ME-MAML!: Multi-Label, Expert-Aided Meta-Learning for Chest X-ray Diagnosis

Minjune Hwang*
Department of Computer Science
Stanford University
mjhwang@stanford.edu

Samar Khanna*
Department of Computer Science
Stanford University
samar.khanna@stanford.edu

Tony Sun*
Department of Computer Science
Stanford University
suntony@stanford.edu

Extended Abstract

Transfer learning and self-supervised learning are widely used in a variety of deep learning tasks to help improve performance in new domains. However, transfer learning methods face difficulty when the source dataset is substantially different from the target dataset. For example, in medical imaging, data labels can be sparse and inconsistent. Meta-learning can be an effective technique to adapt to new tasks given minimal data. In this project, we examine how a meta-learning strategy can effectively adapt across different medical image classification tasks given a few examples (shots) of multi-labeled data. We train and evaluate a Model-Agnostic Meta-Learning (MAML) [2] model for a popular chest X-ray benchmark, CheXpert, in which some of its labels are uncertain or partially unknown.

Medical datasets often can have an imbalance between the number of labels of different diseases, some of which have not been sufficiently studied or previously identified. By recycling expert labels for some traditional, well-known diseases and/or taxonomies, we want to efficiently relabel data for the novel diseases. We therefore propose Multi-Label, Expert-Aided Model-Agnostic Meta-Learning (ME-MAML!), an adaptation of MAML [2] for the expert-aided situation in multi-label classification in which there is an additional learning stage for known labels in the query set of its inner loop. We evaluate combinations of labeling schemes, sampling approaches, and models for chest X-ray image classification. We find that ME-MAML! is able to achieve 75.7% accuracy on a test set of four novel diseases with four support images and that across the board, ME-MAML! outperforms MAML in both post-adapt support and query accuracies.

*Equal Contribution

CS330: Deep Multi-Task and Meta Learning, Fall 2021, Stanford University, CA.

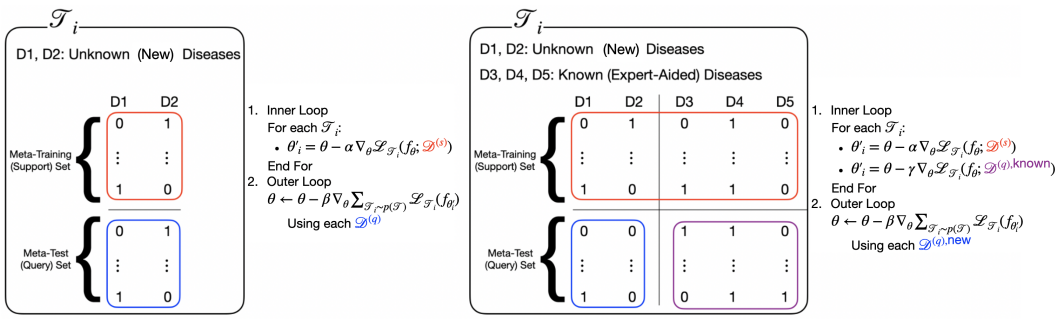


Figure 1: Model-Agnostic Meta-Learning (MAML) vs Multi-Label, Expert-Aided MAML (ME-MAML!). In ME-MAML, we take advantage of expert labels that exist for well-known classes. We add an additional gradient step in the inner loop using known labels for the query set.

1 Introduction

Medical imaging diagnosis is a vital aspect of patient care and disease treatment. It is estimated that 214.2 million X-rays were taken in the US alone in 2018. In recent years, the availability of diverse medical imaging datasets and advances in computational power and deep learning techniques have enabled researchers to investigate applications of computer vision in healthcare, including computer-aided diagnosis, medical image segmentation, and clinical risk prediction.

Most prior work in applying machine learning techniques to healthcare have focused on improving model performance on traditional disease classification tasks [5, 7]. These techniques usually take the form of training a model on a large, fully annotated dataset. There are a few challenges with this paradigm: (1) manually labelling such datasets can be expensive in human labour and time [3, 5], (2) the models trained may be unable to generalise to new data given limited samples.

While transfer learning and self-supervised learning have been used to adapt models to new domains more efficiently than training from scratch, these methods can under-perform if the source dataset is substantially different from the target dataset. In the field of healthcare, this may be especially relevant as new diseases are discovered to which prior models may fail to adapt.

In this project, we examine how a meta-learning strategy can effectively adapt across different medical image classification tasks given a few examples of multi-labeled data. By modeling our data as a distribution of distinct but related "tasks", meta-learning offers tools to adapt to new tasks using limited numbers of instances (or "shots").

While meta-learning has seen success in numerous applications from computer vision to robotics [2], an under-explored avenue is in healthcare, especially in the multi-label setting where labels can be sparse and inconsistent. In traditional single-label classification, meta-learning can be used to quickly adapt to new tasks, where each task can consist of N classes with K instances (or shots) per class. However, this is not directly applicable to the multi-label setting given that instances can belong to multiple classes at once. Moreover, the multi-label setting affords modeling tasks as sets of "labels" such that new tasks may annotate either *existing* or new data with new labels.

This is especially relevant in disease classification, where a patient may be suffering multiple known and as-yet unknown diseases at once. If a new disease such as COVID-19 is discovered by scientists, we consider if we can effectively use a few examples of patients with COVID-19 to retroactively identify if previously seen patients were suffering from this same ailment. Moreover, diseases and health conditions may occur together (for example, coughing and pneumonia with COVID-19). Therefore, can we effectively use information about other *known* health conditions/diseases with which the patient is diagnosed to classify if they are suffering from a *new* disease.

Our contributions in this work are as follows: (1) We propose a framework for multi-label classification under which the ability of models to generalise to new classes (i.e. diseases) can be tested, and (2) we create ME-MAML!, an adaption MAML [2], a popular approach to gradient-based meta-learning, to this framework by using information derived from *known* diseases to improve the classification

of new target diseases given a just a few samples of labeled data. We evaluate both MAML and ME-MAML! on the CheXpert dataset [3] which is a multi-label classification dataset for chest X-ray diagnosis. By adapting the inner-loop of MAML to include a gradient update on previously *known* diseases, we improve validation set performance by 9.3% in absolute terms, as measured by the difference in accuracy for classifying *unseen* diseases. [*]

2 Related Work

Prior work [8] has demonstrated that meta-learning can be used to better predict patient risk to previously unseen target diseases given limited samples of labeled data, even out-performing supervised methods trained on the individual diseases.

One popular approach has been to use transfer learning by fine-tuning a pre-trained model from ImageNet on a given dataset [5, 7]. Zhang et. al. [9] demonstrate that self-supervised pre-training using pairs of medical images and noisy text descriptors improves chest x-ray classification. Moreover, Sowrirajan et. al. [6] demonstrate that pre-training using momentum contrast learning improves pathology detection in chest x-rays, even across different datasets. Mahajan et. al. introduce MetaDerm and demonstrate the effectiveness of gradient-based meta-learning approaches on few-shot disease classification over transfer learning [4]. However, these methods don't consider the multi-label setting and how it may be possible to utilise known disease information to improve detection of novel diseases.

Azizi et. al. [1] demonstrated that self-supervised pre-training can be used to improve medical image classification accuracy on dermatology skin condition and chest x-ray classification. However, pre-training from ImageNet may not be ideal if the domain shift between the characteristics of ImageNet images and medical forms of data are too large [9].

There is a dearth of research into meta-learning for medical data, especially in the multi-label setting. Zhang et. al. consider predicting clinical risk for different diseases given electronic health records (EHR). [8]. However, this method is used only in the single-label setting and uses supervision from source domain tasks during each gradient step for a new target domain task. This is different from our proposed framework given our multi-label setting and since modeling the problem in fixed source or target domains doesn't allow one to flexibly consider learning *unknown* new diseases. Cases where new datasets or instances of a different disease are labeled are under explored.

3 Methodology

3.1 Meta-Learning Setup

In meta-learning, models are adapted to a set of tasks, where each task is sampled from a distribution $\mathcal{T}_i \sim p(\mathcal{T})$. Our goal is to make a model learn how to classify new diseases given a few labeled examples. Given the multi-label framework, we therefore characterise a task τ by a subset Y_{new} of all possible diseases $\mathcal{Y}$. We further distinguish between *known* and *novel* diseases within Y, such that $Y = Y^{\text{known}} \cup Y^{\text{new}}$. This distinction enables us to leverage information from Y^{known} to potentially aid performance on Y^{new}. Note that the goal for each task is still to correctly classify the new diseases in Y^{new}. The known diseases are only provided to be optionally used to improve performance on Y^{new}.

For example, $\mathcal{T}_i$ may be defined with $Y_i = \{\text{Pneumonia}, \text{Edema}, \text{Consolidation}\}$ where "Pneumonia" and "Edema" may be the new diseases we want to classify, and "Consolidation" could be a known disease available as an optional crutch. Another task $\mathcal{T}_j$ could be defined by $Y_j = \{\text{Pleural Effusion}, \text{Atelectasis}, \text{Cardiomegaly}\}$, where "Pleural Effusion" and "Atelactasis" are the target new diseases and "Cardiomegaly" is a known disease. Each of sets Y_i and Y_j is a subset of a set of all possible diseases $\mathcal{Y}$.

Each task τ_i will consist of a **support** set of instance-label pairs $\mathcal{D}_i^{(s)} = \{(\vec{x}_1, \vec{y}_1), \ldots, (\vec{x}_K, \vec{y}_K)\}$, and a corresponding **query** set $\mathcal{D}_i^{(q)} = \{(\vec{x}_1, \vec{y}_1), \ldots, (\vec{x}_Q, \vec{y}_Q)\}$, where each $\vec{y}_j \in \{0, 1\}^{|Y_i|}$ is a binary vector of length $|Y_i|$ containing only the labels in the subset Y_i for task $\mathcal{T}_i$. Therefore, all

[*]Our code is available online at https://github.com/Dieblitzen/cs330-final.

considered, a task is defined as:

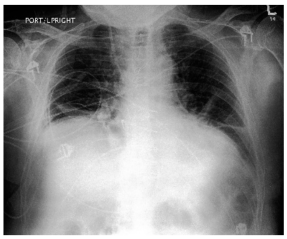
$$\mathcal{T}_i = \{\mathcal{D}_i^{(s)}, \mathcal{D}_i^{(q)}, Y_i = Y_i^{\text{known}} \cup Y_i^{\text{new}}\}\}$$

Our goal is to adapt a model f with meta-parameters θ such that for a new task $\mathcal{T}_j = \{D_j^{(s)}, D_j^{(q)}, Y_j\}$, θ can be quickly adapted to ϕ_j in just 1 or a few gradient steps. We expect that ϕ_j can successfully classify the novel diseases Y_j^{new} of our task.

Before meta-training, we first sample a subset of diseases $Y_{\text{test}} \subset \mathcal{Y}$ and hold this subset out for meta-test time. The classes in Y_{test} will not be seen during meta-training. During meta-training, we sample a task $\mathcal{T}_i \sim p(\mathcal{T})$, where each $Y_i = Y_i^{\text{known}} \cup Y_i^{\text{new}}$ such that $|Y_i^{\text{new}}| = |Y_{\text{test}}|$ (i.e. the number of new diseases per task is the same as the number of new diseases we intend to classify at meta-test time). Our model then adapts the meta-parameters θ on the support set $\mathcal{D}_i^{(s)}$ to get task adapted parameters ϕ_i. We then use f_{ϕ_i} to compute the loss on instances in the query set $\mathcal{D}_i^{(q)}$, and backpropagate to the meta-parameters θ. Once the loss is backpropagated over a batch of tasks, we take a gradient step to update the meta parameters θ.

3.2 Dataset and Sampling

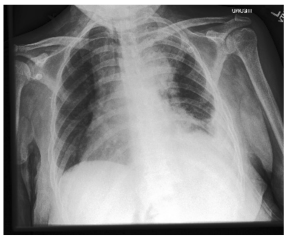
Chest radiographs are the most commonly performed modality for medical image analysis tasks, thanks to their abundance as datasets. One of the most popular Chest X-ray dataset is CheXpert, a large medical imaging dataset that contains 224,316 chest radiographs of 65,240 patients [3]. There are 14 labels for each image in total. We discard two: “No Finding" and “Support Devices", since these aren’t diseases. Therefore, we have $|\mathcal{Y}| = 12$ total diseases.



Positive (1): Lung Opacity, Atelectasis, Pleural Effusion

Negative (0): Pneumothorax

Uncertain (u): Consolidation



Positive (1): Lung Opacity, Lung Lesion, Consolidation, Pleural Other

Negative (0): Pneumothorax, Pleural Effusion

Uncertain (u): Pneumonia

Figure 2: The sample radiographs of patients with their labels. Diseases not mentioned are missing as labels for these instances.

A unique property of the CheXpert dataset is that it captures uncertainties and missing labels inherent in radiograph interpretation, and the imbalance across different diseases. Different methods have been proposed to handle this uncertainty/imbalance and train a model to correctly output probabilities of these observations given the available frontal and lateral X-rays.

Conventionally, a N-way, K-shot classification task refers to a classification task where K input/output pairs for each of N classes are provided for learning the task.

However, in the multi-label classification, each datapoint is associated with multiple labels. In the case of medical imaging diagnosis, a patient can be diagnosed with 2^N different combinations of N diseases (each disease is either present or absent). Following the conventional terminology, we need $2^N K$ datapoints in total, which is rather unrealistic given the scarcity and imbalance of data in CheXpert Dataset and uncertainties in data labels. As such, we use two methods to remedy these issues.

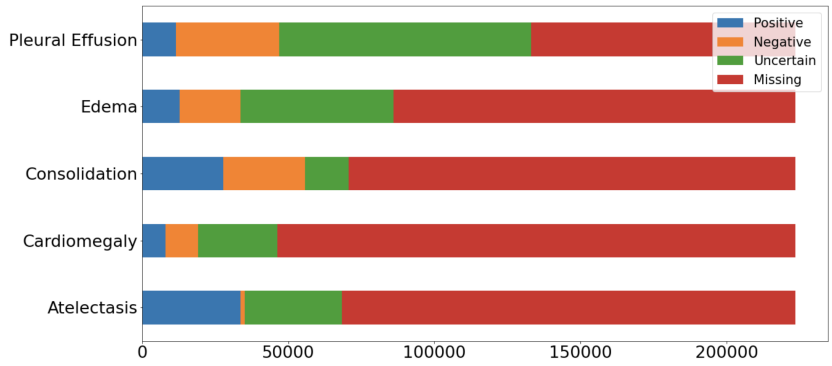


Figure 3: A figure with numbers of datapoints for each disease & label that reflect the imbalance across diseases.

1. **Class-Agnostic Sampling**: For each task τ_i, we first filter for all data points that have non-empty values in at least k of the classes of Y_i^{new}. (For ME-MAML! we also do the same for classes in Y_i^{known}). We found $k = 2$ to be a suitable choice. From the data points available, we then sample $K + Q$ labeled pairs, and then randomly divide them into a support set $\mathcal{D}_i^{(s)}$ of size K and a query set $\mathcal{D}_i^{(q)}$ of size Q.
2. **U-Ones**: The training labels in the dataset for each observation are either 0 (negative), 1 (positive), or u (uncertain). We map all instances of the uncertain label to 1.

As further pre-processing, we also resize every image in the dataset to a pixel resolution $(96, 96)$ to resolve memory issues and speed up training.

3.3 Meta-Learning Algorithms

As our baseline model for meta-learning, we will use Model-Agnostic Meta-Learning (MAML). In MAML, we train the model parameters such that few iterations of gradient descent with few training data from a new task will lead to good generalization performance on that task.

Algorithm 1 MAML

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
randomly initialize θ
while not done **do**
  Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
  for all $\mathcal{T}_i$ **do**
    Sample $\mathcal{D}^{(s)} = \{\mathbf{x}^{(s)}, \mathbf{y}^{(s)}\}$ and $\mathcal{D}^{(q)} = \{\mathbf{x}^{(q)}, \mathbf{y}^{(q)}\}$ from $\mathcal{T}_i$
    Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ using $\mathcal{D}^{(s)}$
    Compute adapted parameters with gradient descent: $\theta'_i = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
  end for
  Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$ using each $\mathcal{D}^{(q)}$
end while

Note that in regular MAML, we are not concerned with Y_i^{known} for a task τ_i. Since the goal is to classify the new targets of the task Y_i^{new} given a few samples, MAML is only concerned with Y_i^{new} in the support set $D_i^{(s)}$ and thus aims to translate that knowledge to effectively classify the diseases Y_i^{new} in the query set $D_i^{(q)}$. The meta-parameters are thus updated in the hope that a model will be able to quickly adapt to a set of new diseases given a new task.

An issue with regular MAML within the multi-label setting is that information about known diseases for examples in the support and query sets remains unused. It is reasonable to expect that knowledge about whether a patient has Pneumonia, for example, may assist in the prediction for whether the patient may have Covid-19. However, if Covid-19 is the new disease we aim to classify, then regular

Algorithm 2 ME-MAML!

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β, γ: step size hyperparameters
1: randomly initialize θ
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Sample $\mathcal{D}^{(s)} = \{\mathbf{x}^{(s)}, \mathbf{y}^{(s)}\}$ and $\mathcal{D}^{(q)} = \{\mathbf{x}^{(q)}, \mathbf{y}^{(q)}\}$ from $\mathcal{T}_i$
6: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta; \mathcal{D}^{(s)})$ and $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta; \mathcal{D}^{(q),\text{known}})$
7: Compute adapted parameters with gradient descent:
$\theta'_i = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta; \mathcal{D}^{(s)}) - \gamma \nabla_{\theta'} \mathcal{L}_{\mathcal{T}_i}(f_\theta; \mathcal{D}^{(q),\text{known}})$
8: **end for**
9: Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$ using each $\mathcal{D}^{(q),\text{new}}$
10: **end while**

MAML doesn't use information about whether the patient has other known diseases (like Pneumonia) to assist in its classification for Covid-19.

We thus aim to address this issue in ME-MAML!. We modify the inner loop in MAML in two ways:

1. **Support set update**: We now update the model parameters θ on the loss computed on both $Y_i^{\text{known}} \cup Y_i^{\text{new}}$ for the support set $D_i^{(s)}$. The intuition is that this allows the model to learn how the classes in Y_i^{new} correlate with those in Y_i^{known}.

2. **Peek the query set**: We allow the model to compute the loss on only the *known* classes of the data in the query set $D_i^{(s)}$. Note that since our tasks are defined by the novelty of the classes Y_i^{new} and not the novelty of the specific data, it is viable to update model parameters on the query set in the task-specific inner loop as long as the *novel* classes are untouched. The intuition is that the model is able to gain information about existing diseases in the query set, which "prepares" it to learn the new diseases in the outer loop update given that it has learned the interactions between Y_i^{known} and Y_i^{new} from the support set.

The combined effect of the modified inner loop can be seen in line 7 of Algorithm 2. The difference between MAML and ME-MAML! can also be seen in 1.

4 Results

We compare a baseline MAML with multiple settings of ME-MAML! on the CheXpert dataset and report the pre-adapt support, post-adapt support, and post-adapt query accuracies. For our model, we take a standard Convolutional Neural Network (CNN) model in which each convolution layer with 64 hidden channels and the kernel size of 3 is followed by batch normalization and ReLu layer. We train the model with Binary Cross Entropy loss, given that we have multi-labels.

After trying out different hyperparameters, we find that using an outer learning rate of 0.1, an inner learning rate of 1, and 8,000 training tasks works best. For each experiment, we take the best post-adapt query accuracy on the validation set and report the corresponding pre-adapt support and post-adapt support accuracies at the specified training step.

We use the following diseases for our test set: Enlarged Cardiomediastinum, Cardiomegaly, Lung Opacity, Lung Lesion. We experiment with varying the support size (Table 1 and number of test classes 2 for both MAML and ME-MAML. In each setting, ME-MAML! was given an additional 6 "known" diseases per task.

We measure performance through accuracy only on the *new* diseases for each, which simply the number of correct predictions divided by the total number of predictions.

Support Size	Pre-Adapt Support		Post-Adapt Support		Post-Adapt Query	
	MAML	ME-MAML	MAML	ME-MAML	MAML	ME-MAML
$n_s = 1$	0.545	0.231	0.684	**0.790**	0.634	**0.725**
$n_s = 2$	0.455	0.253	0.650	**0.799**	0.654	**0.746**
$n_s = 3$	0.548	0.378	0.680	**0.852**	0.658	**0.736**
$n_s = 4$	0.561	0.265	0.702	**0.767**	0.664	**0.757**

Table 1: Results for MAML vs ME-MAML! across varying support sizes.

Test Classes	Pre-Adapt Support		Post-Adapt Support		Post-Adapt Query	
	MAML	ME-MAML	MAML	ME-MAML	MAML	ME-MAML
targets $= 0, 1$	0.433	0.561	0.675	**0.809**	0.631	**0.678**
targets $= 0, 1, 2$	0.568	0.389	0.712	**0.777**	0.671	**0.680**
targets $= 0, 1, 2, 3$	0.561	0.265	0.702	**0.767**	0.664	**0.757**

Table 2: Results for MAML vs ME-MAML! across different test classes with four support images. Targets refer to the index of the disease in the test set (e.g. 0 refers to Enlarged Cardiomediastinum).

We find that across the board, ME-MAML! outperforms the baseline MAML for both post-adapt support and post-adapt query accuracies, while MAML scores higher for pre-adapt support accuracy. This suggests that providing the expert “known" classes during training benefits the model in classifying target “new" diseases.

From Table 1, we see that increasing the number of support shots benefits both MAML and ME-MAML!. This is expected, since more data in the support set allows both versions of the algorithm to better adapt to classifying the new target diseases.

From Table 2, we see that there is a larger drop in performance to ME-MAML! from reducing the number of novel targets, 7.9% for ME-MAML! versus 3.3% for MAML (from targets = { 0, 1, 2, 3} to targets = { 0, 1}). This suggests that there could have been more pertinent information to ME-MAML! in learning about disease correlation with a larger number of target classes.

We also perform a short ablation to study the effect of switching the order of the gradient update in the inner loop of ME-MAML!, and of keeping in the inner loop only (1) the update on the known and unknown classes on the support set (2) the update on the known classes of the query set. These can be found in the appendix 3.

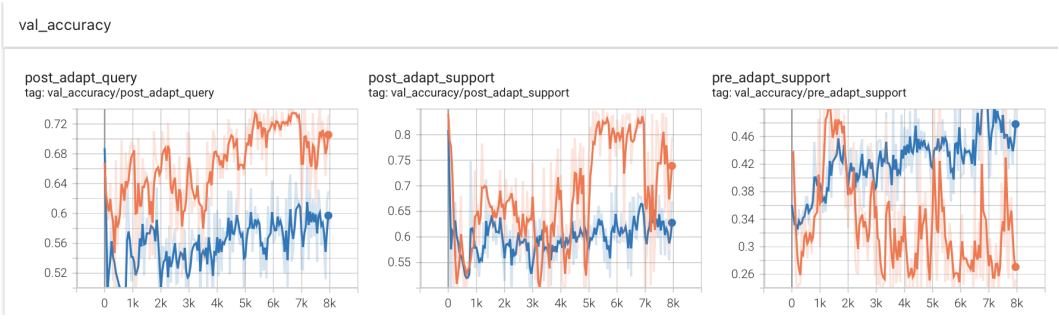


Figure 4: Validation curves for ME-MAML! (orange) and MAML (blue) with targets 0, 1, 2, 3 and $K = Q = 4$. We see ME-MAML! consistenly out-perform MAML for most of the training steps.

5 Conclusion

In this paper, we take a novel approach to few-shot learning with chest X-ray imaging diagnosis in multi-label setting. We experiment with ways to combine gradient updates from two parts of the inner

loop. In particular, we create ME-MAML, a variation of MAML for the expert-aided situation. In the vanilla implementation of MAML, we have a bi-level optimization problem for parameter learning. During inner loop, we optimize parameters ϕ_i using the support set of a mini batch of tasks, and we update θ (parameter vector being meta-learned) using the query set in outer loop. Now, in the case of ME-MAML, we have an additional optimization step within the inner loop for the expert-aided data; we not only train on the support set's known (expert-aided) classes and new classes, but also train on the query set's known (expert-aided) classes in the inner loop. We observed that our meta parameters to more quickly adapt to a given task by leveraging additional expert-aided data. Especially, We found that ME-MAML! outperforms baseline MAML when using higher learning rate and smaller support.

6 Future Work

This worked focused on chest x-ray diagnosis with our proposed multi-label classification algorithm that acquires additional inference about target classes by recycling expert labels for well-known classes in a few-shot learning setting. Future works may apply the algorithm into other few-shot multi-label problems, such as land cover classification or multi-label text classification task.

Another extension of this method can be a case when expert data is not available. We want to test if a pretrained ME-MAML! model can perform well when expert data is not available in a self-supervised manner. The model will first classify well-known diseases to fill up their missing labels and then leverage self-supervised labels to output labels for novel diseases, and we expect this result to be better than the vanilla multi-label MAML case.

7 Team Contributions

Our team plans to meet regularly and work collaboratively on each aspect of the project, ranging from problem scoping and dataset collection to cleaning, model training, evaluation, and presentation. We have confidence that we will be able to lift each other up and bring the best out in one another and that no single person will bear the burden of carrying the whole project.

References

[1] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, Vivek Natarajan, and Mohammad Norouzi. Big self-supervised models advance medical image classification, 2021.

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *CoRR*, abs/1703.03400, 2017.

[3] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 590–597, 2019.

[4] Kushagra Mahajan, Monika Sharma, and Lovekesh Vig. Meta-dermdiagnosis: Few-shot skin disease identification using meta-learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3142–3151, 2020.

[5] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[6] Hari Sowrirajan, Jingbo Yang, Andrew Y. Ng, and Pranav Rajpurkar. Moco pretraining improves representation and transferability of chest x-ray models. In Mattias Heinrich, Qi Dou, Marleen de Bruijne, Jan Lellmann, Alexander Schläfer, and Floris Ernst, editors, *Proceedings of the Fourth Conference on Medical Imaging with Deep Learning*, volume 143 of *Proceedings of Machine Learning Research*, pages 728–744. PMLR, 07–09 Jul 2021.

[7] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.

[8] Xi Sheryl Zhang, Fengyi Tang, Hiroko H. Dodge, Jiayu Zhou, and Fei Wang. Metapred: Meta-learning for clinical risk prediction with limited patient electronic health records. *CoRR*, abs/1905.03218, 2019.

[9] Yuhao Zhang, Hang Jiang, Yasuhide Miura, Christopher D. Manning, and Curtis P. Langlotz. Contrastive learning of medical visual representations from paired images and text. *CoRR*, abs/2010.00747, 2020.

A Additional Experiments

Model	**Pre-Adapt Support**	**Post-Adapt Support**	**Post-Adapt Query**
ME-MAML!	0.289	0.429	0.499
ME-MAML!, support only	0.381	0.436	0.456
ME-MAML!, query only	0.190	0.246	0.404
ME-MAML!, switched order	0.279	0.343	0.472

Table 3: We compare the following ablations on ME-MAML: (1) using only the gradient update on the known diseases of the query set; (2) using only the gradient update on the known and unknown diseases of the support set; (3) switching the order of the gradient updates in the inner loop. These results were on different input sizes for the image and with slightly different hyper-parameters, and so aren't directly comparable with the main table of results. However, we do se a benefit to the order of udpates as seen in 1.