Detecting COVID-19 infection from ultrasound imaging with only five shots: A high-performing explainable deep few-shot learning network

Jessy Song[1], **Ashkan Ebadi**[1,2], **Adrian Florea**[3], **Pengcheng Xi**[1,4], **Alexander Wong**[1,5]
[1]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[2]Digital Technologies Research Centre, National Research Council Canada, Montreal, QC H3T 2B2, Canada
[3]Department of Emergency Medicine, McGill University, Montreal, QC H4A 3J1, Canada
[4]Digital Technologies, National Research Council Canada, Ottawa, Ontario, Canada
[5]Waterloo Artificial Intelligence Institute, Waterloo, ON N2L 3G1, Canada

Abstract

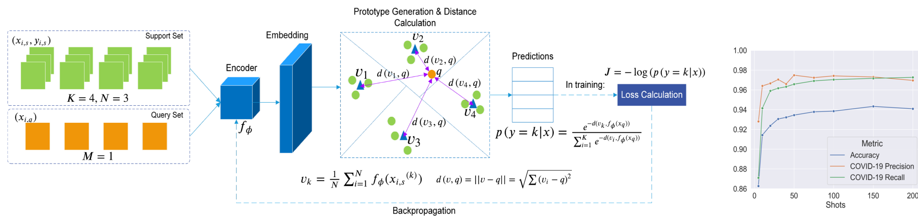
Applications of deep learning solutions, which are usually trained with large amount of dataset, in controlling the spread of Coronavirus Disease 2019 (COVID-19) have shown promising results. Motivated by the lack of large number of well-annotated dataset during the onset of a novel disease, we present a high-performing, interpretable few-shot learning network that detects positive COVID-19 cases with limited examples of ultrasound images. Extensive experiments are conducted to evaluate model performance under different encoder architectures, number of training shots and classification problem complexity. When trained with only 5-shots, network classifies between positive and negative COVID-19 cases with 99.3% overall accuracy, 99.5% recall and 99.25% precision for positive cases. Network explainability is evaluated with two visual explanation tools and reviewed by a practicing clinician to ensure validity of network's decision-making process.

1 Introduction

The adoption of rapid and effective screening methods to prevent further spread of the COVID-19 and lessen the burden on healthcare providers continues to be a necessity. Among all imaging modalities, point-of-care ultrasound (POCUS) imaging is the most accessible and least expensive, and it allows radiologists to identify symptoms and assess severity through visual inspection of the chest ultrasound images [1]. Applications of deep learning in medical image analysis have also shown promising results lately, suggesting that artificial intelligence (AI)-based solutions can accelerate the diagnosis of COVID-19. Multiple works have demonstrated the effectiveness of deep learning in the classification of the computed tomography (CT) scans and X-ray images [2; 3; 4; 5], showing above 97% precision/recall in detecting COVID-19 positive cases. However, deep learning requires a large set of training examples [3; 6; 7].

Due to the nature of novel diseases, the availability of such a huge amount of well-annotated data poses a great challenge. Motivated by this, we present an explainable deep prototypical network that learns to detect COVID-19 positive cases with high precision and recall from a very limited number of lung ultrasound (LUS) images. Our contribution is at least two folds: (1) when trained with only 5 shots, the network classifies between positive and negative COVID-19 cases with 99.3% overall accuracy, 99.5% recall and 99.25% precision for COVID-19 positive cases; (2) to ensure fairness and accountability, the network benefits from an explainability module, assessing decisions with visual explanation tools, i.e., Grad-CAM [8] and GSInquire [9]. Intensive experimentation was conducted (e.g., different image encoders, varying training shots and number of classes) to assess the

36th Conference on Neural Information Processing Systems (NeurIPS 2022).



a) b)

Figure 1: a) Network design. b) Example network performance with increasing shots

performance of the network. Moreover, our contributing clinician carefully verified and validated the pipeline and produced results to ensure the validity of the proposed solution from the clinical perspective.

2 Data and Methods

The COVIDx-US dataset v1.4. [10] is used for this study. The dataset contains 242 videos and 29,651 processed LUS images of patients with COVID-19 infection, non-COVID-19 infection, other lung conditions, and normal control cases. COVIDx-US provides images captured with two kinds of probe, linear probe which produces a square/rectangular image, or convex probe which has a wider field of view [11]. Due to the difference in field of view and low numbers of COVID-19 positive images with linear probe, linear probe data are excluded to reduce noise for the network. A total number of 25,262 convex LUS images are then randomly split into train set containing 90% of images in each class and test set with the remaining 10% of images. All images are rescaled to 224×224 pixels. The dataset is further augmented by rotating each image by 90°, 180°, 270°, resulting in a total of 101,048 images (25,262*4). Image rotation is an appropriate augmentation technique for our problem, as it keeps the images and areas of interest for clinical decisions unaltered and in-bound [12].

Few-shot learning is an approach where the model learns from a limited number of observations [13]. As a metric learning few-shot model, a prototypical network expects that there exists an embedding space in which points cluster around a single prototype representation for each class, and unlabelled data can be classified based on closeness to the class prototypes [13]. We define the problem as a K-way N-shot classification task, where K denotes the number of different classes and N denotes the number of examples presented for each class in the dataset. The few-shot classification with a prototypical network can be summarized into three steps: (1) encoding of the images in a query set of M unlabelled images and a support set of N labelled images, (2) generating class prototypes from support sets, and (3) assigning labels to query samples based on the distance to the class prototypes. Figure 1 shows the network architecture and sample performance results. During the training phase, a convolutional neural network encoder is trained to transform images to an embedding space and learn the non-linear mapping. Class prototypes are then computed as the mean of its support set vectors. The squared Euclidean distances between class prototypes to a query point are computed, and the SoftMax function is applied over distances to the prototypes to compute the probabilities of the query image being in each class. The network learns by minimizing the loss term, i.e., the negative log-SoftMax function. An episodic approach is used to train the network, where in each training episode, the few-shot task is simulated by sampling the data point in mini-batches.

To facilitate effective training process and prevent over-fitting, early stopping is implemented to stop the training process when loss term is not improved after 5 epochs. A total of 10 epochs is set for all training processes and 200 episodes is set for each training epoch. The performance of the network is evaluated using the overall accuracy, and the precision and recall for the COVID-19 class. To comprehensively assess the performance of the proposed network in detecting COVID-19 positive cases from ultrasound images, we evaluated the impact of various training conditions such as (1) image encoders, (2) number of shots available for training, and (3) classification task types

Experiment Parameters		*Description*
Models	1	Pre-trained ResNet 18 with ImageNet, with trainable parameters on the last 4 convolutional layers and final connected layer
	2	Pre-trained ResNet 50 with ImageNet, with trainable parameters on the last 3 convolutional layers and final connected layer
Training Shots		5, 10, 20, 30, 40, 50, 75, 100, 150, 200
Classification Formulation	2-way	Data from all 3 other classes, namely 'normal' class, 'non-COVID-19' class and 'other' class, are viewed as a combined COVID-19 negative class. The network learns to differentiate COVID-19 positive and negative cases.
	3-way	The 'other' class is excluded, as it contains data from multiple lung conditions which results in high in-class variations and may disrupt network's learning process due to the lack of uniformity in the data.
	4-way	The network remains the original dataset setting, and network is trained to classify 'COVID-19', 'normal', 'non-COVID-19' and 'other' class.

a)

		5-shot			*100-shot*		
Classification Type	Model	Accuracy	Precision	Recall	Accuracy	Precision	Recall
2 classes	1	0.993	0.9925	0.995	1	1	1
	2	0.9965	0.9967	0.997	0.9999	0.9999	1
3 classes	1	0.9987	0.9992	0.997	1	1	1
	2	0.9947	0.9942	0.994	1	1	1
4 classes	1	0.9817	0.9975	0.997	0.9884	1	1
	2	0.985	0.9917	0.993	0.9902	1	1

b)

Figure 2: a) Experiment parameters description. b) Classification results for 5-shot and 100-shot conditions.

with increasing class numbers indicates increasing problem complexity. Parameters are described in Fig.2a.[1].

3 Experiments and Results

Across all classification types and models, performance metrics increases from 5-shot and plateaus after 75-shot (ex. Fig.1b). ResNet networks [14] demonstrate the ability to classify COVID-19 with precision and recall above 98% consistently under both 5-shot and above 99% under 100-shot condition (Fig.2b). The increasing classes in 3-way and 4-way classification types reduces the performance of the network as problem complexity increases, but it is compensated as the number of shots increases, since more training examples improves class prototypes' representativeness. Deeper model (i.e., model 2 with ResNet50 as encoder) with re-trained final convolutional layers parameters perform the best in all classification types and shot conditions. Moreover, the 'COVID-19 infection' class is classified with the highest precision and recall among all classes, with 'normal' and 'non-COVID-19 infection' precision and recall being 2-3% lower and the 'other' class having the lowest precision and recall. This is expected since the 'other' class covers various lung conditions.

In order to clinically validate the network output and ensure that it captures clinically relevant patterns in the ultrasound images, our contributing clinician reviewed a randomly selected set of images. Fig.3 presents two Grad-CAM and GSInquire annotated COVID-19 ultrasound example images, enclosing two most common patterns deemed as critical for the final classification decision made by the network. Grad-CAM generates a visual explanation of the input image using the gradient information flowing into the last convolutional layer of the encoder, assigns importance values to each neuron for making a classification decision and outputs a heatmap-overlayed image showing regions that strongly impact the network decision [8]. The other tool GSInquire identifies the critical factors in an input image that are shown to be integral to the decisions made by the network in a generative synthesis approach [9]. As seen, the annotated images identified (1) the lung pleura region at the top (Fig.3(1)), and (2) the bottom region (Fig.3(2)) to have high importance. Clinically, B-lines, or the light comet-tail artifacts extending from pleura to the bottom of the image corresponds to signs of lung consolidation and are indicators of abnormality [15]. Hence, the visual annotations for Fig.3b more closely represent disease-related patterns within the ultrasound image. While the pleura regions at the top of the images are not entirely irrelevant features, the model might have mistaken these dark regions as signs of consolidation. To improve model explainability, we experimented with removing the pleura region of the images by cropping so that network focuses on the disease-defining features at mostly the bottom of the images, and the strategy demonstrates initial improvements in annotations from both explainability tools.

4 Conclusion and Limitations

In this work we propose a deep prototypical network, tailored to detect COVID-19 infection with high accuracy from very few ultrasound images. The proposed network leverages fine-tuned pretrained models to achieve high classification performance when only 5 examples from each class are available for training. Accuracy, precision and recall for the best performing network are over 99%, which

[1]Due to limit in space, not all encoder types are presented in this abstract and tables are kept in figure format.

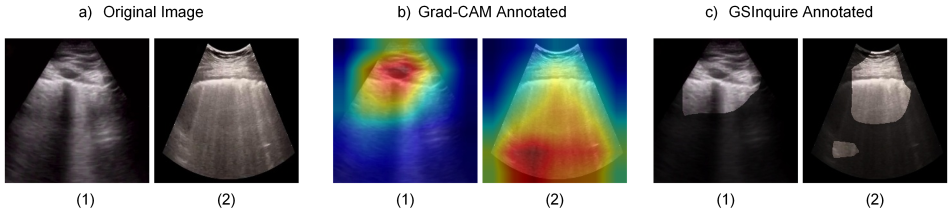


Figure 3: Two COVID-19 positive case examples correctly classified by the network with high confidence, a) original images, b) Grad-CAM annotated images and c) GSInquire annotated images.

are comparable or outperforming other existing work [7; 16]. For example, the work MetaCOVID [17] which also applied a few-shot approach for ultrasound COVID-19 detection, achieved 95.6% accuracy under 3-way, 10-shot setting. Predictive behaviour of our network is also consistent with clinical interpretation, as validated by our contributing clinician.

Several research directions can be explored to further improve the network. First, image augmentation and preparation techniques can be experimented to include linear probe data (which are excluded in this study) and increase the data size. Second, in this work, we experimented with simple cropping to remove the pleura region of the images. A more procedural image segmentation step could be added to include only clinically relevant areas of the images for network construction to further improve network explainability. Lastly, we used a public dataset (i.e., COVIDx-US) that includes data of various sources and quality.

5 Potential Societal Impact

We used the COVIDx-US public dataset for this study. As a potential societal impact, the proposed network and the published model could be used by healthcare providers to augment their decision and facilitate patients' screening, especially during initial onset of new diseases/pandemics where not many data is available. Based on our intensive experiments, although the evaluation results look promising, the model is not yet a production-ready solution and could be only considered as an AI-enabled tool to assist professional medical practitioners and not to replace them. Moreover, the network should also be tested on (at least) another external dataset to ensure network decisions are unbiased.

References

[1] H. Kasban, "A comparative study of medical imaging techniques," *International Journal of Information Science and Intelligent System*, vol. 4, pp. 37–58, 03 2015.

[2] H. Aboutalebi, M. Pavlova, H. Gunraj, M. J. Shafiee, A. Sabri, A. Alaref, and A. Wong, "Medusa: Multi-scale encoder-decoder self-attention deep neural network architecture for medical image analysis," 2021. [Online]. Available: https://arxiv.org/abs/2110.06063

[3] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, p. 19549, Nov 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-76550-z

[4] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in Biology and Medicine*, vol. 121, p. 103792, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482520301621

[5] H. Gunraj, A. Sabri, D. Koff, and A. Wong, "Covid-net ct-2: Enhanced deep neural networks for detection of covid-19 from chest ct images through bigger, more diverse learning," *Frontiers in Medicine*, vol. 8, p. 729287, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmed.2021.729287

[6] H. Gunraj, L. Wang, and A. Wong, "Covidnet-ct: A tailored deep convolutional neural network design for detection of covid-19 cases from chest ct images," *Frontiers in Medicine*, vol. 7, p. 1025, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmed.2020.608525

[7] A. MacLean, S. Abbasi, A. Ebadi, A. Zhao, M. Pavlova, H. Gunraj, P. Xi, S. Kohli, and A. Wong, "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 *FAIR-MICCAI'21*, 2021. [Online]. Available: https://arxiv.org/abs/2108.03131

[8] 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 *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.

[9] Z. Q. Lin, M. J. Shafiee, S. Bochkarev, M. S. Jules, X. Wang, and A. Wong, "Do explanations reflect decisions? A machine-centric strategy to quantify the performance of explainability algorithms," *CoRR*, vol. abs/1910.07387, 2019. [Online]. Available: http://arxiv.org/abs/1910.07387

[10] A. Ebadi, P. Xi, A. MacLean, A. Florea, S. Tremblay, S. Kohli, and A. Wong, "Covidx-us: An open-access benchmark dataset of ultrasound imaging data for ai-driven covid-19 analytics," *Frontiers in Bioscience-Landmark*, vol. 27, no. 7, 2022.

[11] W. Lee and Y. Roh, "Ultrasonic transducers for medical diagnostic imaging," *Biomedical Engineering Letters*, vol. 7, no. 2, p. 91–97, 2017.

[12] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, "Differential data augmentation techniques for medical imaging classification tasks," Apr 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977656/

[13] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[15] D. A. Lichtenstein, G. A. Mezière, J.-F. Lagoueyte, P. Biderman, I. Goldstein, and A. Gepner, "A-lines and b-lines: lung ultrasound as a bedside tool for predicting pulmonary artery occlusion pressure in the critically ill," *Chest*, vol. 136, no. 4, p. 1014–1020, 2009.

[16] M. Karnes, S. Perera, S. Adhikari, and A. Yilmaz, "Adaptive few-shot learning poc ultrasound covid-19 diagnostic system," 2021. [Online]. Available: https://arxiv.org/abs/2109.03793

[17] M. Shorfuzzaman and M. S. Hossain, "Metacovid: A siamese neural network framework with contrastive loss for n-shot diagnosis of covid-19 patients," *Pattern Recognition*, vol. 113, p. 107700, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320320305033