

Open World Active Learning for Echocardiography View Classification

Ghada Zamzmi^{*a}, Tochi Oguguo^{*a}, Sivaramakrishnan Rajaraman^a, and Sameer Antani^a

^aComputational Health Research Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

ABSTRACT

Existing works for automated echocardiography view classification are designed under the assumption that the views in the testing set must belong to a limited number of views that have appeared in the training set. Such a design is called closed world classification. This assumption may be too strict for real-world environments that are open and often have unseen examples, drastically weakening the robustness of the classical view classification approaches. In this work, we developed an open world active learning approach for echocardiography view classification, where the network classifies images of known views into their respective classes and identifies images of unknown views. Then, a clustering approach is used to cluster the unknown views into various groups to be labeled by echocardiologists. Finally, the new labeled samples are added to the initial set of known views and used to update the classification network. This process of actively labeling unknown clusters and integrating them into the classification model significantly increases the efficiency of data labeling and the robustness of the classifier. Our results using an echocardiography dataset containing known and unknown views showed the superiority of the proposed approach as compared to the closed world view classification approaches.

Keywords: Open world, classification, echocardiography, active learning, clustering

1. INTRODUCTION

Echocardiography is a common medical examination that uses high-frequency ultrasound waves to visualize the size and shape of different cardiac regions.¹ A comprehensive echocardiography examination involves imaging the heart in three main modes (Doppler, M-mode, and B-mode) and from different planes and orientations (views). Each echocardiography study contains up to a few hundred echocardiograms acquired to assess the function of the heart's chambers and valves.¹ The first step to ensure an accurate interpretation of echocardiograms is to identify the desired cardiac view from a list of all modes and views.¹ This step is time-consuming and non-trivial as several views differ only subtly from each other. Therefore, computational methods for echocardiography views identification have become an ideal solution for speeding up clinical workflow and obtaining fully automated echocardiogram interpretation in clinical practice.

1.1 Related Work

Several methods can be found in the literature²⁻⁷ for echocardiography modes and views classification. We divide these methods, based on the environment or setting, into closed world and open world methods. Closed world methods are designed for a static environment, which means the classifier is designed under the assumption that all test classes (views) are known (or seen) at training time. All existing methods for echocardiography view classification belong to this group of methods. These conventional methods use classical machine learning (e.g., support vector machine [SVM]) or deep learning (e.g., convolutional neural networks [CNNs]) techniques with datasets containing limited sets of known echocardiography views. For example, Wu et al.³ presented one of the first conventional methods for classifying echo images into eight views including parasternal long-axis view (PLAX) and apical 4 chambers (A4C). Their method extracts spectral energy features from the images using a GIST descriptor. These features are then used to train a SVM for classification. Other methods use SVM

* indicates joint first authors; current affiliation of T. O. is University of Maryland, Baltimore County; send correspondence to G.Z. E-mail: alzamzmiga@nih.gov.

with handcrafted descriptors such as Scale-invariant feature transform (SIFT),² histogram of oriented gradients (HOG),⁴ and bag of visual words (BoWs).⁵ Recent works for echo view classification use state-of-the-art CNNs such as VGG⁸ and ResNet.⁹ For example, Zhang et al.⁶ used a VGG-based model to classify echo images into 23 views including PLAX and A4C. Similarly,⁷ used a VGG-based method for view classification. Their method classifies echo images into 3 modes: B-mode (12 views), M-mode, and Doppler (2 views). We refer the reader to^{10,11} for comprehensive reviews of existing methods for mode and view classification. Although the classical computational methods have already achieved significant success in mode and views classification, these methods are only limited to the known classes (views) that have been learned during training. This drastically weakens their robustness when deployed in open-world settings with unknown or unseen views. Further, obtaining sufficient training data for all cardiac views is challenging due to the lack of training datasets that contain all cardiac views (> 15 views) and subviews (e.g., parasternal short axis view alone has > 5 subviews). Since the current classical view classification methods randomly classify unknown views as one of the known views learned during training, there is a need for a robust open world classifier that labels images of known views into their respective classes while actively recognizing unknown views and learning them (open world active learning). While open world learning has been studied in a variety of machine learning tasks, this approach has not been explored with echocardiography classification. In fact, we are not aware of any work that utilizes open world active learning with medical ultrasound imaging applications, including echocardiography images.

1.2 Contribution

In this work, we propose an open world active learning framework for echocardiography view classification. Specifically, we frame echocardiography view classification as an open set learning problem, where known views are correctly classified into their respective classes and unknown views are labeled as unknown. We train the open world classifier using an initial dataset with known views. To actively update the initially trained open world classifier, the set of unknown views is grouped into different clusters using a clustering algorithm. Then, each cluster is labeled by an echocardiologist into their respective views. Finally, the new labeled views (previously unknown) are added to the initial set of known views and used to update the open world classifier. This allows the use of an initial classifier trained with a relatively limited amount of labeled views while actively recognizing new views (unknowns) that it has not seen before and learning them incrementally to become more and more knowledgeable about the cardiac views. To the best of our knowledge, this work is the first to propose an open world active learning framework for echocardiography view classification.

The rest of this paper is organized as follows. Section 2 provides technical background including definitions of open world, closed world, and cluster-based active learning. Section 3 presents the datasets utilized in this work as well as the proposed framework for open world and active echocardiography view classification. Section 4 presents our experimental results followed by the conclusion and discussion of future directions in Section 5.

2. BACKGROUND

2.1 Closed vs Open World Classification

Although machine learning-based techniques achieved high performance on several visual recognition tasks including image classification^{12,13} and segmentation,^{14,15} the majority of these methods are designed to only learn images belonging to a predefined set of classes given before training (closed world setting). Mathematically, a traditional closed world classifier is trained with $D_{train} = \{(x_i, y_i)\}_{i=1}^N$ and tested with $D_{test} = \{(x_i, y_i)\}_{i=1}^M$, where $x_i \in \mathbb{R}^D$ and $y_i \in Y = \{1, 2, ..C\}$ is a finite set of predefined classes. In closed world setting, we assume that both D_{train} and D_{test} are drawn from the same distribution, and the classifier is trained using D_{train} to minimize an empirical loss function such as cross-entropy.¹⁶ This loss function is optimized to discriminate between known classes. Finally, the trained closed world classifier is tested using D_{test} to label a new image as one of the known classes in Y .

Although the closed world assumption holds in several applications, the majority of real world applications are dynamic and open containing examples from classes that might not appear in training.¹⁷ In such a setting, a closed world classifier would classify an unseen example as one of the known classes. Since the cost of randomly misclassifying an unseen image to a known class can be high, especially in clinical practice, there is a need to design

robust classifiers for open world settings. In such setting, the classifier is still trained using $D_{train} = \{(x_i, y_i)\}_{i=1}^N$, where $x_i \in \mathbb{R}^D$. However, $D_{test} = \{(x_i, y_i)\}_{i=1}^M$ has a set (Y) containing predetermined classes and unknown classes; i.e., $y_i \in Y = \{1, 2, \dots, C, C + 1\}$, where C represents the known classes and $C + 1$ represents the new classes. Similar to the close world classifier, the open world classifier is trained to minimize a loss function (e.g., cross-entropy) with an overall aim to recognize known classes and reject unknown classes or classify them as $C + 1$.

2.2 Deep Open Classification

Open world learning has been integrated into convolutional neural networks (CNNs) to create robust deep open classification (DOC) models.^{18–21} In,¹⁸ Shu et al. integrated open-world learning into CNNs by employing a 1-vs.-rest layer. This layer uses Sigmoid activation functions and Gaussian fitting to classify known classes while rejecting unknown classes. It has N Sigmoid functions for N seen classes; it rejects unseen classes based on thresholding (t_i) as follows:¹⁸

$$y = \begin{cases} reject, & \text{if } Sigmoid(x_i) < t_i, \forall c_i \in Y \\ argmax_{c_i \in Y} Sigmoid(x_i), & otherwise \end{cases} \quad (1)$$

where $Sigmoid(x_i)$ is the function of a class c_i . We reject a given test example (x_i) if all predicted probabilities are less than their corresponding thresholds for that example; otherwise, the class with the highest probability is predicted.

Although DOC has been widely used for open world deep learning classification, other methods have been used. For example, a simpler approach would be using thresholds on the Softmax output; i.e., a given input image is labeled as unknown if none of the classes reaches a predetermined threshold. The performance of this approach is sensitive to the threshold, which has to be estimated empirically from the training dataset. Another method that has been widely used to integrate open world learning into CNNs is OpenMax.²⁰ In this method, the traditional Softmax layer is extended to predict unknown classes using the likelihood of failure and the concept of meta-recognition.²² To estimate if the input is unknown or "far" from known classes, the scores from the penultimate layer of CNNs (i.e., fully connected layer) are used. Then, inputs that are far (in terms of distribution) from known classes are rejected. One limitation of OpenMax is that it requires validation examples from the unseen class for hyperparameters tuning. However, the 1-vs.-rest Sigmoid layer provides a representation of all classes (known and unknown classes). Previous studies²¹ showed that 1-vs.-rest layer achieved superior performance as compared to both Softmax thresholding and OpenMax. In this work, we investigate these three approaches, namely Softmax thresholding, 1-vs.-rest layer, and OpenMax, for open world echocardiography view classification.

2.3 Cluster-Based Active Learning

Although supervised learning methods achieved excellent state-of-the-art performance in several applications including medical image applications, these methods require each sample in the dataset to be annotated during training. In real-world applications, manual labeling of samples is expensive, and might not even be possible due to the dynamic nature of real-world datasets. Therefore, active learning^{23,24} has been proposed to interactively annotate selected data samples that maximizes the model's performance while minimizing labeling. To select the best samples from a pool of unlabeled data, several criteria have been proposed in the literature. These criteria can be divided into uncertainty-based criteria^{25–27} and clustering-based criteria.^{28–30} Clustering-based approaches follow a labeling procedure based on the clustering of the training data. Specifically, these approaches require experts to annotate clusters that contain similar examples instead of annotating single examples, which can significantly boost active learning as well as reduce the labeling cost and the required number of human interactions.

In this work, we combine a clustering method with active learning for echocardiography view classification. In particular, our approach integrates cluster annotation steps into the standard active learning framework, and asks the human expert to label clusters of images instead of single images. To the best of our knowledge, this work is the first to explore cluster-based active learning for echocardiography classification.

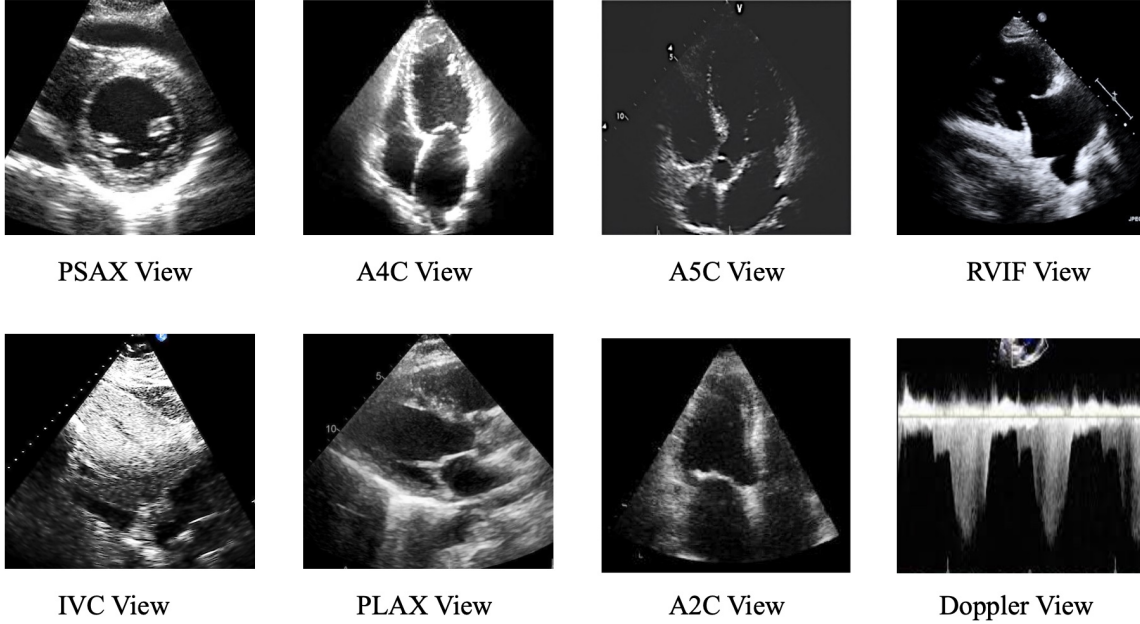


Figure 1: Examples of echocardiography views. PSAX, A4C, A5C, RVIF, IVC, PLAX, and A2C stand for parasternal short axis, apical 4 chamber, apical 5 chamber, right ventricle inflow, inferior vena cava, parasternal long axis, and apical 2 chamber, respectively.

3. MATERIAL & METHOD

3.1 Echocardiography Datasets

We used two datasets to build and evaluate the proposed framework. These datasets are: the publicly available EchoNet-Dynamic dataset³¹ and a private NIH dataset.

3.1.1 EchoNet-Dynamic Dataset

EchoNet-Dynamic dataset³² contains 10,036 B-mode (A4C) videos collected from 10,036 random patients who underwent an echocardiography exam between 2006 and 2018. The number of video frames ranges from 24 to 1002 with a mean acquisition rate of 51 frames per second (FPS). In the processing stage, the videos were cropped, masked to remove protected health information (PHI), and resized to 112×112 pixel resolution. We converted these videos into images by extracting the key frames from all videos in the dataset. We then divided the extracted key frames into training and validation sets. These sets are used to build a self-supervised echo-specific representation. As discussed in the next section, the use of echo-specific representation enhances the performance of the target task (view classification) because it provides better initialization.

3.1.2 NIH Echocardiography View Dataset

We evaluated the proposed open world active learning framework using an echocardiography dataset collected in the Clinical Center at the National Institutes of Health (NIH). The use of the de-identified data was excluded from the Institutional Review Board (IRB) per 45 CFR 46 and the NIH policy for the use of specimens/data (OHSRP#18-NHLBI-00686). Our dataset contains images recorded from different cardiac views including PLAX, A4C, inter vena cava (IVC), Doppler (DP), among others. Our initial dataset, which is used to train the initial classifier, has four known classes: PLAX, A4C, IVC, and Doppler. All other views (e.g., parasternal short-axis views [PSAX] and apical 2 chambers [A2C]) are considered unknown. We used 80% of the dataset for training and validation and the remaining 20% as an independent set for testing. Figure 1 shows examples of common echocardiography views.

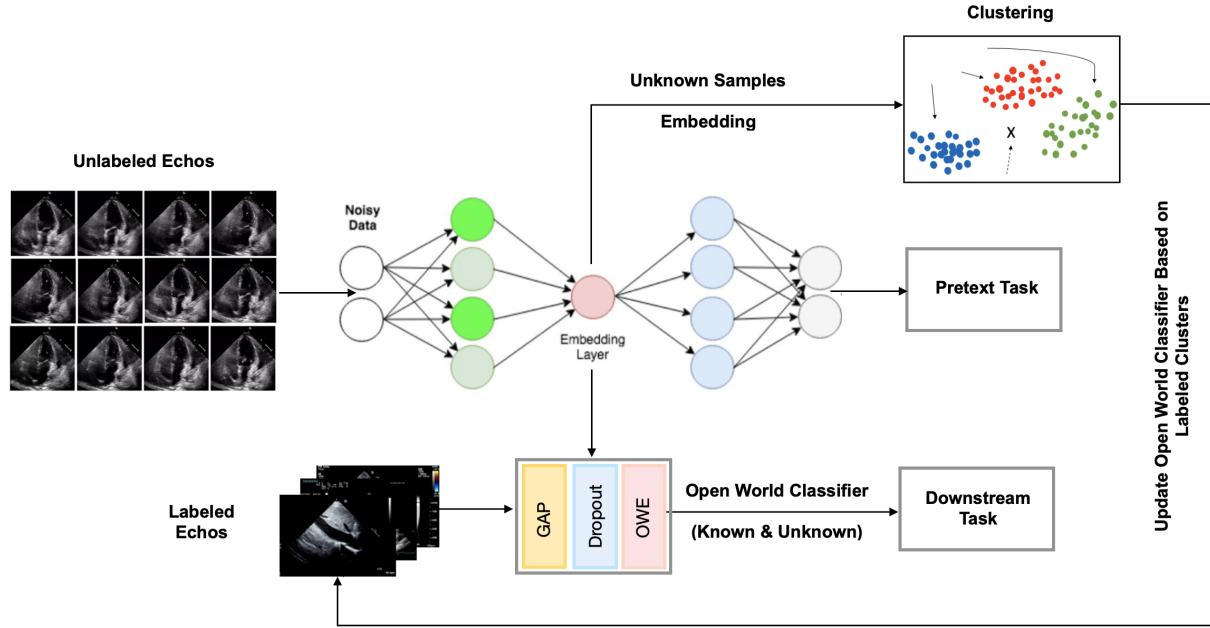


Figure 2: Overview of the open world active learning framework for echocardiography view classification.

3.2 Open World Active Learning for Echocardiography View Classification

Our proposed framework has two main stages: a classification stage and a clustering stage. In the classification stage, an open world view classifier is trained to classify known echocardiography views and detect unknown views. In the clustering stage, similar unknown images are grouped into clusters to be labeled by a human expert before passing them back to the classification stage for model update. Figure 2 presents an overview of the proposed open world active learning framework for echocardiography view classification.

3.2.1 Classification Stage

Our open world classifier is constructed as follows. First, we used a self-supervised denoising autoencoder to learn echo-specific feature representation. Using self-supervised learning to build echo-specific representation allows exploiting publicly available large scale and unannotated datasets for creating better initialization and transferring relevant knowledge (i.e., echo weights) to target tasks that may have relatively small datasets. This can lead to better generalizability and faster convergence as discussed thoroughly in.³³ We used a VGG-based autoencoder, and trained it using a large-scale dataset (EchoNet-Dynamic³²) to learn echo-specific features. We trained the autoencoder to minimize the mean square error (MSE) with a batch size of 64 for 100 epochs, root mean square propagation (RMSprop), and initial learning rate of 1×10^{-2} .

After building the echo-specific autoencoder, we appended the pre-trained echo-specific encoder (the encoder of the autoencoder) with the following classification layers: global average pooling (GAP), empirically determined dropout ratio of 0.5, and fully connected (FC) layer. Then, we added an open world estimator (OWE) to estimate the probability of an image belonging to an unseen view class. For the OWE, we experimented with Softmax thresholding, 1-vs.-rest layer, and OpenMax (see Section 2.2). Finally, we fine-tuned the classifier to minimize the loss using a stochastic gradient descent (SGD) optimizer with an initial learning rate of 1×10^{-3} and momentum of 0.9.

3.2.2 Clustering Stage

We used a clustering algorithm to group similar images of unknown classes into clusters to be labeled by human experts. Several methods can be used to cluster images of unknown classes. Examples of these methods include K-medoids,³⁴ K-Means,³⁴ and K-Centers.³⁴ Although k-means and k-centers have been widely studied in the

literature, these methods do not return real samples as cluster representatives,³⁴ making them unsuitable for active learning.

Hence, we used a simple and fast K-medoids clustering algorithm³⁵ to group unknown images into cluster representatives. Specifically, the embedded features of the unknown images, which are extracted by the autoencoder (see Figure 2), are used by K-medoids to generate k clusters of unknown classes. The optimal number of clusters (k) can be determined empirically (e.g., elbow method) or specified by a human’s expert. After grouping unknown images into different clusters, a certified echocardiologist labeled each cluster group of unknown images instead of labeling all the unknown images, leading to a significant reduction in the required time and number of human interactions. Finally, the newly labeled images (previously unknown) were sent back to the classification stage and used to update the open world classifier.

The iterative procedure of training and updating the proposed framework can be summarized as follows:

1. Train an initial open world classifier to classify echocardiography view images as known views or unknown (unseen) views.
2. Use the features embeddings of unknown images (extracted by the autoencoder) and K-medoid clustering algorithm to group similar unknown images into k clusters.
3. Present the k clusters of unknown images to human experts for labeling.
4. Add the newly labeled images to the initially labeled dataset.
5. Re-train the open-world classifier (step 1) using the new labeled dataset.

We repeat this process of labeling clusters and using them to update the open world classifier whenever new clusters of unknown images are created.

4. EXPERIMENTS & RESULTS

We conducted different experiments to evaluate the proposed open world active learning framework. For the analysis, we partitioned the echocardiography view dataset into four known classes (IVC, Doppler, PLAX, and A4C) and four unknown classes.

4.1 Open World Classification

To evaluate the performance of echocardiography view classification in closed and open world settings, we performed four experiments to train four different classifiers. In each experiment, we trained the classifier using the echo-specific weights, which are learned by the self-supervised representation, and compared its performance with random and ImageNet weights.

4.1.1 Experiment 1 (Closed world classifier)

This classifier is trained with the classical Softmax to recognize four classes: PLAX, IVC, A4C, and Doppler. If the classifier encounters an unknown class, it classifies it as one of the known classes. To evaluate the impact of the echo-specific representation, we trained three versions of this closed world classifier. The first version is initialized using the echo weights learned from the echo-specific representation. The second and third ones are trained using random weights and ImageNet weights, respectively.

4.1.2 Experiment 2 (Open world classifier with Softmax)

This classifier is trained only on known classes and uses a specific threshold to detect unknown classes. Specifically, this classifier labels an input image as unknown if it has an output score less than a given threshold (t). Similar to the closed world classifier, we trained this classifier with echo-specific, random, and ImageNet weights to evaluate the impact of the echo-specific representation on the classification performance. In all experiments, we set the threshold $t = 0.5$.

Table 1: Performance (F-score) of echocardiography view classification using the traditional closed world classifier and three open world classifiers. We initialized each classifier with echo-specific, ImageNet, and random weights to evaluate the impact of our self-supervised echo-specific representation on the performance of view classification. The \star symbol indicates that the performance of using echo-specific weights is significantly ($P < 0.05$) higher than other weights. \times symbol indicates that the performance of open world (1-vs.-rest) is significantly ($P < 0.05$) higher than closed world, open world (Softmax), and open world (OpenMax).

Classifier	Echo-specific Weights	ImageNet Weights	Random Weights
Cloed world (traditional)	0.698±0.11 \star	0.685±0.17	0.636 ±0.22
Open world (Softmax)	0.755±0.08 \star	0.723±0.010	0.702±0.14
Open world (OpenMax)	0.858±0.21 \star	0.824±0.16	0.776±0.22
Open world (1-vs.-rest)	0.954±0.09 $\star\times$	0.911±0.13	0.875±0.11

4.1.3 Experiment 3 (Open world classifier with OpenMax)

This classifier is trained for an open world setting using the OpenMax layer. To generate the OpenMax layer, we computed the activation vectors based on the fully connected layer (penultimate) as follows. For each class in our dataset, we computed the activation vector based on the output of the penultimate network layer for all correctly classified examples. Then, we computed the Euclidean distance between each correctly classified training example and their corresponding activation vector to generate class-specific distance distribution. From these distances, we estimated the parameters of Weibull distribution. After estimating Weibull distributions for each class, the classical Softmax layer is replaced with a new OpenMax layer that outputs a distribution among $C + 1$ classes. During testing, a distance with respect to the activation vectors is computed and used to revise OpenMax activations. These OpenMax activations are used to accept known classes and reject unknown ones.

4.1.4 Experiment 4 (Open world classifier with 1-vs.-rest)

In this classifier, we included a 1-vs.-rest layer after the fully connected layer. This layer rejects an input image or labels it as unknown using a Gaussian fitting (GF)-based threshold (see Section 2.2). This classifier is trained three times with echo-specific weights, random weights, and ImageNet weights.

Table 1 presents the performance of the four experiments. Previous studies³⁶ reported that F measure is better than other metrics (e.g., accuracy) when reporting the performance of open world classification as it is not inflated by true negatives; hence, we used F measure to report the performance. From the table, we can see that the classifiers in the four experiments achieved significantly ($P < 0.05$) higher performance when initialized using echo-specific weights. This is attributed to the similarity of the transferred echo-specific knowledge to the target task (view classification) as compared to the knowledge (or weights) obtained randomly or from stock photographic images (ImageNet). These results are consistent with previous works in the literature^{33,37-39} that report enhanced performance and generalization of target tasks when initialized with modality-specific weights. We can also see from Table 1 that the closed world classifier achieved the lowest performance. This is attributed to the random classification of unknown views as one of the known views. Comparing the open world classifiers, we can observe that the 1-vs.-rest open world classifier achieved the highest performance. The receiver operating characteristic (ROC) curves for all classifiers are shown in Figure 3. As shown in the figure, the curve of the open world 1-vs.-rest classifier lies above all other curves, suggesting its superiority. These results support previous works^{21,40} that reported the superior performance of 1-vs.-rest layer (a.k.a, DOC) for both open image and open text classification.

4.2 Cluster-based Active Learning

After the best performing open-world classifier (i.e., 1-vs.-rest) identifies the unknown views, these views are 1) clustered as described in Section 3.2.2, 2) labeled by a human expert, and 3) used to update the initial open world classifier with the new known views (previously unknown).

Figure 4 shows the confusion matrices of the open world classifier with two updates. In the left confusion matrix, we can see that the updated open world classifier can classify the images of the four known classes

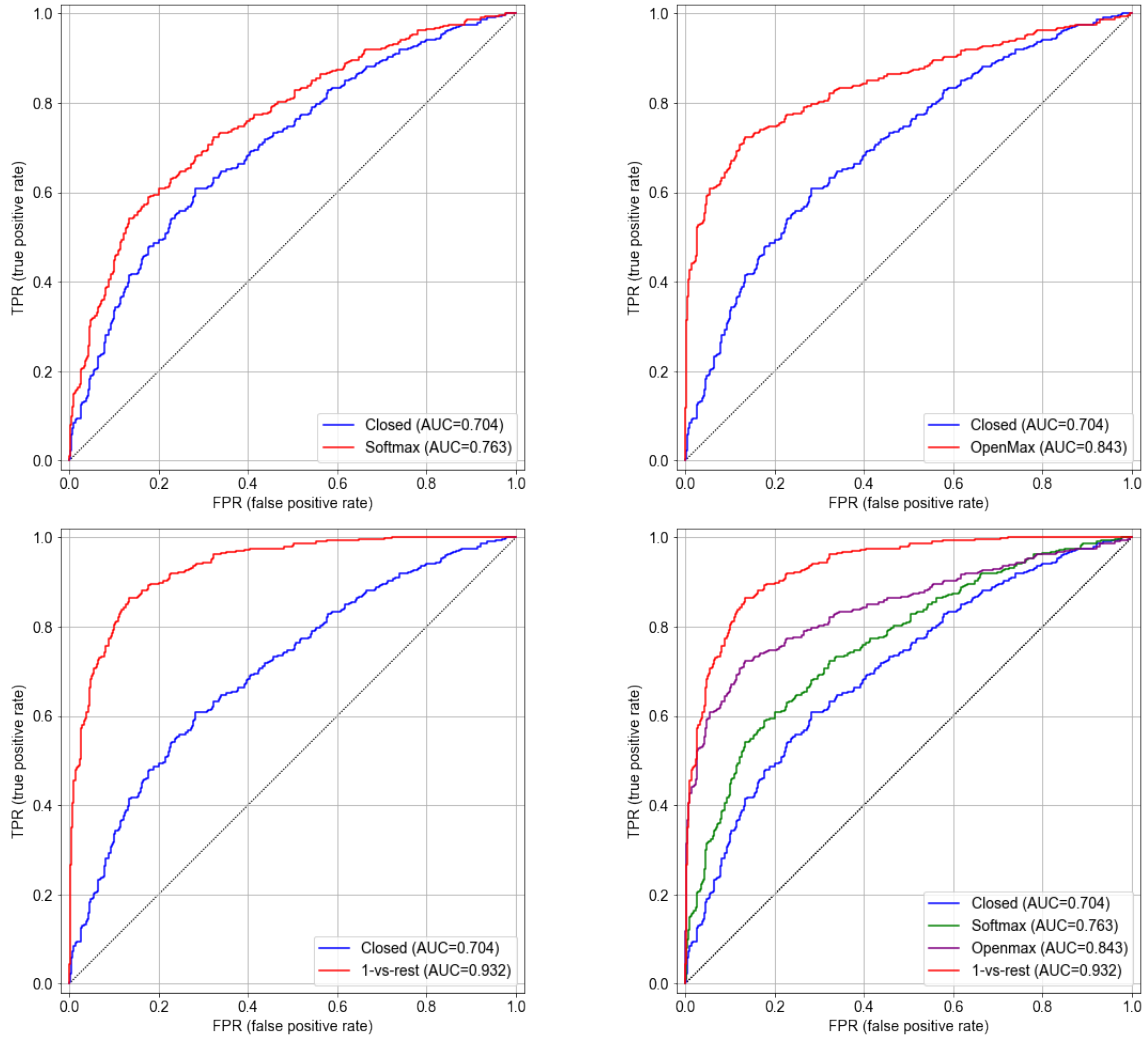


Figure 3: ROC plots for closed world and open world classifiers.

as well as the images from two new classes (A2C and PSAX) with high accuracy. It can also label unseen images as unknown. Then, these new unknown images are clustered, labeled, and used for another update. The performance of the second classifier's update is reported in the right confusion matrix. Similarly, we can see that the updated classifier classified the images of the six classes as well as the images from two new classes (A3C and M mode) with high accuracy. These results prove the ability of the proposed classifier to accurately classify known views while actively recognizing and learning unknown views in a dynamic way. All current echocardiography view classifiers are static and closed (i.e., only classify views seen during training). These results are encouraging and prove the robustness, reliability, and superiority of the proposed framework as compared to the traditional approaches for echocardiography view classification.

5. CONCLUSION

As the real world clinical environment is dynamic and open containing images from classes that might not appear during training, it is important to design a robust open world classifier that classifies medical images from the seen classes while rejecting or labeling images from unseen classes as unknown. This paper presents the first application of an open world active learning framework in echocardiography view classification. Our

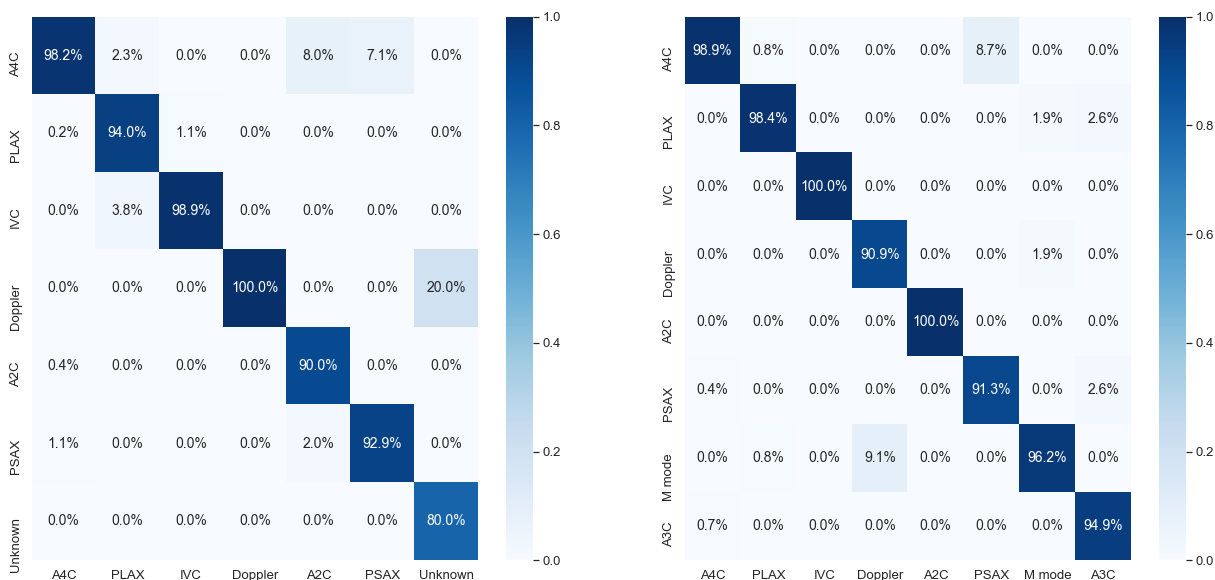


Figure 4: Confusion matrices for open world active learning classifier. First column: the matrix for the open classifier after an update, with the initial 4 known classes and 2 new known classes (A2C and PSAX). Second column: the matrix for the open classifier after a second update with the initial 4 known classes, 2 known classes from the first update (A2C and PSAX), and 2 new known classes (A3C and M-mode). From these matrices, we can see that the classifier is dynamically and actively learning and updating its knowledge about the views while identifying unknown images. Further, we can observe that the performance of the known classes improved after each update of the classifier.

framework uses a self-supervised autoencoder trained on unlabeled echo dataset to learn echo-specific feature representation. The learned representation is then used to enhance the performance and generalizability of the echocardiography view classifier. To make this classifier suitable for the open world setting, we experimented with three approaches: Softmax thresholding, OpenMax, and 1-vs.-rest layer. Further, we used a cluster-based approach to actively cluster images of unknown classes, present these clusters to experts for annotation, and use them to update the open world classifier. As experts annotate clusters instead of single images, this approach significantly reduces the required number of human interactions to train and update the classifier. Our experiments on an echocardiography dataset with known and unknown views show that the proposed classifier significantly outperforms the closed world echocardiography view classifier and achieves new state-of-the-art results.

The proposed framework can be improved in different ways. For example, we can explore different CNNs architectures, hyperparameter optimization methods, and other clustering methods. Also, we can explore presenting a representative subset of images with fewer samples instead of presenting all images in the cluster to reduce the cluster’s visual complexity. We believe this framework can be easily extended and applied to other medical imaging modalities including chest X ray (CXR), Magnetic resonance imaging (MRI), and computerized tomography (CT).

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Library of Medicine (NLM) at the National Institutes of Health (NIH). We would like to acknowledge our partners in the National Heart Lung and Blood Institute (NHLBI) for providing access to the dataset as well as guidance on clinical matters.

REFERENCES

- [1] Wasserman, M. A., Shea, E., Cassidy, C., Fleishman, C., France, R., Parthiban, A., and Landeck, B. F., “Recommendations for the adult cardiac sonographer performing echocardiography to screen for critical congenital heart disease in the newborn: from the american society of echocardiography,” *Journal of the American Society of Echocardiography* **34**(3), 207–222 (2021).
- [2] Qian, Y., Wang, L., Wang, C., and Gao, X., “The synergy of 3d sift and sparse codes for classification of viewpoints from echocardiogram videos,” in [*MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support*], 68–79, Springer (2012).
- [3] Wu, H., Bowers, D. M., Huynh, T. T., and Souvenir, R., “Echocardiogram view classification using low-level features,” in [*2013 IEEE 10th International Symposium on Biomedical Imaging*], 752–755, IEEE (2013).
- [4] Agarwal, D., Shriram, K., and Subramanian, N., “Automatic view classification of echocardiograms using histogram of oriented gradients,” in [*2013 IEEE 10th International Symposium on Biomedical Imaging*], 1368–1371, IEEE (2013).
- [5] Penatti, O. A., Werneck, R. d. O., de Almeida, W. R., Stein, B. V., Pazinato, D. V., Júnior, P. R. M., Torres, R. d. S., and Rocha, A., “Mid-level image representations for real-time heart view plane classification of echocardiograms,” *Computers in biology and medicine* **66**, 66–81 (2015).
- [6] Zhang, J., Gajjala, S., Agrawal, P., Tison, G. H., Hallock, L. A., Beussink-Nelson, L., Fan, E., Aras, M. A., Jordan, C., Fleischmann, K. E., et al., “A computer vision pipeline for automated determination of cardiac structure and function and detection of disease by two-dimensional echocardiography,” *arXiv preprint arXiv:1706.07342* (2017).
- [7] Madani, A., Arnaout, R., Mofrad, M., and Arnaout, R., “Fast and accurate view classification of echocardiograms using deep learning,” *NPJ digital medicine* **1**(1), 6 (2018).
- [8] Simonyan, K. and Zisserman, A., “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556* (2014).
- [9] He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,” in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).
- [10] Østvik, A., Smistad, E., Aase, S. A., Haugen, B. O., and Lovstakken, L., “Real-time standard view classification in transthoracic echocardiography using convolutional neural networks,” *Ultrasound in medicine & biology* **45**(2), 374–384 (2019).
- [11] Smistad, E., Østvik, A., Salte, I. M., Melichova, D., Nguyen, T. M., Haugaa, K., Brunvand, H., Edvardsen, T., Leclerc, S., Bernard, O., et al., “Real-time automatic ejection fraction and foreshortening detection using deep learning,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **67**(12), 2595–2604 (2020).
- [12] Das, A., Mohapatra, S. K., and Mohanty, M. N., “Design of deep ensemble classifier with fuzzy decision method for biomedical image classification,” *Applied Soft Computing* **115**, 108178 (2022).
- [13] Mahapatra, D., Bozorgtabar, B., and Ge, Z., “Medical image classification using generalized zero shot learning,” in [*Proceedings of the IEEE/CVF International Conference on Computer Vision*], 3344–3353 (2021).
- [14] Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., and Martel, A. L., “Loss odyssey in medical image segmentation,” *Medical Image Analysis*, 102035 (2021).
- [15] Yan, X., Tang, H., Sun, S., Ma, H., Kong, D., and Xie, X., “After-unet: Axial fusion transformer unet for medical image segmentation,” in [*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*], 3971–3981 (2022).
- [16] Hui, L. and Belkin, M., “Evaluation of neural architectures trained with square loss vs cross-entropy in classification tasks,” *arXiv preprint arXiv:2006.07322* (2020).
- [17] Tyagi, A. K. and Rekha, G., “Challenges of applying deep learning in real-world applications,” in [*Challenges and Applications for Implementing Machine Learning in Computer Vision*], 92–118, IGI Global (2020).
- [18] Shu, L., Xu, H., and Liu, B., “Doc: Deep open classification of text documents,” *arXiv preprint arXiv:1709.08716* (2017).
- [19] Prakhya, S., Venkataram, V., and Kalita, J., “Open set text classification using convolutional neural networks,” in [*International Conference on Natural Language Processing, 2017*], (2017).

- [20] Bendale, A. and Boulton, T. E., “Towards open set deep networks,” in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1563–1572 (2016).
- [21] Chen, Z. and Liu, B., “Lifelong machine learning,” *Synthesis Lectures on Artificial Intelligence and Machine Learning* **12**(3), 1–207 (2018).
- [22] Scheirer, W. J., Rocha, A., Micheals, R. J., and Boulton, T. E., “Meta-recognition: The theory and practice of recognition score analysis,” *IEEE transactions on pattern analysis and machine intelligence* **33**(8), 1689–1695 (2011).
- [23] Settles, B., “Active learning literature survey,” (2009).
- [24] Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., Chen, X., and Wang, X., “A survey of deep active learning,” *ACM Computing Surveys (CSUR)* **54**(9), 1–40 (2021).
- [25] Burbidge, R., Rowland, J. J., and King, R. D., “Active learning for regression based on query by committee,” in [*International conference on intelligent data engineering and automated learning*], 209–218, Springer (2007).
- [26] Wang, R., Chen, D., and Kwong, S., “Fuzzy-rough-set-based active learning,” *IEEE Transactions on Fuzzy Systems* **22**(6), 1699–1704 (2013).
- [27] Liu, J., Li, X., Zhou, J., and Shen, J., “Prediction stability as a criterion in active learning,” in [*International Conference on Artificial Neural Networks*], 157–167, Springer (2020).
- [28] Wang, M., Min, F., Zhang, Z.-H., and Wu, Y.-X., “Active learning through density clustering,” *Expert systems with applications* **85**, 305–317 (2017).
- [29] Liu, J., Wang, Y., Hooi, B., Yang, R., and Xiao, X., “Active learning for node classification: The additional learning ability from unlabelled nodes,” *arXiv preprint arXiv:2012.07065* (2020).
- [30] Urner, R., Wulff, S., and Ben-David, S., “Plal: Cluster-based active learning,” in [*Conference on Learning Theory*], 376–397, PMLR (2013).
- [31] Ouyang, D., He, B., Ghorbani, A., Lungren, M. P., Ashley, E. A., Liang, D. H., and Zou, J. Y., “Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning,” in [*NeurIPS ML4H Workshop: Vancouver, BC, Canada*], (2019).
- [32] Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C. P., Heidenreich, P. A., Harrington, R. A., Liang, D. H., Ashley, E. A., et al., “Video-based ai for beat-to-beat assessment of cardiac function,” *Nature* **580**(7802), 252–256 (2020).
- [33] Zamzmi, G., Rajaraman, S., and Antani, S., “Ums-rep: Unified modality-specific representation for efficient medical image analysis,” *Informatics in Medicine Unlocked* , 100571 (2021).
- [34] Madhulatha, T. S., “Comparison between k-means and k-medoids clustering algorithms,” in [*International Conference on Advances in Computing and Information Technology*], 472–481, Springer (2011).
- [35] Park, H.-S. and Jun, C.-H., “A simple and fast algorithm for k-medoids clustering,” *Expert systems with applications* **36**(2), 3336–3341 (2009).
- [36] Scheirer, W. J., de Rezende Rocha, A., Sapkota, A., and Boulton, T. E., “Toward open set recognition,” *IEEE transactions on pattern analysis and machine intelligence* **35**(7), 1757–1772 (2012).
- [37] Rajaraman, S., Zamzmi, G., and Antani, S. K., “Novel loss functions for ensemble-based medical image classification,” *Plos one* **16**(12), e0261307 (2021).
- [38] Zamzmi, G., Rajaraman, S., and Antani, S., “Accelerating super-resolution and visual task analysis in medical images,” *Applied Sciences* **10**(12), 4282 (2020).
- [39] Rajaraman, S., Zamzmi, G., Folio, L., Alderson, P., and Antani, S., “Chest x-ray bone suppression for improving classification of tuberculosis-consistent findings,” *Diagnostics* **11**(5), 840 (2021).
- [40] Shu, L., Xu, H., and Liu, B., “Unseen class discovery in open-world classification,” *arXiv preprint arXiv:1801.05609* (2018).