Lung Ultrasound Pathology Classification for ICU Patient Management in LMIC

Phung Tran Huy Nhat[1,2], Hamideh Kerdegari[1], Angela McBride[2], Luigi Pisani[3], Nguyen Van Hao[4], Le Dinh Van Khoa[2], Shujie Deng[1], Le Ngoc Minh Thu[4], Duong Bich Thuy[4], VITAL Consortium[6], Marcus J. Schultz[4], Reza Razavi[1], Andrew P. King[1], Louise Thwaites[2,5], Sophie Yacoub[2,5], and Alberto Gomez[1]

[1] School of Biomedical Engineering Imaging Sciences, King's College London, UK
[2] Oxford University Clinical Research Unit, Ho Chi Minh city, Vietnam
[3] Mahidol Oxford Tropical Medicine Research Unit, Thailand
[4] Hospital of Tropical Diseases, Ho Chi Minh city, Vietnam
[5] Centre for Tropical Medicine and Global Health, University of Oxford
[6] a corporate author; the list of members is at the end of this manuscript.
{nhat.phung,alberto.gomez}@kcl.ac.uk

Abstract. Lung ultrasound (LUS) allows to assess lung condition in a wide variety of diseases, and can be particularly useful to manage patients in the intensive care unit, particularly in low and middle income countries (LMIC) where prevalence of infectious diseases such as dengue, malaria and sepsis is high. However, utilization of LUS is particularly challenging since it involves interpretation of artefacts in images caused by abnormalities in the lungs. As a result, efficient use of LUS is limited to expert LUS operators, very scarce in LMIC ICUs.
In this paper we propose an automatic method to classify the main 5 LUS features that are used to manage dengue and sepsis patients in the ICU: A-lines, B-lines, confluent B-lines, consolidation and pleural effusion. Using a deep neural network that combines a convolutional feature extractor submodule, a recurrent temporal assimilation submodule, and a classification submodule, our method can classify the aforementioned artefacts in LUS videos to an accuracy of 83.2%±10.5%.

Keywords: Lung ultrasound · Classification · Deep learning · LMIC

1 Introduction

In recent years, point-of-care ultrasound (POCUS) has been proved to be a useful imaging technique for bedside assessment of the critically ill patients for diagnosis and therapeutic management [21, 23]. POCUS lung ultrasound (LUS) plays a vital role in Intensive Care Unit in high-income countries as diagnostic and monitoring tool for the critically ill patients [18]. Moreover, LUS does not expose patients to radiation, and has also shown to be more sensitive and specific to diagnosing many pulmonary pathologies than chest x-ray (CXR) [4]. Ultrasound

is fast, low-cost and able to detect and diagnose lung problems such as diffuse pulmonary edema, cardiogenic edema, inflammatory interstitial lung diseases and pleural effusion, hence the potential for application of LUS in low and middle income countries (LMICs) can certainly increase [5, 15]. Respiratory distress in conditions most prevalent in LMICs such as malaria, dengue or sepsis and more recently covid-19. A major cause of respiratory distress in these common infectious diseases in LMICs is acute respiratory distress syndrome (ARDS) [9, 11, 8].

For example, although a recently performed meta-analysis [22] confirmed high sensitivity (96%) and specificity (93%) of LUS for detecting pneumonia in children, a common challenge in LMICS for daily management of pneumonia is the inability to regularly perform LUS on patients with suspected pneumonia. This limitation is mainly due to the unsustainability of traditional radiological equipment and expert in LMICs.

Lung ultrasound focuses on imaging artefacts, and mainly the artefacts produced at the pleural surface.These artefacts change over time with respiration cycle so identifying the right frame is challenging especially in critically ill patients with tachypnea (rapid, shallow breathing). In addition, like other ultrasound techniques, LUS is operator-dependent and requires training for image acquisition and interpretation. The lack of qualified ultrasound professionals, most likely due to lack of training program in developing countries, is an obstacle to the implementation of lung ultrasound.

Such challenges can be addressed by automatic computer-assisted classification and detection algorithms, such as the one proposed in this paper. In addition to reducing variability, automatic classification can help reduce the time spent by the physician on image interpretation, especially in the circumstances where resources are limited, and high trained healthcare staff is unavailable such as LMIC clinics and particularly the ICU.

Related work: Deep learning and machine learning methods are being increasingly investigated for many medical imaging applications, and particularly to assist LUS image analysis [6, 19, 1]. Most existing work covers one of two themes: binary classification of a single artefact (B-lines), or, since recently, multi-class classification for lung disease, focusing on COVID19.

B-line identification can be challenging skill for novice ultrasound users. Kerdegari et al [10] showed that the combination of a convolutional neural network (CNN) with a long short-term memory (LSTM) network and a temporal attention mechanism can classify videos into presence or absence of B-lines (to an accuracy of 81%), and temporally localise the frames which have B-lines in them (to an accuracy of 87%). However this method is limited to B-lines and does not consider other artefacts, as described above, which are crucial for efficient patient management in the ICU.

Most of the research related to classification of multiple artefacts in LUS were published recently focusing on COVID-19 patients. Roy at al [17] introduced several models to classify and localise COVID-19 markers (A-lines, vertical artifacts, consolidation and white lung) in POCUS lung ultrasound, and

included frame-based score, video-based score prediction and semantic segmentation. For video-based classification, the best model achieved F1, Precision and recall of 61, 70 and 60 percent, respectively. Other study conducted by Liu et al [14] proposed a new multi-symptom multi-label model with Active Learning methods [13] that are able to learn from less annotated data to classify multiple COVID-19 lung symptoms, achieving an accuracy of 100 %, 95.72 %, and 80.98 % for A-line, B-line and pleural lesion, respectively. Both studies have made the code and dataset publicly available. Arntfield et al [2] developed a CNN to classify LUS images with B-lines of different etiologies (COVID-19, Non-COVID respiratory distress syndrome and hydrostatic pulmonary oedema), and showed a better performance than physicians. Most of the previous studies mentioned above work on static LUS images that have been identified, selected and saved by the clinician. Two studies, conducted by Liu et al [14] and Born et al [3], propose video-based classification for COVID19 related image features and differentiate COVID19 from bacterial pneumonia and healthy subjects.To date, no method that can classify all 5 required features has been proposed in the literature.

Contributions:

The main contribution of this paper is to extend the model in [10] to perform multi-class classification of all five artefacts that are relevant for patient management in the ICU: A-line, B-line, confluent B-lines, consolidation and pleural effusion(examples shown in Figure 1). This constitutes the first video-based method for automatic LUS multi-artefact classification, and for the first time shows the potential to support lung function assessment in full, providing information about all clinically relevant features to ICU staff in LMIC for better management of dengue and sepsis patients.

2 Methodology

2.1 Dataset and Annotation

The dataset contains 3078 lung ultrasound videos (4 seconds) from 60 patients with dengue shock or septic shock admitted to the [*hospital name removed for anonymity*] between June 2019 and June 2020. The research was approved by the [*name removed for anonymity*] Ethics Committee and the [*name removed for anonymity*] Review Boards. LUS examination were carried out using Sonosite M-Turbo machine (Fujifilm Sonosite, Inc., Bothell, WA) with a low medium frequency (3.5-5 MHz) convex probe by qualified sonographers. LUS was performed using a standardised operating procedure based on the Kigali ARDS protocol [16]: assessment for B-lines [20, 12], consolidation and pleural effusion, performed at 6 points on each side of the chest (2 anterior, 2 lateral and 2 posterolateral).

These videos were annotated by expert sonographers using the VGG annotator tool [7]. Five lung patterns were selected for multi-class classification, as introduced above and illustrated in Fig. 1: A-line (normal lung), B-line, Confluent B-line, Consolidation, Pleural effusion. The distribution of overall data is shown in Table 1. Because the class imbalance distribution in the data, during

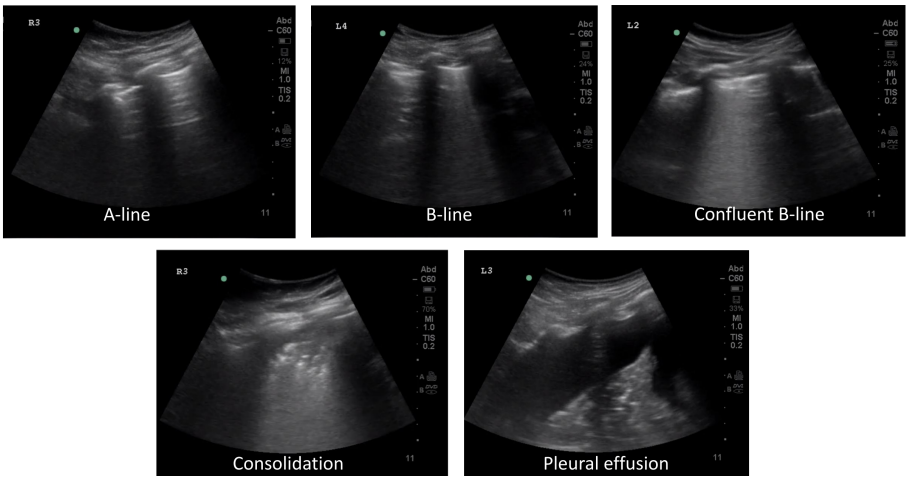

Fig. 1. Sample LUS images. A-line (healthy lung, horizontal reverberation artifacts), B-line (vertical artifacts deriving from the pleural line, moving synchronously with lung), Confluent B-line (many B-lines merge together), Consolidation (an echo-poor image juxtaposed to the pleural line (white arrow) and delimited by irregular boundaries), Pleural effusion (hypoechoic space between the parietal and visceral pleura)

training we weighted each class when computing the loss, and computed class weights w_c as: $w_c = \sum_c N_c/N_c$, where N_c is the number of samples per class.

Class	**Number of video**	**Weight,** w_c
A-line	1825	1.0
B-line	102	17.8
Confluent B-line	138	11.4
Consolidation	21	67.3
Pleural effusion	993	1.6

Table 1. Number of samples per class and weight values w_c of each class c.

AVI-format videos were cropped and masked to remove text and information outside of the scanning sector. The 640x480 pixels videos were down sampled using OpenCV into 64x64 pixels. For training, each four-second clip was converted into shorter clips of one second with an overlap of 20% between consecutive frames in the video.

2.2 Model architecture and training

Our proposed model is an extension from the model from [10] that was used for B-line classification and localization in LUS videos. The model architecture

is depicted in Fig. 2. It consists of three parts: convolutional neural network (CNN) to extract frame-wise spatial features, bidirectional long short-term memory (LSTM) network to extract temporal features from video and a temporal attention mechanism to weight up frames that contribute more to the classification task. In this paper, we replaced the classification subnet, after the temporal attention mechanism, with a fully connected layer (with ReLU non-linear activation and Dropout) and a 5-element final layer that produces a 1-hot 5D vector for 5 class classification.

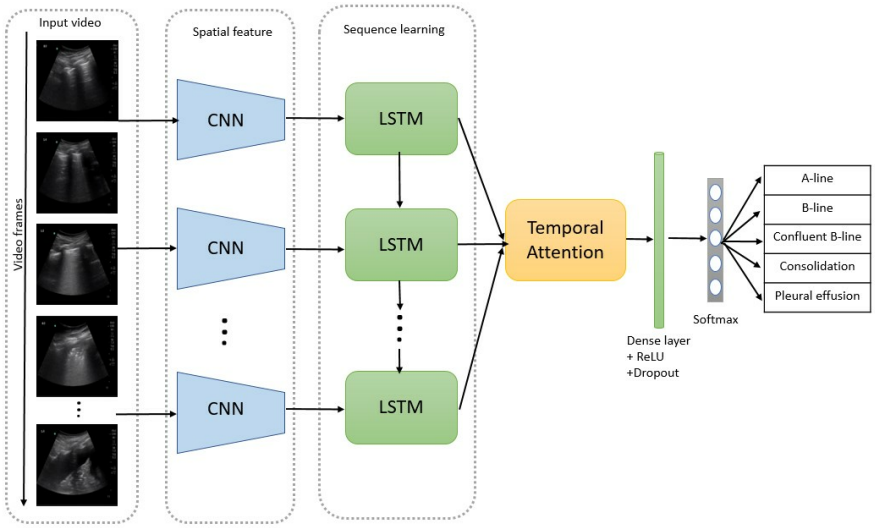

Fig. 2. Model architecture.

The model was implemented using Keras library with a Tensorflow backend. It was trained using the Adam optimizer with a learning rate of 0.001. A batch size of 16 and batch normalization were applied for both CNN and LSTM network layers. Dropout of 0.2 and L2 =10-5 for regularization were utilized. LUS video data were augmented by adding horizontally-flipped frames to the training set. We used 5-fold cross validation and trained the network for 100 epochs. After the models were trained, they were evaluated on the test set.

3 Experiments and evaluation

To assess the performance of the proposed method, we used the following five metrics on the test set: 1) Overall accuracy, calculated as the number of correctly classified clips as a fraction of the total number of clips; 2) Average accuracy,

calculated as the average over all lung pathologies of per-pathology accuracy; 3) F-score; 4) precision ; and 5) sensitivity (recall). Confusion matrices were calculated and are also reported.

4 Results

Our model achieved an overall test accuracy of 86.4%, with the test accuracy for each class A-line, B-line, Confluent B-line, Consolidation, Pleural Effusion are 89.2%, 77.5%, 84.8%, 66.7%, and 97.9%, respectively. This resulted in an average accuracy of 83.2%±10.5%. The F1-scores, precision and recall are reported in Table 2. The confusion matrix (in relative numbers), colour coded by value is shown in Table 3.

	F1	**Precision**	**Recall**
A-line	93%	98%	89%
B-line	70%	66%	77%
Confluent B-line	84%	82%	85%
Consolidation	64%	61%	67%
Pleural effusion	91%	86%	98%
Average (± std)	80%±11%	77%±13%	83%±10%

Table 2. Classification performance (F1-score, precision and recall) on the test set.

	A	B	CB	Cn	PE
A	0.89	0.02	0.01	0.0	0.08
B	0.1	0.77	0.04	0.02	0.07
CB	0.04	0.04	0.85	0.01	0.07
Cn	0.05	0.0	0.1	0.67	0.19
PE	0.02	0.0	0.0	0.0	0.98

Table 3. Confusion matrix, colour coded by value.

5 Conclusion and Discussion

We proposed a method to classify LUS videos into the main 5 features that are used in the ICU to manage patients with different diseases including dengue, malaria or sepsis: A-line (i.e. normal lung), B-line, confluent B-line, consolidation, and pleural effusion. The method achieved an average classification accuracy of 83%. This is the first method to provide accurate automatic classification for all lung features that are clinically relevant for these patients, and particularly aiming at a LMIC setting. Therefore, the proposed method has the potential to enable widespread, regular and reliable LUS assessment in LMICs ICUs.

Although we used a class-weighted loss, data imbalance had an impact on accuracy: accuracy was highest for LUS patterns with more training data (in this case A-line and Pleural effusion), and lowest for LUS patterns that have less training data, especially for B-line and Consolidation. Particularly, B-line was wrongly classified as A-line at 14 percent, the reason may be due to the fact that B-lines appear and disappear during the respiratory cycle.

The results are very promising towards proving usefulness in a resource-limited ICU. However, there are several challenges for clinical adoption that

need to be investigated: how will this method fit in the clinical workflow? How to reassure clinicians that the model predictions are trustworthy? And how can this technology be integrated within the constrains and IT infrastructure of a LMIC ICU? To address these issues, future work will include a clinical usability study; investigation of interpretability and explainability methods to provide insight into the prediction process; and integration of the algorithm into portable, small-footprint computers that can be connected to ICU ultrasound systems.

Acknowledgement

The VITAL Consortium: **OUCRU**: Sayem Ahmed, Dang Trung Kien, Doan Bui Xuan Thy, Dong Huu Khanh Trinh, Joseph Donovan, Du Hong Duc, Ronald Geskus, Ho Bich Hai, Ho Quang Chanh, Ho Van Hien, Hoang Minh Tu Van, Huynh Trung Trieu, Evelyne Kestelyn, Lam Minh Yen, Le Dinh Van Khoa, Le Thanh Phuong, Luu Phuoc An, Angela Mcbride, Nguyen Lam Vuong, Nguyen Quang Huy, Nguyen Than Ha Quyen, Nguyen Thanh Ngoc, Nguyen Thi Giang, Nguyen Thi Le Thanh, Nguyen Thi Phuong Dung, Ninh Thi Thanh Van, Pham Thi Lieu, Phan Nguyen Quoc Khanh, Phung Khanh Lam, Phung Tran Huy Nhat, Guy Thwaites, Louise Thwaites, Tran Minh Duc, Trinh Manh Hung, Hugo Turner, Jennifer Ilo Van Nuil, Sophie Yacoub. **Hospital for Tropical Diseases, Ho Chi Minh City**: Cao Thi Tam, Duong Bich Thuy, Ha Thi Hai Duong, Ho Dang Trung Nghia, Le Buu Chau, Le Ngoc Minh Thu, Le Thi Mai Thao, Luong Thi Hue Tai, Nguyen Hoan Phu, Nguyen Quoc Viet, Nguyen Thanh Nguyen, Nguyen Thanh Phong, Nguyen Thi Kim Anh, Nguyen Van Hao, Nguyen Van Thanh Duoc, Nguyen Van Vinh Chau, Pham Kieu Nguyet Oanh, Phan Thi Hong Van, Phan Tu Qui, Phan Vinh Tho, Truong Thi Phuong Thao. **University of Oxford**: David Clifton, Mike English, Shadi Ghiasi, Heloise Greeff, Jannis Hagenah, Ping Lu, Jacob McKnight, Chris Paton, Tingting Zhu. **Imperial College London**: Pantellis Georgiou, Bernard Hernandez Perez, Kerri Hill-Cawthorne, Alison Holmes, Stefan Karolcik, Damien Ming, Nicolas Moser, Jesus Rodriguez Manzano. **King's College London**: Alberto Gomez, Hamideh Kerdegari, Marc Modat, Reza Razavi. **ETH Zurich**: Abhilash Guru Dutt, Walter Karlen, Michaela Verling, Elias Wicki. **Melbourne University**: Linda Denehy, Thomas Rollinson.

References

1. Akkus, Z., Cai, J., Boonrod, A., Zeinoddini, A., Weston, A.D., Philbrick, K.A., Erickson, B.J.: A survey of deep-learning applications in ultrasound: Artificial intelligence–powered ultrasound for improving clinical workflow. Journal of the American College of Radiology **16**(9), 1318–1328 (2019)
2. Arntfield, R., VanBerlo, B., Alaifan, T., Phelps, N., White, M., Chaudhary, R., Ho, J., Wu, D.: 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 **11**(3), e045120 (2021)

3. Born, J., Wiedemann, N., Cossio, M., Buhre, C., Brändle, G., Leidermann, K., Aujayeb, A., Moor, M., Rieck, B., Borgwardt, K.: Accelerating detection of lung pathologies with explainable ultrasound image analysis. Applied Sciences **11**(2), 672 (2021)
4. Brogi, E., Bignami, E., Sidoti, A., Shawar, M., Gargani, L., Vetrugno, L., Volpicelli, G., Forfori, F.: Could the use of bedside lung ultrasound reduce the number of chest x-rays in the intensive care unit? Cardiovascular ultrasound **15**(1), 1–5 (2017)
5. Buonsenso, D., Rose, C.: Implementation of lung ultrasound in low-to-middle income countries: a new challenge global health? (2021)
6. Drukker, L., Noble, J., Papageorghiou, A.: Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. Ultrasound in Obstetrics & Gynecology **56**(4), 498–505 (2020)
7. Dutta, A., Gupta, A., Zissermann, A.: Vgg image annotator (via). URL: http://www. robots. ox. ac. uk/˜ vgg/software/via (2016)
8. Gibson, P.G., Qin, L., Puah, S.H.: Covid-19 acute respiratory distress syndrome (ards): clinical features and differences from typical pre-covid-19 ards. Medical Journal of Australia **213**(2), 54–56 (2020)
9. Graça, L., Abreu, I.G., Santos, A.S., Graça, L., Dias, P.F., Santos, M.L.: Descriptive acute respiratory distress syndrome (ards) in adults with imported severe plasmodium falciparum malaria: A 10 year-study in a portuguese tertiary care hospital. Plos one **15**(7), e0235437 (2020)
10. Kerdegari, H., Nhat, P.T.H., McBride, A., Razavi, R., Van Hao, N., Thwaites, L., Yacoub, S., Gomez, A.: Automatic detection of b-lines in lung ultrasound videos from severe dengue patients. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 989–993. IEEE (2021)
11. Kwizera, A., Dünser, M.W.: A global perspective on acute respiratory distress syndrome and the truth about hypoxia in resource-limited settings (2016)
12. Lichtenstein, D.A., Meziere, G.A.: Relevance of lung ultrasound in the diagnosis of acute respiratory failure*: the blue protocol. Chest **134**(1), 117–125 (2008)
13. Liu, B., Ferrari, V.: Active learning for human pose estimation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4363–4372 (2017)
14. Liu, L., Lei, W., Wan, X., Liu, L., Luo, Y., Feng, C.: Semi-supervised active learning for covid-19 lung ultrasound multi-symptom classification. In: 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI). pp. 1268–1273. IEEE (2020)
15. Riaz, A., Cambaco, O., Ellington, L.E., Lenahan, J.L., Munguambe, K., Mehmood, U., Lamorte, A., Qaisar, S., Baloch, B., Kanth, N., et al.: Feasibility, usability and acceptability of paediatric lung ultrasound among healthcare providers and caregivers for the diagnosis of childhood pneumonia in resource-constrained settings: a qualitative study. BMJ open **11**(3), e042547 (2021)
16. Riviello, E.D., Kiviri, W., Twagirumugabe, T., Mueller, A., Banner-Goodspeed, V.M., Officer, L., Novack, V., Mutumwinka, M., Talmor, D.S., Fowler, R.A.: Hospital incidence and outcomes of the acute respiratory distress syndrome using the kigali modification of the berlin definition. American journal of respiratory and critical care medicine **193**(1), 52–59 (2016)
17. Roy, S., Menapace, W., Oei, S., Luijten, B., Fini, E., Saltori, C., Huijben, I., Chennakeshava, N., Mento, F., Sentelli, A., et al.: 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 (2020)
18. Shrestha, G.S., Weeratunga, D., Baker, K.: Point-of-care lung ultrasound in critically ill patients. Reviews on recent clinical trials **13**(1), 15–26 (2018)

19. Van Sloun, R.J., Cohen, R., Eldar, Y.C.: Deep learning in ultrasound imaging. Proceedings of the IEEE **108**(1), 11–29 (2019)
20. Volpicelli, G., Elbarbary, M., Blaivas, M., Lichtenstein, D.A., Mathis, G., Kirkpatrick, A.W., Melniker, L., Gargani, L., Noble, V.E., Via, G., et al.: International evidence-based recommendations for point-of-care lung ultrasound. Intensive care medicine **38**(4), 577–591 (2012)
21. Volpicelli, G., Mayo, P., Rovida, S.: Focus on ultrasound in intensive care (2020)
22. Xin, H., Li, J., Hu, H.Y.: Is lung ultrasound useful for diagnosing pneumonia in children?: a meta-analysis and systematic review. Ultrasound quarterly **34**(1), 3–10 (2018)
23. Zieleskiewicz, L., Lopez, A., Hraiech, S., Baumstarck, K., Pastene, B., Di Bisceglie, M., Coiffard, B., Duclos, G., Boussuges, A., Bobbia, X., et al.: Bedside pocus during ward emergencies is associated with improved diagnosis and outcome: an observational, prospective, controlled study. Critical Care **25**(1), 1–12 (2021)