19 detection*: Developing personalized AI models tailored to individual patient characteristics, such as age, gender, and pre-existing conditions, can improve diagnostic accuracy. Future work can focus on building adaptive AI models that account for individual variability in disease progression.
- *Application of multi-task learning*: Multi-task learning could allow AI models to perform multiple related tasks simultaneously, such as COVID-19 detection, severity assessment, and segmentation of lung lesions. Research in this area could lead to more efficient and versatile models for COVID-19 diagnosis using ultrasound.
- *AI for ultrasound image quality improvement*: Many existing studies rely on suboptimal or noisy ultrasound images. Future research could focus on leveraging AI techniques to enhance the quality and resolution of ultrasound images before they are used for diagnostic purposes, potentially improving the performance of downstream AI models.
- *AI-powered telemedicine for remote ultrasound diagnostics*: AI models can be integrated into telemedicine platforms to enable remote diagnosis of COVID-19 using portable ultrasound devices. This future direction could help extend diagnostic capabilities to rural or underserved areas where access to advanced imaging tools is limited.

11. Conclusions

In this comprehensive review, we provide a detailed survey of LUS-based AI COVID-19 studies that have utilized both publicly available and private LUS datasets. The main contributions of this review include an exhaustive survey of articles using publicly available LUS datasets for COVID-19, a listing and review of these datasets, and the organization of LUS-based AI studies by dataset.

Additionally, this review analyzes and tabulates studies across several dimensions, such as data preprocessing, AI models, cross-validation, and evaluation criteria, and summarizes all reviewed works in a tabular format to facilitate easier comparison among studies. The search strategy employed was thorough, with a total of 60 articles reviewed, with 41 using public datasets and the remainder using private data. We selected articles based on criteria including full-text availability, use of AI techniques for COVID-19 detection or analysis from LUS data, hypothesis support through qualitative and quantitative results, and adherence to a minimum standard of quality.

However, it is important to note some limitations of our review. Despite our rigorous search and selection process, some relevant studies may have been missed due to the exclusion of non-English publications and articles not indexed in the major databases we focused on. Additionally, the variability in AI methodologies and dataset characteristics could mean that some findings may not be directly comparable. These limitations suggest that future research should aim for a more inclusive and standardized approach to enhance the generalizability of AI models for COVID-19 detection using lung ultrasound data. This review provides valuable insights into the current state of LUS-based AI COVID-19 studies and serves as a crucial resource for researchers in this field. The findings can aid in developing more accurate and efficient AI models for COVID-19 detection and diagnosis, ultimately improving patient care and outcomes.

References

1. Adedigba AP, Adeshina SA (2021) Deep learning-based classification of COVID-19 lung ultrasound for tele-operative robot-assisted diagnosis. In: 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), IEEE, pp 1–6

2. Al-Jumaili S, Duru AD, Uçan ON (2021) COVID-19 ultrasound image classification using svm based on kernels deduced from convolutional neural network. In: 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, pp 429–433

3. Al-Zogbi L, Singh V, Teixeira B, et al (2021) Autonomous robotic point-of-care ultrasound imaging for monitoring of COVID-19-induced pulmonary diseases. Frontiers in Robotics and AI 68(645756). doi:10.3389/frobt.2021.645756

4. Allinovi M, Parise A, Giacalone M, et al (2020) Lung ultrasound may support diagnosis and monitoring of COVID-19 pneumonia. Ultrasound in medicine & biology 46(11):2908–2917

5. Almeida A, Bilbao A, Ruby L, et al (2020) Lung ultrasound for point-of-care COVID-19 pneumonia stratification: computer-aided diagnostics in a smartphone. first experiences classifying semiology from public datasets. In: 2020 IEEE International Ultrasonics Symposium (IUS), pp 1–4, doi:10.1109/IUS46767.2020.9251716

6. Alzubaidi M, Zubaydi HD, Bin-Salem AA, et al (2021) Role of deep learning in early detection of COVID-19: Scoping review. Computer methods and programs in biomedicine update 1:100,025

7. de Araujo AF, Constantinou CE, Tavares JMR (2016) Smoothing of ultrasound images using a new selective average filter. Expert Systems with Applications 60:96–106

8. Arntfield R, VanBerlo B, Alaifan T, et al (2020) Development of a deep learning classifier to accurately distinguish COVID-19 from look-a-like pathology on lung ultrasound. medRxiv

9. Awasthi N, Dayal A, Cenkeramaddi LR, et al (2021) Mini-covidnet: Efficient lightweight deep neural network for ultrasound based point-of-care detection of COVID-19. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 68(6):2023–2037. doi:10.1109/TUFFC.2021.3068190

10. Azad R, Kazerouni A, Heidari M, et al (2024) Advances in medical image analysis with vision transformers: a comprehensive review. Medical Image Analysis p 103000

11. Azimi H, Ebadi A, Song J, et al (2022) Covid-net uv: An end-to-end spatio-temporal deep neural network architecture for automated diagnosis of COVID-19 infection from ultrasound videos. arXiv preprint arXiv:220508932

12. Barros B, Lacerda P, Albuquerque C, et al (2021) Pulmonary COVID-19: Learning spatiotemporal features combining cnn and lstm networks for lung ultrasound video classification. Sensors 21(16). doi:10.3390/s21165486, https://www.mdpi.com/1424-8220/21/16/5486

13. Born J, Brändle G, Cossio M, et al (2020) Pocovid-net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (pocus). arXiv preprint arXiv:200412084

14. Born J, Wiedemann N, Brändle G, et al (2020) Accelerating COVID-19 differential diagnosis with explainable ultrasound image analysis. CoRR abs/2009.06116. https://arxiv.org/abs/2009.06116, https://arxiv.org/abs/2009.06116

15. Born J, Wiedemann N, Cossio M, et al (2021) Accelerating detection of lung pathologies with explainable ultrasound image analysis. Applied Sciences 11(2):672

16. Bottenus N, Üstüner KF (2015) Acoustic reciprocity of spatial coherence in ultrasound imaging. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 62(5):852–861

17. Camacho J, Muñoz M, Genovés V, et al (2022) Artificial intelligence and democratization of the use of lung ultrasound in COVID-19: On the feasibility of automatic calculation of lung ultrasound score. International Journal of Translational Medicine 2(1):17–25

18. Carrer L, Donini E, Marinelli D, et al (2020) Automatic pleural line extraction and COVID-19 scoring from lung ultrasound data. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 67(11):2207–2217

19. Castelao J, Graziani D, Soriano JB, et al (2021) Findings and prognostic value of lung ultrasound in COVID-19 pneumonia. Journal of Ultrasound in Medicine 40(7):1315–1324

20. Che H, Radbel J, Sunderram J, et al (2021) Multi-feature multi-scale cnn-derived COVID-19 classification from lung ultrasound data. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, pp 2618–2621

21. Chen J, He C, Yin J, et al (2021) Quantitative analysis and automated lung ultrasound scoring for evaluating COVID-19 pneumonia with neural networks. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 68(7):2507–2515

22. Custode LL, Mento F, Tursi F, et al (2023) Multi-objective automatic analysis of lung ultrasound data from covid-19 patients by means of deep learning and decision trees. Applied Soft Computing 133:109,926

23. Dastider AG, Sadik F, Fattah SA (2021) An integrated autoencoder-based hybrid cnn-lstm model for COVID-19 severity prediction from lung ultrasound. Computers in Biology and Medicine 132:104,296

24. Diaz R, Marathe A (2019) Soft labels for ordinal regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4738–4747

25. Diaz-Escobar J, Ordóñez-Guillén NE, Villarreal-Reyes S, et al (2021) Deep-learning based detection of COVID-19 using lung ultrasound imagery. PLOS ONE 16:1–21. doi:10.1371/journal.pone.0255886

26. Dong D, Tang Z, Wang S, et al (2020) The role of imaging in the detection and management of COVID-19: a review. IEEE reviews in biomedical engineering 14:16–29

27. Durrani N, Vukovic D, van der Burgt J, et al (2022) Automatic deep learning-based consolidation/collapse classification in lung ultrasound images for COVID-19 induced pneumonia. Preprint at https://doi.org/10.21203/rs.3.rs-1531881/v1

28. Ebadi A, Xi P, MacLean A, et al (2022) Covidx-us: An open-access benchmark dataset of ultrasound imaging data for ai-driven COVID-19 analytics. Frontiers in Bioscience-Landmark 27(7):198. doi:10.31083/j.fbl2707198

29. Ebadi SE, Krishnaswamy D, Bolouri SES, et al (2021) Automated detection of pneumonia in lung ultrasound using deep video classification for COVID-19. Informatics in Medicine Unlocked 25:100,687

30. Esmaeili V, Feghhi MM (2023) Covid-19 diagnosis: Ulbpfp-net approach on lung ultrasound data. Iranian Journal of Electrical & Electronic Engineering 19(3)

31. Faita F, De Rosa L, Barbieri G, et al (2024) Covid-19 score severity prediction using 3d-based deep learning models on lung ultrasound video: could the system stand the test of time and of disease's evolution? In: Medical Imaging 2024: Computer-Aided Diagnosis, SPIE, pp 637–642

32. Frank O, Schipper N, Vaturi M, et al (2021) Integrating domain knowledge into deep networks for lung ultrasound with applications to COVID-19. IEEE transactions on medical imaging 41(3):571–581

33. Gare GR, Schoenling A, Philip V, et al (2021) Dense pixel-labeling for reverse-transfer and diagnostic learning on lung ultrasound for COVID-19 and pneumonia detection. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), IEEE, pp 1406–1410

34. Gudigar A, Raghavendra U, Nayak S, et al (2021) Role of artificial intelligence in COVID-19 detection. Sensors 21(23):8045

35. Gürsoy E, Kaya Y (2023) An overview of deep learning techniques for COVID-19 detection: methods, challenges, and future works. Multimedia Systems 29(3):1603–1627. doi:10.1007/s00530-023-01083-0, https://doi.org/10.1007/s00530-023-01083-0

36. Hou D, Hou R, Hou J (2020) Interpretable saab subspace network for COVID-19 lung ultrasound screening. In: 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp 0393–0398, doi:10.1109/UEMCON51285.2020.9298069

37. Howell L, Ingram N, Lapham R, et al (2024) Deep learning for real-time multi-class segmentation of artefacts in lung ultrasound. Ultrasonics 140:107,251

38. Huang C, Wang Y, Li X, et al (2020) Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. The lancet 395(10223):497–506

39. Huang Q, Lei Y, Xing W, et al (2022) Evaluation of pulmonary edema using ultrasound imaging in patients with COVID-19 pneumonia based on a non-local channel attention resnet. Ultrasound in Medicine & Biology 48(5):945–953

40. Huang S, Yang J, Fong S, et al (2021) Artificial intelligence in the diagnosis of covid-19: challenges and perspectives. International journal of biological sciences 17(6):1581

41. Hussain MA, Anas EMA, Alam SK, et al (2012) Direct and gradient-based average strain estimation by using weighted nearest neighbor cross-correlation peaks. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 59(8):1713–1728

42. Hussain MA, Shourov RM, Khan SN (2015) Towards real-time 3d geometric nonlinear diffusion filter and its application to ct and mr imaging. In: 2015 18th International Conference on Computer and Information Technology (ICCIT), IEEE, pp 462–467

43. Hussain MA, Amir-Khalili A, Hamarneh G, et al (2017) Segmentation-free kidney localization and volume estimation using aggregated orthogonal decision cnns. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 612–620

44. Hussain MA, Mirikharaji Z, Momeny M, et al (2022) Active deep learning from a noisy teacher for semi-supervised 3d image segmentation: Application to covid-19 pneumonia infection in ct. Computerized Medical Imaging and Graphics 102:102,127

45. Karar ME, Reyad O, Abd-Elnaby M, et al (2021) Lightweight transfer learning models for ultrasound-guided classification of COVID-19 patients. Computers, Materials and Continua pp 2295–2312

46. Karar ME, Shouman MA, Chalopin C (2021) Adversarial neural network classifiers for COVID-19 diagnosis in ultrasound images. Computers, Materials and Continua pp 1683–1697

47. Karnes M, Perera S, Adhikari S, et al (2021) Adaptive few-shot learning poc ultrasound COVID-19 diagnostic system. In: 2021 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE, pp 1–6

48. Khan U, Mento F, Giacomaz LN, et al (2022) Deep learning-based classification of reduced lung ultrasound data from COVID-19 patients. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 69(5):1661–1669

49. Khan U, Afrakhteh S, Mento F, et al (2023) Benchmark methodological approach for the application of artificial intelligence to lung ultrasound data from covid-19 patients: From frame to prognostic-level. Ultrasonics 132:106,994

50. Kimura BJ, Waltman DR, Han PJ, et al (2024) Effectiveness of audio output from an artificial intelligence method for layperson recognition of pulmonary edema or covid lung infection on ultrasound images. Journal of the American Society of Echocardiography 37(1):112–115

51. Kuroda Y, Kaneko T, Yoshikawa H, et al (2023) Artificial intelligence-based point-of-care lung ultrasound for screening covid-19 pneumoniae: Comparison with ct scans. PLoS One 18(3):e0281,127

52. La Salvia M, Secco G, Torti E, et al (2021) Deep learning and lung ultrasound for COVID-19 pneumonia detection and severity classification. Computers in biology and medicine 136:104,742

53. Li Z, Yang X, Lan H, et al (2024) Knowledge fused latent representation from lung ultrasound examination for covid-19 pneumonia severity assessment. Ultrasonics 143:107,409

54. Liu J (2021) Review of deep learning-based approaches for COVID-19 detection. In: 2021 2nd International Conference on Computing and Data Science (CDS), IEEE, pp 366–371

55. Liu L, Lei W, Wan X, et al (2020) Semi-supervised active learning for COVID-19 lung ultrasound multi-symptom classification. In: 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), IEEE, pp 1268–1273

56. Lucassen RT, Jafari MH, Duggan NM, et al (2023) Deep learning for detection and localization of b-lines in lung ultrasound. IEEE Journal of Biomedical and Health Informatics 27(9):4352–4361

57. MacLean A, Abbasi S, Ebadi A, et al (2021) Covid-net us: A tailored, highly efficient, self-attention deep convolutional neural network design for detection of COVID-19 patient cases from point-of-care ultrasound imaging. In: Domain Adaptation and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health. Springer, p 191–202

58. MacLean A, Ebadi A, Florea A, et al (2021) An initial study into the feasibility of deep learning-based COVID-19 severity classification using point-of-care ultrasound imaging. Journal of Computational Vision and Imaging Systems 7(1):31–33

59. Madhu G, Kautish S, Gupta Y, et al (2024) Xcovnet: An optimized xception convolutional neural network for classification of covid-19 from point-of-care lung ultrasound images. Multimedia Tools and Applications 83(11):33,653–33,674

60. Martinez Redondo J, Comas Rodriguez C, Pujol Salud J, et al (2021) Higher accuracy of lung ultrasound over chest x-ray for early diagnosis of COVID-19 pneumonia. International Journal of Environmental Research and Public Health 18(7):3481

61. Mateu M, Montiel JO, Escalante-Ramírez B (2022) COVID-19 detection from lung ultrasound images. In: Optics, Photonics and Digital Technologies for Imaging Applications VII, SPIE, pp 75–83

62. McLaughlan J, Howell L, Ingram N (2024) Lung ultrasound covid phantom dataset used for training machine learning model. Research Data Leeds Repository

63. Mento F, Perrone T, Fiengo A, et al (2021) Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study. The Journal of the Acoustical Society of America 149(5):3626–3634

64. Mento F, Perrone T, Macioce VN, et al (2021) On the impact of different lung ultrasound imaging protocols in the evaluation of patients affected by coronavirus disease 2019: how many acquisitions are needed? Journal of Ultrasound in Medicine 40(10):2235–2238

65. Momeny M, Neshat AA, Hussain MA, et al (2021) Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep cnn for the detection of covid-19 in x-ray images. Computers in Biology and Medicine 136:104,704

66. Mondal MRH, Bharati S, Podder P (2021) Diagnosis of COVID-19 using machine learning and deep learning: a review. Current Medical Imaging 17(12):1403–1418

67. Muhammad G, Hossain MS (2021) COVID-19 and non-COVID-19 classification using multi-layers fusion from lung ultrasound images. Information Fusion 72:80–88

68. Nabalamba I (2022) Machine learning-aided classification of COVID-19 in lung ultrasound images. PhD thesis, Makerere University

69. Nehary EA, Rajan S, Rossa C (2023) Lung ultrasound image classification using deep learning and histogram of oriented gradients features for covid-19 detection. In: 2023 IEEE Sensors Applications Symposium (SAS), IEEE, pp 1–6

70. Pal SK, Bhardwaj A, Shukla A (2021) A review on despeckling filters in ultrasound images for speckle noise reduction. In: 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), IEEE, pp 973–978

71. Panicker MR, Chen YT, Narayan KV, et al (2021) An approach towards physics informed lung ultrasound image scoring neural network for diagnostic assistance in COVID-19. arXiv preprint arXiv:210606980

72. Park HJ, Jang HY, Kim SY, et al (2020) Non-enhanced magnetic resonance imaging as a surveillance tool for hepatocellular carcinoma: comparison with ultrasound. Journal of hepatology 72(4):718–724

73. Perera S, Adhikari S, Yilmaz A (2021) Pocformer: A lightweight transformer architecture for detection of COVID-19 using point of care ultrasound. In: 2021 IEEE International Conference on Image Processing (ICIP), IEEE, pp 195–199

74. Perrone T, Soldati G, Padovini L, et al (2021) A new lung ultrasound protocol able to predict worsening in patients affected by severe acute respiratory syndrome coronavirus 2 pneumonia. Journal of Ultrasound in Medicine 40(8):1627–1635

75. Prentice R, Wright EK, Flanagan E, et al (2022) The use of fecal calprotectin and intestinal ultrasound in the evaluation and management of stricturing crohn's disease in pregnancy. Inflammatory Bowel Diseases 28(1):e13–e16

76. Qian X, Wodnicki R, Kang H, et al (2020) Current ultrasound technologies and instrumentation in the assessment and monitoring of COVID-19 positive patients. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 67(11):2230–2240

77. Quentin Muller M, Aleandro Eccel M, Arnaud Robert M (2020) Extracting high value lung ultrasound images from video for the diagnosis and prognosis of COVID-19. EPFL Project Report

78. Rahhal MMA, Bazi Y, Jomaa RM, et al (2022) Contrasting efficientnet, vit, and gmlp for covid-19 detection in ultrasound imagery. Journal of Personalized Medicine 12(10):1707

79. Rojas-Azabache C, Vilca-Janampa K, Guerrero-Huayta R, et al (2021) Detection of COVID-19 disease using deep neural networks with ultrasound imaging. arXiv preprint arXiv:210401509

80. Roshankhah R, Karbalaeisadegh Y, Greer H, et al (2021) Investigating training-test data splitting strategies for automated segmentation and scoring of COVID-19 lung ultrasound images. The Journal of the Acoustical Society of America 150(6):4118–4127

81. Roy S, Menapace W, Oei S, et al (2020) Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. IEEE transactions on medical imaging 39(8):2676–2687

82. Sadik F, Dastider AG, Fattah SA (2021) SpecMEn-DL: spectral mask enhancement with deep learning models to predict COVID-19 from lung ultrasound videos. Health Information Science and Systems 9(1):28. doi:10.1007/s13755-021-00154-8

83. Sagreiya H, Jacobs MA, Akhbardeh A (2023) Automated lung ultrasound pulmonary disease quantification using an unsupervised machine learning technique for covid-19. Diagnostics 13(16):2692

84. Siddiqui MA, Ali MA, Deriche M (2021) On the early detection of covid19 using advanced machine learning techniques: A review. In: 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD), IEEE, pp 1–7

85. Soldati G, Demi M, Smargiassi A, et al (2019) The role of ultrasound lung artifacts in the diagnosis of respiratory diseases. Expert review of respiratory medicine 13(2):163–172

86. Soldati G, Smargiassi A, Inchingolo R, et al (2020) 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 39(7):1413–1419

87. Song J, Ebadi A, Florea A, et al (2023) Covid-net uspro: An explainable few-shot deep prototypical network for covid-19 screening using point-of-care ultrasound. Sensors 23(5):2621

88. Stokes EK, Zambrano LD, Anderson KN, et al (2020) Coronavirus disease 2019 case surveillance—united states, january 22–may 30, 2020. Morbidity and Mortality Weekly Report 69(24):759

89. Torti E, Gazzoni M, Marenzi E, et al (2024) Gpu-based key-frame selection of pulmonary ultrasound images to detect covid-19. Journal of Real-Time Image Processing 21(4):113

90. Ulhaq A, Born J, Khan A, et al (2020) COVID-19 control by computer vision approaches: A survey. Ieee Access 8:179,437–179,456

91. Vafaeezadeh M, Behnam H, Gifani P (2024) Ultrasound image analysis with vision transformers. Diagnostics 14(5):542

92. Vinod DN, Kapileswar N, Simon J, et al (2024) Prognosis of covid-19 using ultrasound scans augmented by generative adversarial networks. International Journal of Intelligent Engineering & Systems 17(4)

93. Wang J, Yang X, Zhou B, et al (2022) Review of machine learning in lung ultrasound in COVID-19 pandemic. Journal of Imaging 8(3):65

94. Wang Y, Zhang Y, He Q, et al (2021) A semi-automatic ultrasound image analysis system for the grading diagnosis of COVID-19 pneumonia. arXiv preprint arXiv:211102676

95. Willemink MJ, Koszek WA, Hardell C, et al (2020) Preparing medical imaging data for machine learning. Radiology 295(1):4–15

96. Worldometer (2024) COVID-19 CORONAVIRUS PANDEMIC. https://www.worldometers.info/coronavirus/. Accessed: 9-10-2024.

97. Wynants L, Van Calster B, Collins GS, et al (2020) Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. bmj 369

98. Xue W, Cao C, Liu J, et al (2021) Modality alignment contrastive learning for severity assessment of COVID-19 from lung ultrasound and clinical information. Medical image analysis 69:101,975

99. Yuan J, Ye D, Chen S, et al (2021) Therapeutic ultrasound-enhanced immune checkpoint inhibitor therapy. Frontiers in Physics 9:636,985

100. Zeng EZ, Florea A, Wong A (2022) Covid-net us-x: Enhanced deep neural network for detection of COVID-19 patient cases from convex ultrasound imaging through extended linear-convex ultrasound augmentation learning. doi:10.48550/ARXIV.2204.13851

101. Zeng EZ, Ebadi A, Florea A, et al (2024) Covid-net l2c-ultra: An explainable linear-convex ultrasound augmentation learning framework to improve covid-19 assessment and monitoring. Sensors 24(5):1664

102. Zhao L, Lediju Bell MA (2022) A review of deep learning applications in lung ultrasound imaging of COVID-19 patients. BME Frontiers 2022

103. Zhao L, Fong TC, Bell MAL (2024) Detection of covid-19 features in lung ultrasound images using deep neural networks. Communications Medicine 4(1):41