Case-based Image Retrieval in Point-of-care Lung Ultrasonography

Bárbara Teixeira[1]
up202209742@edu.fe.up.pt

João Pedrosa[1,2]
joao.m.pedrosa@inesctec.pt

Sandro Queirós[3]
sandroqueiros@med.uminho.pt

[1]Faculty of Engineering of the University of Porto (FEUP)
Porto, Portugal

[2]Institute for Systems and Computer Engineering,
Technology and Science (INESC TEC)
Porto, Portugal

[3]Life and Health Sciences Research Institute (ICVS)
School of Medicine, University of Minho, Braga, Portugal

Abstract

Point-of-care lung ultrasonography (LUS) plays a vital role in rapid lung assessment. Despite its utility in the clinical evaluation of pulmonary diseases, its use is currently limited due to the lack of qualified professionals to interpret these images. Therefore, this study harnesses content-based image retrieval (CBIR) techniques for LUS examinations, focusing on binary and multi-label classifications. Two pre-trained models were used for binary and multi-label classification, and feature-based retrieval was carried out. The used models demonstrate 88.3% binary retrieval accuracy, while achieving 65.6% accuracy for multi-label retrieval. Promising results have been achieved, but there remains a wealth of opportunities to explore, and it is of utmost importance to do so, in order to be able to accurately assist a health professional in their decision-making.

1 Introduction

Point-of-care lung ultrasonography (LUS) is a safe, low-cost, portable imaging technique practical in emergency rooms for a rapid bedside examination of a patient's lungs. Ultrasound is often considered not applicable for lung evaluation due to its air-filled structure. However, artifacts like the pleura as a horizontal hyperchoic line and hyperchoic A-lines can help assess lung health [2]. The absence of these artifacts is often associated with pathology, while B-lines, which represent higher-density tissues or fluids, are representative of pathologies. The severity of pathology is directly correlated with the quantity of B-lines [1]. Despite its clinical value in assessing pulmonary disease, limited availability of skilled professionals to interpret challenging ultrasound phenomena and artifacts constrains its current use. The ability to closely study similar cases and draw from prior diagnostic experience is essential for ultrasonographer training and daily decision-making. Developing image retrieval systems for LUS exams becomes crucial in creating this vital tool, especially given the scarcity of related studies.

Image retrieval is a crucial aspect of image matching, aiming to retrieve images from databases similar to a query image. Similarity calculation determines the performance of any image retrieval system, requiring discriminative, robust, and efficient computation [4]. Content-based image retrieval (CBIR) analyzes the content of images, including color, texture, shapes, and structure [5]. CBIR has experienced rapid research growth since the early 80s, with recent trends focusing on deep learning techniques to understand complex functions and map raw input data to content features, eliminating the need for human experts [4][5].

In this work, the objective was to investigate deep learning-based CBIR techniques for LUS videos.

2 Methods

2.1 Dataset

The dataset consists of 3649 LUS videos in which the frame rate varies from 25 to 60 frames per second, with most of the videos having 6 seconds each. All the videos were manually annotated by one of four medical doctors with point-of-care ultrasound experience with respect to the finding(s) present in the video. They are divided into two main groups: normal and indicative of an underlying pathology. The presence of any indicative finding overshadows the presence of a normal finding. Normal findings include scattering artifacts and A-lines patterns. Indicative findings include non-pathological negative B-lines (presence of less than 3 B-lines), pathological positive B-lines (3 or more B-lines or coalescent B-lines), and other pathologies (e.g., consolidation and pleural effusion). The dataset was then split into 80% for training and 20% for testing.

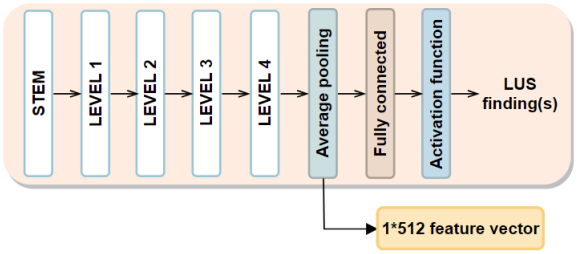

Figure 1: Feature representation extraction.

2.2 Automatic lung POCUS interpretation framework

This work uses a pre-trained model in [6] developed for LUS video classification, consisting of a pre-processing block and a supervised learning block. The pre-processing block standardizes input data by removing information outside the field-of-view and scaling the videos to 128×128 pixels. Each video was divided into 32-frame clips at 8 Hz, considering the 4 seconds needed for LUS examinations. In turn, the supervised block is a R2+1D convolutional neural network with 18 layers, based on ResNet. It factors 3D filters into spatial and temporal blocks, inserting shortcut connections to bypass layers and propagate input directly [8]. The R2+1D network receives as input the pre-processed LUS clip, applies a stem block, followed by four levels of residual convolutions. It then computes global spatio-temporal average pooling, resulting in a 1D vector of 512 features, followed by a fully connected layer and the activation function (softmax or sigmoid if multi-class or multi-label, respectively).

The network was trained for 2 classification scenarios: binary classification, in model C1, between normal and indicative findings, and for multi-label classification, in model M1, considering scattering, A-lines, B-negative, B-positive, and other pathologies. Normal findings cannot coexist with indicative findings, and only one B-line label can be selected.

2.3 Image retrieval

CBIR can be divided in two parts: 1) feature representation (fetch the vector that represents the clip) and 2) feature indexing (sort the vectors by order of similarity based on a given metric) [3]. The model described in section 2.2 was implemented as a feature extractor, by fetching feature vectors from the average pooling layer, as seen in Fig.1. First, a Z-score normalization was performed, given by $x' = \frac{x_i - \mu_i}{\sigma_i}$, resulting in zero mean and unit variance data, minimizing the effect of outliers and ensuring accurate data analysis. To obtain the closest clips to a query clip, the distance between a query clip and each of the other clips was calculated using the cosine similarity, which is given by:

$$Cosine\ distance = 1 - \frac{xy^T}{||x||.||y||}, \quad (1)$$

where x and y are row vectors of the two samples to be compared [7].

Finally, the first nearest neighbour, i.e. the training set sample whose feature vector presents the smallest distance towards the feature vector of the query sample from the test set, was selected as the reference clip most similar to the query clip to be retrieved.

	Accuracy	Recall
Normal	0.883	0.889
Indicative		0.879

Table 1: Class consistency in model C1.

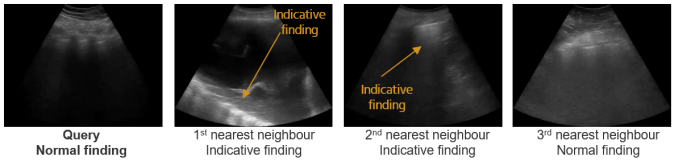

Figure 2: Example of retrieval for model C1.

	Accuracy	Recall
Scattering	0.656	0.646
A-lines		0.797
B-negative		0.499
B-positive		0.797
Other pathologies		0.513

Table 2: Class consistency in model M1.

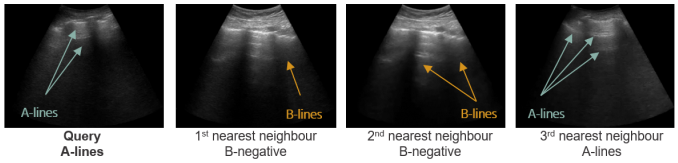

Figure 3: Example of retrieval for model M1.

2.4 Performance evaluation

To evaluate the performance of the retrieval, accuracy and recall were measured. The accuracy is defined as the number of correctly retrieved items out of the total number of items. Note that retrieval is considered correct if the closest selected training sample is from the same class as the query sample. Recall, in turn, is calculated as the number of well classified samples within the set of samples in each class.

3 Results and Discussion

3.1 Model C1

First, the dataset distribution was evaluated. A similar prevalence of both "normal" and "indicative" samples was observed across the entire dataset, which includes 11252 training samples and 3425 test samples (47% and 53%, respectively).

As seen in Table 1, the model's accuracy in retrieval is 0.883. In turn, the recall for each class shows a similar ability to retrieve clips from the database when query clips are of the "normal" class and "indicative" class (0.889 and 0.879, respectively).

Figure 2 shows an example of retrieval for model C1 in which the first 2 nearest neighbours retrieved have different labels from the query (indicative and normal, respectively), with the label of the third neighbour being the same as the query.

3.2 Model M1

Regarding class distribution, considering the complete dataset, which contains 11328 and 3368 training and test samples, respectively, a lower prevalence of "other pathologies" (13%) was observed, with "A-lines" and "B-positive" recording the highest prevalence (29% and 27%, respectively). In turn, the classes "scattering" and "B-negative" both have a prevalence of 15%.

The performance of the retrieval model was evaluated by calculating its general accuracy and recall for each of the 5 classes, as shown in Table 2. In this multi-label problem, a correct retrieval occurs when the label profile of the sample selected from the training set is equal to the label profile of the query. The model's accuracy is 0.656, which is lower than model C1. This may indicate that the model performs worse in multi-label classification scenarios than in binary classification scenarios. However, this task is more complex, as it aims to retrieve a sample with several identical findings to the query sample. The recall for each class is lowest for class "B-negative" and highest for classes "A-lines" and "B-positive". Distinguishing clips with A-lines or many B-lines is easier, but distinguishing clips with few B-lines or scattering is more difficult due to subjectivity of labeling. The smaller number of training samples may also affect the results for "other pathologies".

In the example of retrieval for model M1, in Figure 3, the query presents A-lines, while the first 2 nearest neighbours present negative B lines and the third neighbour has A-lines.

4 Conclusions

This project explored an image retrieval method for LUS exams using a pre-trained classifier. The approach included feature extraction, cosine similarity-based distance calculation between query and training set clips, and selection of the most similar clip. The model was trained for binary (C1) and multi-label (M1) classification. Results demonstrated superior retrieval ability for binary classification. In M1, retrieval was weaker for "B-negative" clips, but stronger for "A-lines" and "B-positive" classes. Retrieving "B-positive" clips is vital for pathology detection. Improving video retrieval is essential for accurate LUS exam evaluation by healthcare professionals and aiding in similar case assessment.

Acknowledgements

This work was funded by National Funds, through the FCT – Portuguese Foundation for Science and Technology, within the scope of project THOR DSAIPA/AI/0083/2020, LA/P/0063/2020 and PTDC/EMD-EMD/1140/2020, grant CEECIND/03064/2018 (S.Q.) and grant application reference 2022.06138.CEECIND (J.P.).

References

[1] Giacomo Baldi, Luna Gargani, Antonio Abramo, Luigia D'Errico, Davide Caramella, Eugenio Picano, Francesco Giunta, and Francesco Forfori. Lung water assessment by lung ultrasonography in intensive care: A pilot study. *Intensive Care Medicine*, 39(1):74–84, jan 2013. ISSN 03424642.

[2] Luna Gargani and Giovanni Volpicelli. How i do it: Lung ultrasound. *Cardiovascular Ultrasound*, 12(1):1–10, jul 2014. ISSN 14767120.

[3] Kazuma Kobayashi, Ryuichiro Hataya, Yusuke Kurose, Mototaka Miyake, Masamichi Takahashi, Akiko Nakagawa, Tatsuya Harada, and Ryuji Hamamoto. Decomposing normal and abnormal features of medical images for content-based image retrieval of glioma imaging. *Medical Image Analysis*, 74:102227, dec 2021. ISSN 1361-8415.

[4] Afshan Latif, Aqsa Rasheed, Umer Sajid, Jameel Ahmed, Nouman Ali, Naeem Iqbal Ratyal, Bushra Zafar, Saadat Hanif Dar, Muhammad Sajid, and Tehmina Khalil. Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review. 2019.

[5] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23, feb 2004. ISSN 1386-5056.

[6] Bárbara Malainho Pereira. *Automatic interpretation of point-of-care lung ultrasound*. Master's thesis, University of Minho, 2022.

[7] Mohammed Senoussaoui, Patrick Kenny, Themos Stafylakis, and Pierre Dumouchel. A study of the cosine distance-based mean shift for telephone speech diarization. *IEEE Transactions on Audio, Speech and Language Processing*, 22(1):217–227, 2014. ISSN 15587916.

[8] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann Lecun, and Manohar Paluri. A Closer Look at Spatiotemporal Convolutions for Action Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 6450–6459, dec 2018. ISSN 10636919.