Deep learning methods for point-of-care ultrasound examination

Giacomo Ignesti*‡, Chiara Deri† Gennaro D'Angelo†, Lorenza Pratali†,
Antonio Bruno* Antonio Benassi*, Ovidio Salvetti*, Davide Moroni*, Massimo Martinelli*
*Institute of Information Science and Technologies
National Research Council of Italy, Pisa, Italy
e-mail corresponding author: e-mail: {*name.surname*}@*isti.cnr.it*
†Institute of Clinical Physiology
National Research Council of Italy, Pisa, Italy
e-mail{*name.surname*}@*ifc.cnr.it*
‡University of Pisa
Department of Computer Science, Pisa, Italy
e-mail{*name.surname*}@*phd.unipi.it*

***Abstract*—Point-of-care Test (POCT) is the delivery of medical care at or near the patient's bedside. Primarily employed in emergencies, where rapid diagnosis and treatment are critical, POCT is now being used in domestic telehealth solutions, as in the TiAssisto project, thanks to technological advances such as the development of portable and affordable devices, high-speed Internet connections, video conferencing, and Artificial Intelligence (AI). Ultrasound (US) images of internal organs and structures are valuable tools in POCT medicine since this examination is portable, quick, and cost-effective. USs can help diagnose different conditions, including heart problems, abdominal pain, and pneumonia. Deep learning algorithms have proven to be highly effective in image recognition, enabling physicians to make informed decisions on-site. This article presents a pipeline approach providing remarkable and reliable results to handle point-of-care ultrasound examinations, making use of methods for: a) automating text cleaning for privacy based on an Optical Character Recognition (OCR) algorithm; b) scrolling through the video frames and annotating them using an ad hoc implemented tool; c) classifying various signs in US using a state of the art deep learning algorithm, that is an adaptive efficient method ensembling two EfficientNet-b0 weak models; d) benchmarking medical plausibility to address transparency and human in the loop setting using a post hoc explanation visual explanation method, i.e. Grad-CAM.**

The involved physician's feedback remarks that this system can detect important signs in pulmonary US imaging. However, the dataset is not yet the final one since the TiAssisto project is still ongoing, with a planned conclusion in February 2024. Our ultimate goal is not merely to develop a classification system but to create an effective healthcare support system that can be used beyond primary healthcare facilities.

***Index Terms*—Point-of-care testing Ultrasound Telemedicine Multi-pathology Artificial Intelligence Explainable Artificial Intelligence (XAI) Optical Character Recognition (OCR) Machine Learning Decision Support System (DSS)**

Funded by Tuscany Region COVID19 Call.
This publication was produced with the co-funding European Union - Next Generation EU, in the context of The National Recovery and Resilience Plan, Investment 1.5 Ecosystems of Innovation, Project Tuscany Health Ecosystem (THE), CUP: B83C22003920001.

I. INTRODUCTION

TiAssisto is a telemedicine project funded by the COVID-19 Tuscany Region call. Its main objectives include exploring and validating the potential benefits of telemedicine in healthcare. A crucial element of this approach is the adoption of delocalized point-of-care testing (POCT) to minimize unnecessary hospital admissions.

Telemedicine projects are increasingly studying the use of specialized examinations, such as ultrasound (US), that can be conducted outside traditional medical facilities. Beyond the confines of conventional medical practice, there is a growing emphasis on integrating these examinations with artificial intelligence (AI). This integration aims to enhance the capabilities of healthcare professionals and provide more effective care to their patients [1].

POCT US has achieved a fundamental role in the last years, becoming the "fifth pillar to bedside physical examination" [2]. Especially during the COVID-19 pandemic, its application to lungs has been crucial in many medical settings: from the emergency department, where it was employed for the differential diagnoses of dyspnea, to the COVID-19 patient's home, where it helped to detect lung involvement. Although USs don't allow lung parenchyma studies due to air presence, some artefacts appear in pathological situations. The main pulmonary deaeration sign shown by ultrasound is the B-line, a vertical hyperechoic artefact arising from the pleural line that is visible when the air in the lung is replaced by water, as in heart failure or fibrosis. Even if B-lines are not specific signs, their features may suggest the presence of a particular disease [3]. In heart failure, B-lines are multiple, bilateral, and more present in dependent zones; moreover, their number is often correlated to possible lung congestion.

On the other hand, in acute distress respiratory syndrome, B-lines show a more irregular distribution pattern, with spared areas and parenchymal consolidations. Pulmonary fibrosis is also characterized by B-lines and an irregular thickened pleural

line. When the density of lung parenchyma increases, lung ultrasound shows consolidated areas, typical of pneumonia, atelectasis or pulmonary contusions. Another condition that can be detected by lung US is pleural effusion: its sonographic features can help detect its nature. All the conditions mentioned above can be easily and effectively monitored through lung US during the hospitalization and in the follow-up, given the portability of the devices.

Many protocols have been developed to quantify lung involvement in different pathological conditions [4]. Regarding heart failure, an 8-scanning site scheme is recommended. The total number of B-lines expresses the severity of congestion [3] Concerning COVID-19, different standardized approaches to lung ultrasound examination have been suggested [5], the majority of them consider a scan area of the chest area considering the possibilities of following different anatomical reference lines. The different scan criteria generated different score criteria, all agreeing on evaluating how many B-lines are in the ultrasound field of view and how much space they occupy. More filled space and more lines indicate a severe score.

Deep learning approaches to study ultrasound signs have risen during the pandemic [6], but their obvious benefits opened a previously undermined branch. Deep learning architectures have shown remarkable ability in detecting these patterns, and so it is possible to train a customized network to detect B-lines, also known as comet tails artefact [7]. While network robustness is pivotal in any technical study, nowadays, AI applications in medicine should also guarantee trustworthiness criteria [8]. Therefore, a study on methodology granting privacy, transparency, and explainability related to deep learning applications should be included in any project of applied AI, especially in medicine. The proposed pipeline involves a self-developed video annotation tool, then it adopts an optical character recognition (OCR) algorithm to inspect and clean the acquired ultrasounds. The collected images are then forwarded to a state-of-the-art (SOTA) deep learning network, a fine-tuned EfficientNet-b0 adaptive efficient ensemble, able to quickly recognize any signs potentially related to pathologies. Classification outputs are then assessed with the Grad-CAM algorithms, to evaluate if the proper medical signs were identified, offering a quick and effective second opinion

In the next Section II, we introduce the subject of deep learning processing in Point-of-Care US (POCUS). This is followed by a presentation of the materials accumulated during the TiAssisto project, as detailed in Section III. The methods applied to address this research inquiry are delineated in Section IV, and the results achieved to date are outlined in Section V. As is customary in scientific literature, the article will end with discussions and recommendations for future research, which are the content of Section VI.

II. Related work

POCUS can be used in conjunction with telemedicine, allowing physicians to get real-time US images from patients in remote locations and transmit them to an expert for their interpretation [9]. The feasibility of using deep learning to assist specialists in the POCUS examination further is now been also investigated [10]. Such a trend is more prominent for other medical image modalities [11] (MRI and CT) but is now transitioning to US studies [12], also based on advances in embedded-AI and in-device machine learning. As for any other medical domain, the lack of available data makes it difficult to benchmark a deep learning algorithm, especially in the framework of trustworthy AI, where data privacy is mandatory. A major problem that can be extrapolated from the literature is that no active trial study on the use of POCUS, telemedicine and AI has yet concluded and shared its results [13]. The majority of the studies employ clinical trials to check the cost-benefit of tele-POCUS or integrated deep learning algorithms in datasets of already concluded projects.

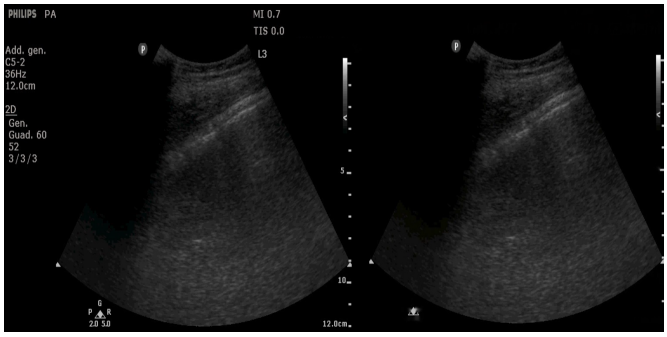

Fig. 1. Annotated ultrasound in the original view and cleaned storing to ensure privacy and improve the quality of the dataset

III. Materials

The primary focus of TiAssisto is the collection of patients' vital signs and medical examinations, particularly US videos. The current dataset comprises lung examinations obtained throughout the TiAssisto initiative. Particular emphasis has been placed on those videos displaying "comet tail" artefacts. The context or clinical scenarios attached to each video, presented as vignettes or brief histories, offered insight into the severity of the illness of the subjects. This information was used to label this data, automatically avoiding other processing time from experts. After creating the label, images were fully anonymized1. The current dataset features a total of 30 lung US videos from 30 distinct subjects enrolled as patients in TiAssisto. These videos specifically highlight the presence or absence of signs of pulmonary disease detection; some of them depict patients with evident pulmonary oedema, while others display lungs with minimal or no B-lines at all. Images were extracted by the above frame-to-frame commented videos, obtaining the dataset of Table III. The labels were agreed upon with the involved physician and depicted how much space comet tails occupied in the image.

IV. Methods

The first step of the proposed approach involves a video annotation tool. An ad hoc implemented Java program enables frame-to-frame video processing, allowing medical staff to

TABLE I
DISTRIBUTION OF IMAGES IN THE DATASET

Label	Number of Images
0 : No B-Lines	183
1 : B-Lines in 30% FOV	754
2 : B-Lines in 50% FOV	196
3 : B-Lines in more then 50% FOV	76
Total	**1209**

annotate the entire dataset. All the acquired US images containing text are anonymised through an inpainting method, but only outside the line of view of US images, avoiding diagnostic content alterations. The used inpainting algorithm is based on the Keras Optical Character Recognition (OCR) module [14]. While traditional OCR systems rely on hand-engineered features and heuristics, deep learning-based approaches have vastly outperformed them by automatically learning relevant features from data [15]. Keras OCR utilizes different methods as the recognition one. This approach identifies a series of quadrilateral zones in the image in which text is presented, outputting four couple of points to identify the area with text. This detection is followed also by a recognition approach that uses a defined dictionary to recognize letters and words in the image; in this case, we are interested only in the detection part. The used structure is the one of Convolutional Recurrent Neural Networks (CRNN) [16]. Convolutional Neural Networks (CNNs), Figure 2, are commonly used for image classification tasks. For OCR, CNNs can be used to identify, and classify, each region of an image as a particular character or symbol, defining the bounding boxes. Given the sequential nature of texts, Recurrent Neural Networks, RNNs, are used to capture the characters. In this case, it is used a bi-directional Long-Short-Term-Memory, LSTM, architecture. When the algorithm ends scanning the image, every bounding box is sent to a method that evaluates the area and the position of the box and simply covers it with a mask whose dimension and position have already been evaluated with the same procedure.

The classification task is instead performed with a simple EfficientNet-b0 [17]. Data augmentation was still employed by avoiding any transformation that could alter the diagnostic content.

1) Resizing: The first transformation resizes any input image to a consistent dimension of 512x512 pixels.
2) Rotation: the images are subjected to a random rotation. Specifically, each image might be rotated by up to 20 degrees in either direction.
3) Random Affine Translation: The third transformation is a random affine transformation that moves the image slightly in one of the four primary directions (up, down, left, or right). However, the rotation is set to 0 degrees, meaning the image will not be rotated during this step. The translation is random up to a maximum of 10% of the image's height in the vertical direction.
4) Random Shearing: Following the translation, another affine transformation is applied for shearing the image. Shearing is a distortion that skews the image, introducing another form of variability. In this case, the shear is random and can be up to 10% in any direction.
5) Random Scaling and Resizing: Next, the image undergoes random scaling, where its size is adjusted between 80% to 120% of its current size. After this scaling operation, the image is again resized to 512x512 pixels to maintain a consistent size.

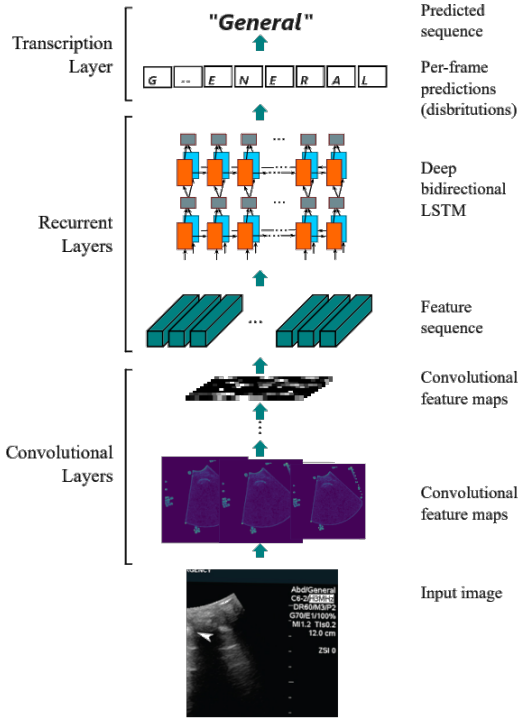

Fig. 2. OCR PIpeline implemented from Keras module

After the last transformation, the pixels of each image are tensored between $[0 < x < 1]$ and then undergo normalization. Mean and standard deviation are calculated separately for each of the three channels over the entire dataset. Given that the normalization occurs in the tensor space, the values are averaged over a pixel range of 255.

The EfficientNet and the network are then trained, using the method described in [18]: 5 weak EfficientNet-b0 models and then 5 ensembles inheriting the deep features of the best two weak models are validated and tested. Although ensemble methods provide higher classification performances, their architecture is generally more complex than their single counterparts. For this reason, they have been scarcely used in the past. Our method greatly simplifies ensembling using an adaptive efficient approach. As shown in Table II, our ensemble method [19] remarkably improves the results of EfficientNet-b0 weak learners. The main characteristic of our method is the linear combination of two convolutional blocks of already trained EfficientNets; out of the five weak models, only the best two are used to build the new ensemble architecture. The chosen layers are then frozen and only the

linear combination of the convolutional layer outputs is used. This operation is the concatenation of the features extracted by the two weak learners and their weights. At last, this new layer is fine-tuned to solve the classification task.

A patience method of ten epochs is used to avoid overfitting and optimize run time. The learning rate is fixed at 0.001 and AdaBelief optimizer is used.

The last step in the methodology pipeline is the application of Grad-CAM [20] saliency map algorithm on every image in the dataset after it's fed into the network. This procedure simply evaluates the gradient which activates the classification of the image and evaluates it over the feature maps that contribute more to the class activation [21].

V. Results

The autonomous process of anonymization of the dataset took around 45 minutes with an average process time for each image of 2 seconds: this delay is entirely compatible with a point-of-care application of the process. One of the two proposed methods, while promising, did not achieve complete accuracy in the classification of US images as in previous study [18]. The weak learners in Table II do not surpass the 95% benchmark in the cleaned or in the original dataset, while the ensemble average performance is 99% with a peak of perfect accuracies on the test set in Table III.

TABLE II
Test, Val, and Train Accuracies for Weak and Ensemble Models (Truncated to Two Decimal Places)

Run	Weak			Ensemble		
	Test	Val	Train	Test	Val	Train
1	0.92	0.89	1.00	1.00	0.99	0.99
2	0.91	0.90	1.00	1.00	0.99	0.99
3	0.91	0.90	1.00	1.00	0.99	0.99
4	0.91	0.90	1.00	0.99	1.00	0.99
5	0.91	0.90	1.00	0.99	0.99	0.99

TABLE III
Classification Report of the Best Ensemble

Class	Precision	Recall	F1-Score	Support
0	1.0	1.0	1.0	19
1-3	1.0	1.0	1.0	76
4-5	1.0	1.0	1.0	20
6-8	1.0	1.0	1.0	7
Accuracy			1.0	122
Macro avg	1.0	1.0	1.0	122
Weighted avg	1.0	1.0	1.0	122

There is not enough evidence to see if there is any benefit in training the cleaned image set against the original one. Images are processed with the Grad-CAM approach and then commented on by a team of physicians with a range of expertise between average and expert.

Overall, the physicians were satisfied with the first version of the classification system, agreeing with a significant portion of the explanations provided by the system.

In particular, the system was able to catch meaningful anatomical detail in the classes annotated by label 1 and label 3, Figures 3 and 4. While it seems to struggle in the identification of the label 2. This thesis is supported by the classification report from the test set in which we can see the poor classification performance on the same label, Figure 5.

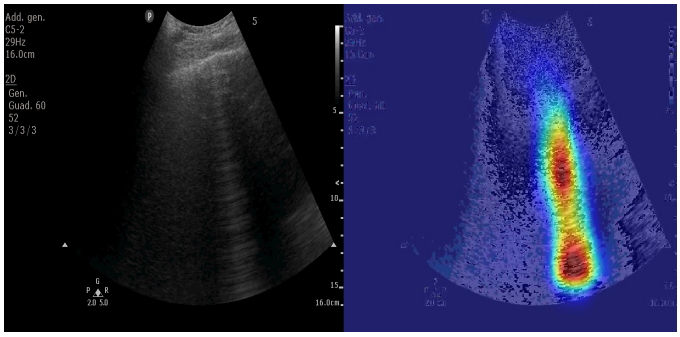

Fig. 3. The system seems to correctly detect the B-Line and almost segment it.

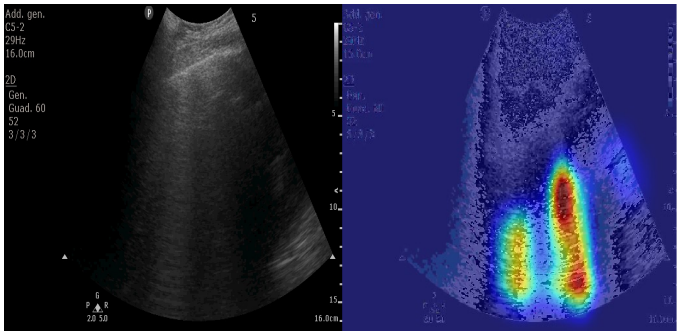

Fig. 4. The system seems to correctly detect the two B-Lines and almost segment it.

VI. Discussion/ Future work

The volume of US images collected during the project suggests a potential for further data acquisition. The TiAssisto welfare initiative encompasses a diverse group and is not exclusively focused on individuals with illnesses. While it doesn't readily provide a comprehensive medical database for benchmarking future studies, synergizing with data from related endeavours can furnish ample material. The proposed

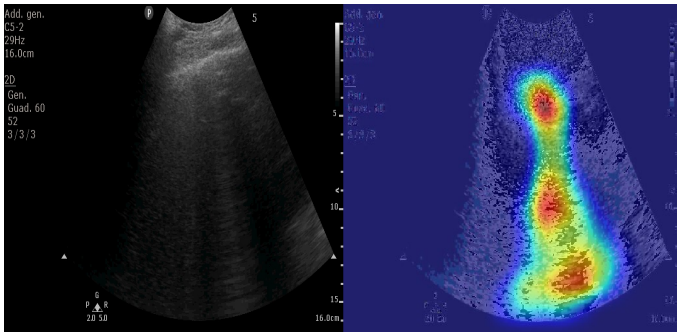

Fig. 5. The system is not able to correctly detect the two B-Line on the left side of the images, Label 2, being outperformed by physicians.

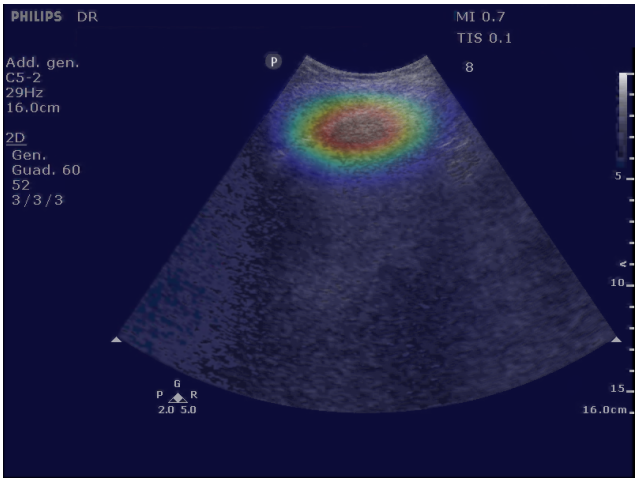

Fig. 6. The displayed ultrasound result is classified as a label 3 image, which should contain over 50% of comet line in the ultrasound line of view, while, after the GradCAM explainability inspection from physicians, no comet tail was detected. It seems anyway that the system is trying to classify the image as a label 0, i.e. no comet tail.

approach seems to be eligible to be applied in a POCUS telemedicine study. The reason for using EfficienNET architecture is vastly summarized by the literature. The network inverted bottleneck block, coupled with the compound scaling approach, defines the focal point of this architecture's efficiency. Notably, this configuration ensures that the network operates with optimal parameters at all times, minimizing computational overhead without compromising performance. The structure adeptly captures a diverse range of features from images, extracting both intricate details and broader patterns [22]. The proposed system's biggest downfall seems to be partly attributable to the dataset's status, which is still in the development stage and is still not heavily commented on. Nevertheless, the training algorithm provided some examples of robustness toward this error. As in Figure 6 labelled as 3, upon evaluation, the system appears to detect no evidence of comet tails, as observed by the attending physician, likely a case of incorrect human or automatic annotation.

The performance of the proposed ensemble architecture is in line with previous studies [18], [22]. Regarding Table II, comparable values are found for the not anonymised dataset. These highlight different aspects. Among them, the MBblock of EfficientNet seems to correctly extract the patterns of the US test without considering texts in case of the presence of a visible separation between image and text. Nothing can be said for images in which the text covers the field of view, since, in this situation, the deletion and reconstruction alter the diagnostic content of the test. Moreover, in many cases, ensembles are viable alternatives. Still, this approach allows the use of reduced training and computing capabilities. This ensembling method employs two already trained networks and fine-tunes a new final layer on the same task, achieving minimality, efficiency and adaptability requirements, since it uses the least number of parameters and uses an architecture trainable by gradient descent (linear combination layer) and therefore optimisable. The result in Table II highlights an improvement in classification since performance consistently rises in train, testing and validation among five different seeds. Even if Table III shows significant results, a limit is that support images are few, and further tests are needed using a larger dataset. To this aim, the sonographers of the project are annotating new ultrasounds we will use for additional experiments.

Other studies on autonomous classification or segmentation of medical images seem to be wrongly influenced by text information on the image [23], but that does not seem to be the case for our study in which the network appears not to be influenced by text presence. Privacy is a mandatory setting, and therefore this issue should be further investigated. Federated learning could be a possible solution since the possibility of running the model on a decentralized dataset grants privacy and enables blind studies on the image status impact on the model training. Another big improvement can be obtained with direct video segmentation and classification instead of a frame-by-frame approach. Vision transformer architectures have demonstrated a remarkable capacity for identifying temporal information in image sequences [24]. The problem with this kind of network is that it has a more significant number of parameters. It must be investigated if it is possible to find the proper set of hyperparameters to work in POC situations where resources can be limited. The video parsing and the federated learning approaches open the way to the last major improvement and future studies in this field which is continuous learning. Indeed, conventional models suffer from the "catastrophic forgetting" phenomenon when exposed to new data streams; continuous/lifelong learning models are instead architected to learn continuously [25]: they adapt and evolve, integrating new knowledge without obliterating the old. We believe that in the vast and dynamic world of healthcare, where patient demographics shift and treatment protocols change, continuous learning has enormous potential to improve medical deep learning algorithms.

REFERENCES

[1] S. Bhaskar, S. Bradley, S. Sakhamuri, S. Moguilner, V. K. Chattu, S. Pandya, S. Schroeder, D. Ray, and M. Banach, "Designing futuristic telemedicine using artificial intelligence and robotics in the covid-19 era," Frontiers in public health, p. 708, 2020.

[2] J. Narula, Y. Chandrashekhar, and E. Braunwald, "Time to add a fifth pillar to bedside physical examination: inspection, palpation, percussion, auscultation, and insonation," JAMA cardiology, vol. 3, no. 4, pp. 346–350, 2018.

[3] G. Volpicelli, M. Elbarbary, M. Blaivas, D. A. Lichtenstein, G. Mathis, A. W. Kirkpatrick, L. Melniker, L. Gargani, V. E. Noble, G. Via et al., "International evidence-based recommendations for point-of-care lung ultrasound," Intensive care medicine, vol. 38, pp. 577–591, 2012.

[4] F. Mojoli, B. Bouhemad, S. Mongodi, and D. Lichtenstein, "Lung ultrasound for critically ill patients," American journal of respiratory and critical care medicine, vol. 199, no. 6, pp. 701–714, 2019.

[5] Y. Tung-Chen, S. Ossaba-Vélez, K. S. Acosta Velásquez, M. L. Parra-Gordo, A. Díez-Tascón, T. Villén-Villegas, E. Montero-Hernández, A. Gutiérrez-Villanueva, Á. Trueba-Vicente, I. Arenas-Berenguer et al., "The impact of different lung ultrasound protocols in the assessment of lung lesions in covid-19 patients: Is there an ideal lung ultrasound protocol?" Journal of Ultrasound, pp. 1–9, 2022.

[6] L. Zhao and M. A. Lediju Bell, "A review of deep learning applications in lung ultrasound imaging of covid-19 patients," BME frontiers, vol. 2022, 2022.

[7] R. Arntfield, B. VanBerlo, T. Alaifan, N. Phelps, M. White, R. Chaudhary, J. Ho, and D. Wu, "Development of a convolutional neural network to differentiate among the etiology of similar appearing pathological b lines on lung ultrasound: a deep learning study," BMJ open, vol. 11, no. 3, p. e045120, 2021.

[8] A. Holzinger, M. Dehmer, F. Emmert-Streib, R. Cucchiara, I. Augenstein, J. Del Ser, W. Samek, I. Jurisica, and N. Díaz-Rodríguez, "Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence," Information Fusion, vol. 79, pp. 263–278, 2022.

[9] M. Hermann, C. Hafner, V. Scharner, M. Hribersek, M. Maleczek, A. Schmid, E. Schaden, H. Willschke, and T. Hamp, "Remote real-time supervision of prehospital point-of-care ultrasound: a feasibility study," Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, vol. 30, no. 1, p. 23, 2022.

[10] J. Diaz-Escobar, N. E. Ordóñez-Guillén, S. Villarreal-Reyes, A. Galaviz-Mosqueda, V. Kober, R. Rivera-Rodriguez, and J. E. Lozano Rizk, "Deep-learning based detection of covid-19 using lung ultrasound imagery," Plos one, vol. 16, no. 8, p. e0255886, 2021.

[11] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on mri," Zeitschrift für Medizinische Physik, vol. 29, no. 2, pp. 102–127, 2019.

[12] K. Jabeen, M. A. Khan, M. Alhaisoni, U. Tariq, Y.-D. Zhang, A. Hamza, A. Mickus, and R. Damaševičius, "Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion," Sensors, vol. 22, no. 3, p. 807, 2022.

[13] G. Ignesti, A. Bruno, C. Deri, G. D'Angelo, L. Bastiani, L. Pratali, S. Memmini, D. Cicalini, A. Dini, G. Galesi et al., "An intelligent platform of services based on multimedia understanding and telehealth for supporting the management of sars-cov-2 multi-pathological patients," in 2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, 2022, pp. 553–560.

[14] F. Chollet et al. (2015) Keras. [Online]. Available: https://github.com/fchollet/keras

[15] N. d. V. Dalarmelina, M. A. Teixeira, and R. I. Meneguette, "A real-time automatic plate recognition system based on optical character recognition and wireless sensor networks for its," Sensors, vol. 20, no. 1, p. 55, 2019.

[16] A. Onan, "Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 5, pp. 2098–2117, 2022.

[17] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020.

[18] A. Bruno, G. Ignesti, O. Salvetti, D. Moroni, and M. Martinelli, "Efficient lung ultrasound classification," Bioengineering, vol. 10, no. 5, p. 555, 2023.

[19] A. Bruno, D. Moroni, and M. Martinelli, "Exploring ensembling in deep learning," Pattern recognition and image analysis, no. 32, pp. 519–521, 2022.

[20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.

[21] F. Bodria, F. Giannotti, R. Guidotti, F. Naretto, D. Pedreschi, and S. Rinzivillo, "Benchmarking and survey of explanation methods for black box models," Data Mining and Knowledge Discovery, pp. 1–60, 2023.

[22] A. Bruno, D. Moroni, and M. Martinelli, "Efficient adaptive ensembling for image classification," arXiv preprint arXiv:2206.07394, 2022.

[23] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," 2015.

[24] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein et al., "The future of digital health with federated learning," NPJ digital medicine, vol. 3, no. 1, p. 119, 2020.

[25] O. S. Pianykh, G. Langs, M. Dewey, D. R. Enzmann, C. J. Herold, S. O. Schoenberg, and J. A. Brink, "Continuous learning ai in radiology: implementation principles and early applications," Radiology, vol. 297, no. 1, pp. 6–14, 2020.