NYU SHANGHAI 上海纽约大学

Data Science

Capstone Report - Fall 2022

Automatic Diagnosis of Myocarditis in Cardiac Magnetic Resonance Images Using Deep Learning Models

Yimeng Zhang,
Ruiqi Xue

supervised by
Professor Li Guo

Preface

Medical imaging problems are never in short supply in the field of machine learning. They are particularly meaningful because the solutions obtained could make a contribution to improving human health and saving lives. Machine learning techniques can be very helpful for this type of problems. In this work, we choose myocarditis and Cardiac magnetic resonance (CMR) images as the target disease and images type. This idea is inspired by the passion and dedication of the authors to apply skills learnt from Data Science major to healthcare field, as well as by the personal experience of one author, which is especially meaningful to her.

The target audience is designed to be our fellow students, professors and scholars from the field of data science, computer science and/or medicine and health science, or general professionals who have a background in these mentioned and relevant areas. In this work, we explore deep learning models that have not been applied in the diagnosis of myocarditis with CMR images before according to existing publications, and the results may give implications for methods chosen in future practice.

Acknowledgements

We want to thank our supervisor Professor Li Guo who has been very patient and supportive. During our weekly meetings, we felt her caring for students as a Professor and affability as a friend. Doing this project with professor Guo is a truly memorable experience for both of us.

Abstract

Cardiac magnetic resonance (CMR) imaging is a popular technique for the diagnosis of myocarditis. However, there are some challenges for specialist physicians to identify myocarditis with CMR due to factors such as low contrast, various noises, and invisible mass or lesion in the images. On one hand, deep learning models requires a large amount of data to achieve good results, but expert manually-annotated medical images are expensive to acquire. On the other hand, deep learning pre-trained models suffer from domain discrepancy and overfitting when trained on a small-sized dataset. In this work, we aim to mitigate data deficiency and improve the classification accuracy of CMR images by developing sample-efficient methods to train highly-performant deep learning models. We synergistically integrated self-supervised learning with transfer learning to learn powerful and unbiased feature representations to improve the classification accuracy. We also experimented with a label smoothing loss function for reducing the risk of overfitting and demonstrated the effectiveness of our proposed methods even when the number of training CMR images is limited.

Keywords

Deep learning; Medical imaging; Contrastive learning; Pre-trained models; Myocarditis; CMR; Convolutional neural networks

Contents

1 Introduction

Myocarditis is a type of cardiovascular disease that is leads to inflammation of the heart muscle and threatens human health. This disease is caused by factors such as viral or bacterial infections including COVID-19. Cardiac magnetic resonance (CMR) imaging provides the possibility of anatomical and functional imaging and accurate assessment of the heart. CMR imaging is the one of the most commonly used technologies in myocarditis identification. Three diagnostic targets for the three recommended CRI criteria are myocardial edema, hyperemia/capillary leak, and scar. However, there are some challenges for specialist physicians to identify myocarditis with CMR due to factors such as low contrast, various noises, and invisible mass or lesion in the images. In recent years, classification using machine learning has been a powerful tool and application in medical imaging to help disease diagnosis.

Two paradigms of deep learning approaches for improving classification accuracy have been proposed: transfer learning and self-supervised learning. Transfer learning aims to leverage data-rich source tasks to help with the learning of a data-deficient target task. However, it might lead to two problems which are domain discrepancy and overfitting. Additionally, expert manually-annotated medical images are expensive to acquire. This issue can be partially addressed by self-supervised learning which aims to learn meaningful representations of input data without using human annotations. Self-supervised representation learning has greatly advance unsupervised training of deep image models. Some recent works study self-supervised representation learning based on contrastive learning [1], and Khosla et al. extend the concept of the self-supervised contrastive learning to the fully-supervised setting [2], allowing to effectively leverage of label information.

In recent years, other techniques built upon transfer learning and self-supervised learning are also proposed. In this work, we aim to mitigate data deficiency and improve the classification accuracy of CMR images by developing sample-efficient methods to train highly-performant deep learning models on the basis of these two methodologies. Specifically, using the benchmark Z-Alizadeh-Sani myocarditis dataset which contains 4686 myocarditis CMRs and 2449 normal CMRs, we synergistically integrated self-supervised learning with transfer learning to learn powerful and unbiased feature representations to improve the classification accuracy. We also experimented with a label smoothing loss function for reducing the risk of overfitting and demonstrated the effectiveness of our proposed methods even when the number of training CMR images is limited. We achieved a 94.79% prediction accuracy of myocarditis using CMR images.

2 Related Work

Our work draws on existing literature in transfer learning, self-supervised representation learning, and supervised contrastive learning. Here we focus on the most relevant papers. The strengths and limitations of these methods are discussed in detail in the following sections.

2.1 Transfer Learning

Transfer learning aims to use pre-trained models obtained from sources with rich data to help with solving the problem of a new task involved with a deficient data source (CMR-based diagnosis of myocarditis in our case). It is normally performed by taking a standard neural architecture along with its pretrained weights on large-scale datasets such as ImageNet [3], and then fine-tuning the weights on the target task [4]. This method is efficient and effective in tackling new problems since the pretrained models are likely to have mature features already and thus can give superior performance. In the medical domain, transfer learning has been widely used in medical image classification and recognition tasks, such as tumor classification [5], pneumonia detection [6], and skin cancer classification [7]. The most commonly used approach is Convolutional Neural Network (CNN).

Convolutional neural networks are constructed using several convolution layers that use learnable filters or kernels to identify patterns in images such as edges, texture, color, and shapes. It is a deep learning technique that can be used in image classification, by taking an image as input and allocating importance (biases and learnable weights) to different objects/aspects in the image [8]. There are plenty of models available for medical imaging use, among which the commonly used ones are as follows: VGGNet [9], GoogleNet [10], ResNet [11], DenseNet [12], and EfficientNet [13]. In experiments done in [14], which use all of the five models mentioned above to diagnose skin lesions from photographs, all give 94% or more in terms of accuracy, with EfficientNet giving 96.7% as the highest.

Closely related to our work, Sharifrazi1 et al. [15] used the same Z-Alizadeh Sani myocarditis dataset and applied CNN combined with clustering (CNN-KCL) for the same image classification task. With k-means clulstering performed before three convolutional layers, they achieved a classification accuracy of 97.41%. Our work built upon the ResNet18 backbone which has more convolutional layers and a heavier structure. Recently, Shoeibi et al. [16] also used the same dataset but with CycleGAN data augmentation and employed various deep learning pretrained models such as ResNet and EffecientNet and achieved good results as well.

While effective in general, transfer learning may be suboptimal due to the fact that the source data may have a large discrepancy with the target data in terms of visual appearance of images and class labels, which causes the feature extraction network biased to the source data and generalizes less well on the target data. A recent study in [17] explores the properties of transfer learning for medical imaging tasks and finds that the standard large networks pretrained on ImageNet are often over-parameterized and may not be the optimal solution for medical image diagnosis, specifically in terms of the category of images (e.g. medical images vs. general images from ImageNet) and the methods of medical imaging (e.g. CMR vs. other common methods such as CT or X-ray). Additionally, due to the large scale of the data-rich sources, models pretrained on them usually have high complexity and therefore it is easy to cause the overfitting problem when being applied to the target data [4].

The cross-entropy loss is the most widely used loss function for supervised learning of deep classification models. Several works have explored shortcomings of this loss, such as lack of robustness to noisy labels [18] and the possibility of poor margins [19], leading to reduced generalization performance. Alternative losses have been proposed, but the most effective ideas in practice have been approaches that change the reference label distribution, such as label smoothing [20, 21], data augmentations such as Mixup [22], and knowledge distillation [23].

Label smoothing is a regularization technique that addresses both overfitting and overconfident problems. It has been used in many state-of-the-art models, including image classification[21]. In practice, the model applies the softmax function to the penultimate layer's logit vectors and compute cross entropy only with the "soft" targets from the dataset, and with a weighted mixture of "hard" targets with the uniform distribution, to improve the accuracy [20]. The largest logit gaps can be put into the softmax function with one-hot encoded labels, and as a result, large logit gaps with restricted gradient will reduce the model's prediction confidence necessarily. Label smoothing helps the representations of training examples from the same class to group in tight clusters [21]. [21] also empirically showed that apart from improving generalization, label smoothing can also improve model calibration, which can significantly improve beam-search.

2.2 Self-supervised Learning

The idea of self-supervised learning is to construct some auxiliary tasks based on the data without using human-annotated labels and force the network to learn meaningful representations by performing the auxiliary tasks well. Initial works in self-supervised representation learning focused on the problem of learning embeddings without labels such that a low-capacity (commonly linear) classifier operating on these embeddings could achieve high classification accuracy [24].

Although self-supervised learning has only recently become viable on standard image classification datasets, it has already seen some application within the medical domain. Some works have attempted to design domain-specific pretext tasks. For example, Zhu et al. propose a pretext task Rubik's cube+ [25], forcing networks to learn translation and rotation invariant features from the original 3D medical data and tolerate the noise of the data at the same time. Compared to the strategy of training from scratch, fine-tuning from the Rubik's cube+ pre-trained weights can remarkably boost the accuracy of 3D neural networks on various tasks, such as cerebral hemorrhage classification and brain tumor segmentation, without the use of extra data.

2.2.1 Self-supervised Contrastive Learning

Contrastive learning applied to self-supervised representation learning has seen a resurgence in recent years, leading to state of the art performance in the unsupervised training of deep image models. Modern batch contrastive approaches subsume or significantly outperform traditional contrastive losses such as triplet, max-margin and the N-pairs loss.

Some recent works study self-supervised representation learning based on instance discrimination with contrastive learning [1]. Given an original image in the dataset, contrastive self-supervised learning (CSSL) performs data augmentation of this image and obtains two augmented images where the first one is referred to as query and the second one as key [26]. Two networks are used to obtain latent representations of the two images respectively. A query and a key belonging to the same image are labeled as a positive pair. A query and a key belonging to different images are labeled as a negative pair. The auxiliary task is: given a (query, key) pair, judging whether it is positive or negative. Given a new pair $(\boldsymbol{q}_j, \boldsymbol{k}_j)$ obtained from a new image, a contrastive loss can be defined as

$$\mathcal{L} = -\log \frac{\exp(\boldsymbol{q}_j \cdot \boldsymbol{k}_j/\tau)}{\exp(\boldsymbol{q}_j \cdot \boldsymbol{k}_j/\tau) + \sum_i \exp(\boldsymbol{q}_j \cdot \boldsymbol{k}_j)} \tag{1}$$

where τ is an annealing parameter. The weights in the encoders are learned by minimizing the losses of such a form.

Oord et al. propose contrastive predictive coding (CPC) to extract useful representations from high-dimensional data [27]. Bachman et al. propose a self-supervised representation learning approach based on maximizing mutual information between features extracted from multiple views of a shared context [28]. Momentum Contrast (MoCo) expands the idea of contrastive learning with an additional dictionary and a momentum encoder [29]. More recently, Chen et al. present a simple framework for contrastive learning (SimCLR) with larger batch sizes and extensive data augmentation [30]. SimCLR learns representations by maximizing agreement between differently augmented views of the same data example via a contrastive loss in a hidden representation of neural nets. These methods were the first to achieve linear classification accuracy approaching that of end-to-end supervised training, and they can be used in different domains where data annotations is scarce or challenging.

Self-supervised contrastive learning have also been widely exploited in the medical domain. Chaitanya et al. studied contrastive learning of global and local features for medical image segmentation with limited annotations [31]. Azizi et al. proposed a Multi-Instance Contrastive Learning (MICLe) method that uses multiple images of the underlying pathology per patient case to construct more informative positive pairs for self-supervised learning in multiple disease diagnosis [32].

2.2.2 Supervised Contrastive Learning

More recently, Khosla et al. propose a novel extension to the contrastive loss function that allows for multiple positives per anchor [2], thus adapting contrastive learning to the fully supervised setting, which allows to leverage label information more effectively. This model shows consistent outperformance over cross-entropy on other datasets and two ResNet variants. Furthermore, it provides a unifying loss function that can be used for either self-supervised or supervised learning. Our work is closely related to and builds upon the architecture of supervised contrastive learning proposed by Khosla and we experiment a more flexible loss function beyond supervised contrastive loss.

According to [2], due to the presence of labels, Eq. 1 is incapable of handling the case where more than one sample belong to the same class. Generalization to an arbitrary numbers of positives will result in a choice between multiple possible functions. The following equation presents the most straightforward way to generalize the previous contrastive loss to incorporate supervision.

$$\mathcal{L}_{out}^{sup} = \sum_{i \in I} \mathcal{L}_{out,i}^{sup} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_p / \tau)}{\sum_{a \in A(i)} \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_a / \tau)} \tag{2}$$

The supervised contrastive learning [2] provides good generalization to an arbitrary number of positives with increased contrastive power given more negatives. It also has the intrinsic ability to perform hard positive/negative mining. However, it has not seen any application in the medical imaging domain yet. To the best of our knowledge, we are the first to integrate the supervised contrastive learning into disease diagnosis tasks using the Z-Alizadeh Sani myocarditis dataset or other medical imaging datasets.

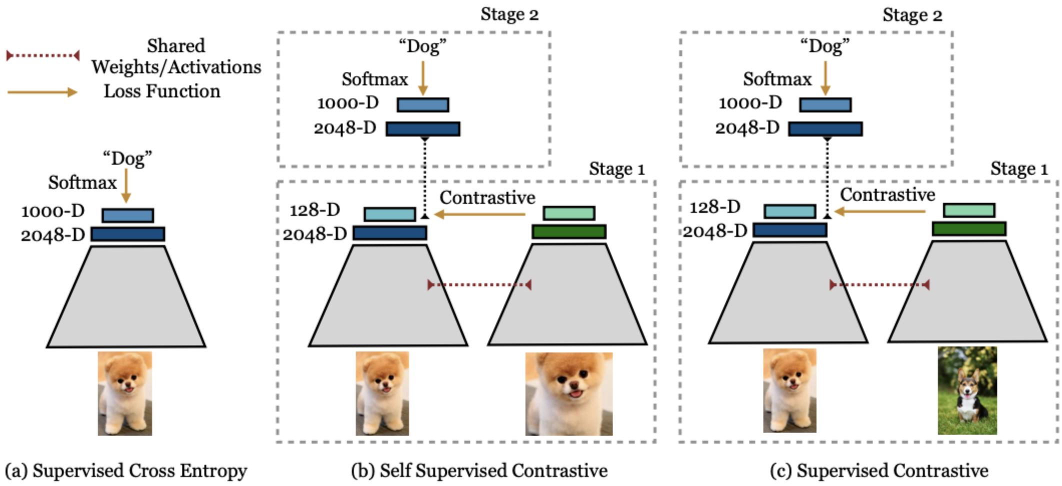

Figure 1: The supervised contrastive loss (right) learns representations using a contrastive loss, but uses label information to sample positives in addition to augmentations of the same image [2].

3 Solution

3.1 Label Smoothing

To deal with the problem of overfitting, we add label smoothing to the cross entropy loss, with the temperature scalar being 0.1. The model applies the softmax function and compute cross entropy with a weighted mixture of "hard" targets and the uniform distribution, to improve the accuracy and reduce prediction confidence necessarily.

3.2 Supervised Contrastive Learning

Our method is structurally similar to that used in [2] for supervised contrastive learning. Given an input batch of data, we first apply data augmentation twice to obtain two copies of the batch.

Both copies are forward propagated through the encoder network to obtain a 512-dimensional normalized embedding. In training, this representation is further propagated through a projection network that is discarded at inference time. The supervised contrastive loss is computed on the outputs of the projection network. However, different from the Supervised Contrastive Learning architecture which trains a linear classifier on top of the frozen representations using a cross-entropy loss. We combine cross-entropy loss with supervised contrastive loss with different weights and cocurrently train the model based on the integrated loss. This approach is innovative because it integrates two separate, static stages and provides more flexibility in parameter tuning. The overall loss function of our SupConCE is

$$SupConCELoss = \beta ContrastiveLoss + CrossEntropyLoss \tag{3}$$

Specifically, the main components of our framework are:

(1) Data Augmentation. For each input sample, x, we generate two random augmentations using resize or random horizontal flip, each of which represents a different view of the data and contains some subset of the information in the original sample.

(2) Encoder Network, which maps x to a representation vector r. Both augmented samples are separately input to the same encoder, resulting in a pair of representation vectors. The representation vector is then normalized to the unit hypersphere with a 512-dimensional embedding.

(3) Projection Network, which maps r to a vector z. We instantiate a single hidden layer of size 128. We discard the projection head at the end of contrastive training.

(4) Classification Head, which takes the embedding and outputs logits.

An illustration of our model is shown below.

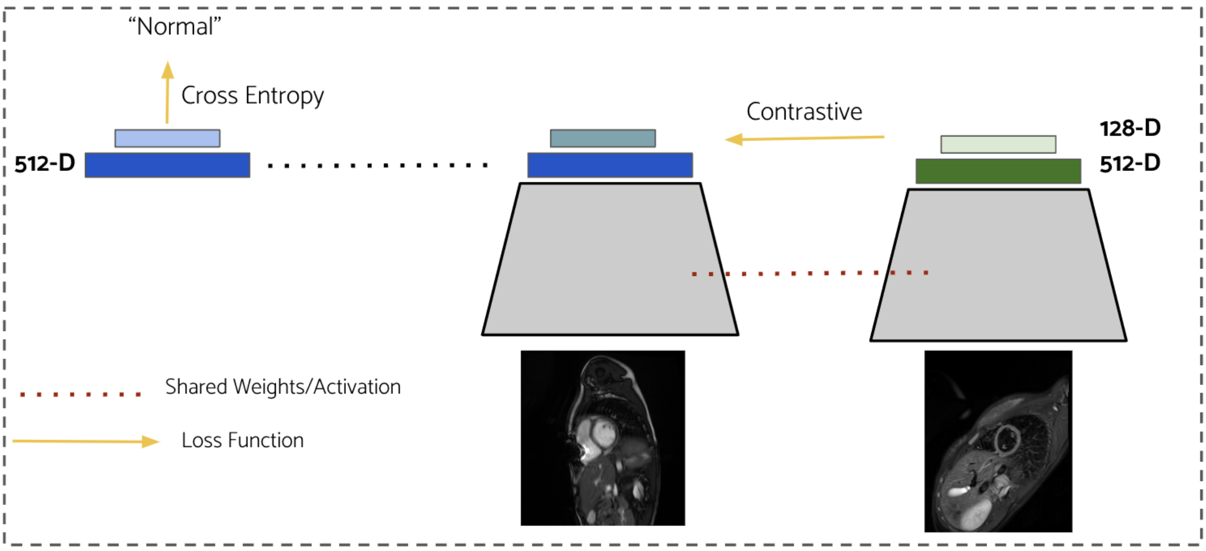

Figure 2: The Supervised Contrastive Cross Entropy SupConCE Architecture

4 Results

4.1 Experimentation protocol

We evaluate our SupConCE loss by measuring classification accuracy on the common myocarditis image classification benchmark Z-Alizadeh Sani myocarditis dataset. We also show how performance varies with changes to hyperparameters. For the backbone architecture we experimented with two commonly used encoder architectures: ResNet-18 and ResNet-50. The normalized activations of the final pooling layer are used as the representation vector. We experimented with two implementations of the data augmentation module: Resize and Random Horizontal Flip.

We use the common metrics for image classification such as accuracy, precision, recall, specificity, and F-1 score. Experiment results for different backbone models, with or without label smoothing, with or without supervised contrastive loss are summarized below.

4.2 Data table

Loss	Architecture	Acc (%)	Prec (%)	Rec (%)	Spec (%)	F1 (%)
Cross Entropy	ResNet18	92.32	98.18	82.77	99.14	89.82
Label Smoothing	ResNet18	**94.97**	100.00	87.69	100.00	93.44
Label Smoothing	ResNet50	91.00	99.61	78.31	100.00	87.68
SupConCE + Label Smoothing	ResNet18	93.27	100.00	83.54	100.00	**97.62**

Table 1: Model Performance

The table above shows the comprehensive results using different network architectures with different loss functions and their combinations, in terms of accuracy, precision, recall rate, specificity and F1 score. Compare the first two rows, we found that adding label smoothing in addition to the cross entropy loss gives considerable improvement for all metrics of measurement. Especially for accuracy, it achieves 94.97%, it is not only 2.65% higher than without the label smoothing, but also the highest of all four experiments. While changing the architecture from ResNet18 to ResNet50, it yields weaker performance, which is contrary to common cases and expectations. The reason is discussed in the following section in detail. As a result, we chose ResNet18 for the following experiments. Further adding supervised contrastive learning based on the loss function used in experiment of row two, we found that the accuracy has slightly declined, but F1 score has been significantly improved, achieving a 97.62%, which is also the highest of all four experiments.

4.3 Graphs

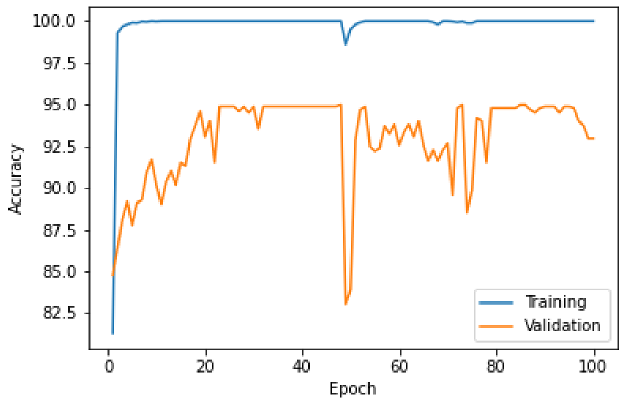

Figure 3: Training and Validation Accuracy with Label Smoothing

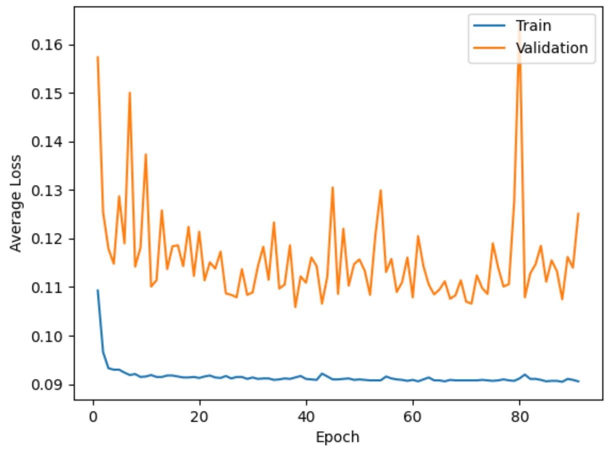

Figure 4: Training and Validation Loss with Supervised Contrastive Loss and Label Smoothing

As shown in Figure 3, for the model with ResNet18 backbone and label smoothing, the training accuracy keeps increasing and until reaching 100%. The validation accuracy climbs up from 85% to slightly over 95% and stablizes there. As shown in Figure 4, for the model with both contrastive learning and label smoothing, the validation loss has a decreasing trend at the beginning but is volatile throughout all the training epochs. This shows that the model does not generalize very well and might suffer from overfitting. A more detailed discussion about this model is listed in the next section.

5 Discussion

For CNN methods, compared to papers that used the same dataset as ours, the CNN-KCL model proposed in [15] outperforms our model in terms of accuracy. Since its model uesd a lighter structured CNN (a self-built network with only 3 convolutional layers), it is a simpler and lighter network than ours. Considering the characteristics of the MRI image dataset, it is possible that simpler network can give better performance. What's more, [16] used CycleGan as data augmentation and it also outperforms our result. However, we think their method is not intuitive and may not give best result. To some extent, data augmentation generated by CycleGan is not suitable for this particular dataset, and it does not match the traditional cross entropy loss. Because the key difference point for diagnosis of a disease from MRI images are usually uniform, and such augmentation method could unintentionally crop out the key part and thus miss the correct classification. Therefore, we believe simpler augmentation methods we used, such as random flipping, are more suitable for this dataset.

One interesting finding that is worth noticing is that when we applied CNN with their pretrained weights, we found ResNet18 outperformed ResNet50, although the former is a lighter weighted network. This may be caused by the following reasons. Firstly, ResNet models are trained based on ImageNet, in which the objects used for training are all common objects, such as chairs and dogs. However, our dataset is of medical data and specifically MRI images. So the difference of domain of data may be the primary reason for this phenomenon. Moreover, RestNet18 has 11,689,512 parameters while RestNet50 has 25,557,132 parameters, which is almost double in the number of parameters. Since our training data is a relatively small data set with 4507 images, the parameter updates of ResNet50 might not be as efficient as ResNet18 in backpropagation. Besides, since we trained our model using the Google Colaboratory, we only train for 100 epochs, which might not reach the potential of every model we have proposed. It is possible that better performance and results are to be obtained with more training epochs. But if time permits, we will continue our experiments on a greater scale.

In this work, we experimented with the supervised contrastive learning approach on the Z-Alizadeh Sani myocarditis dataset with a ResNet backbone model. In fact, other backbone models such as the vision transformer(ViT) can be another feasible method of transfer learning apart from CNN, according to our initial literature review. It has been effectively used in some classification tasks of medical imaging, including various image modalities. Since there is evidence showing that it can outperform CNN under some scenarios, it could be used in the future study and get its results competitive as the ones of CNNs.

Moreover, we mainly study the supervised contrastive learning. In fact, self-supervised learning methods might also be helpful in producing better prediction accuracy such as self-supervised contrastive learning with memory bank and semi-supervised learning, both of which have successful practices on other medical data set. For example, He et al. propose an Self-Trans approach, which integrates contrastive self-supervision [4] into the transfer learning process. They first train their model on unlabeld COVID CT images to remove some domain discrepency brought by pre-trained weights from ImageNet. Then, they construct an auxiliary task of judging whether two images created via random data augmentation are augments of the same original image on the COVID CT dataset with a memory bank to fully explore the structure and information of the CT scans. Inspired by this, we are going to explore other self-supervised learning techniques on the myocarditis dataset to explore more possibilities if given more time.

6 Conclusion

In this work, we synergistically integrate contrastive learning with transfer learning to learn powerful and unbiased feature representations on the benchmark Z-Alizadeh-Sani myocarditis dataset. To the best of our knowledge, we are the first to apply supervised contrastive learning on medical imaging for disease diagnosis. We also contribute to the model architecture of supervised contrastive learning with an integrated loss function combining both contrastive loss and cross entropy loss together. This will make the model more flexible in training and has the potential to achieve better result. Moreover, we integrated label smoothing as a regularization technique to reduce the risk of overfitting and demonstrate the effectiveness of our proposed methods even when the number of training CMR images is limited. We achieved a highest 94.97% accuracy in myocarditis identification from CMR images.

References

[1] Zhirong Wu, Yuanjun Xiong, Stella X. Yu, and Dahua Lin. 2018. Unsupervised Feature Learning via Non-parametric Instance Discrimination. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, (May 2018), 3733-3742.
DOI:https://doi.org/10.1109/cvpr.2018.00393

[2] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. arXiv:2004.11362. Retrieved from https://arxiv.org/abs/2004.11362

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li,Li Kai and Fei-Fei Li. 2009. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248-255.
DOI: https://doi.org/10.1109/CVPR.2009.5206848

[4] Xuehai He, Xingyi Yang, Shanghang Zhang, Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie. 2020. Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans. *medRxiv*, (Apr. 2020).
DOI: https://doi.org/10.1101/2020.04.13.20063941

[5] Benjamin Q. Huynh, Hui Li, and Maryellen L. Giger. 2016. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging* 3, 3 (Aug. 2016), 034501.
DOI:https://doi.org/10.1117/1.jmi.3.3.034501

[6] Vikash Chouhan, Sanjay Kumar Singh, Aditya Khamparia, Deepak Gupta, Prayag Tiwari, Catarina Moreira, Robertas Damaševičius, and Victor Hugo C. De Albuquerque. 2020. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. *Applied Sciences* 10, 2 (Jan. 2020), 559. DOI:https://doi.org/10.3390/app10020559

[7] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639 (Jan. 2017), 115-118. DOI:https://doi.org/10.1038/nature21056

[8] Hadi Mahami, Navid Ghassemi, Mohammad Tayarani Darbandy, Afshin Shoeibi, Sadiq Hussain, Farnad Nasirzadeh, Roohallah Alizadehsani, Darius Nahavandi, Abbas Khosravi, Saeid Nahavandi. 2020. Material Recognition for Automated Progress Monitoring using Deep Learning Methods. arXiv:2006.16344. Retrieved from https://arxiv.org/abs/2006.16344

[9] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556. Retrieved from https://arxiv.org/abs/1409.1556

[10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Jun. 2015). DOI:https://doi.org/10.1109/cvpr.2015.7298594

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. 2015. Deep residual learning for image recognition. arXiv 1512.03385. Retrieved from https://arxiv.org/abs/1512.03385

[12] Gao Huang, Zhuang Liu, Laurens Van der Maaten, and Kilian Q. Weinberger. 2017. Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Aug. 2017). DOI:https://doi.org/10.1109/cvpr.2017.243

[13] Mingxing Tan, Quoc V. Le. 2019. Rethinking model scaling for convolutional neural networks. arXiv:1905.11946.
Retrieved from https://arxiv.org/abs/1905.11946

[14] Emma Rocheteau, Doyoon Kim. 2020. Deep Transfer Learning for Automated Diagnosis of Skin Lesions from Photographs. arXiv: 2011.04475. Retrieved from
https://arxiv.org/abs/2011.04475

[15] Danial Sharifrazi, Roohallah Alizadehsani, Javad Hassannataj Joloudari, Shahab S. Band, Sadiq Hussain, Zahra Alizadeh Sani, Fereshteh Hasanzadeh, Afshin Shoeibi, Abdollah Dehzangi, Mehdi Sookhak, Hamid Alinejad-Rokny. CNN-KCL: Automatic myocarditis diagnosis using convolutional neural network combined with k-means clustering. *Mathematical Biosciences and Engineering*, 19, 3 (Jan. 2022), 2381-2402. DOI: 10.3934/mbe.2022110

[16] Shoeibi, A., Ghassemi, N., Heras, J., Rezaei, M., Gorriz, J.M. (2022). Automatic Diagnosis of Myocarditis in Cardiac Magnetic Images Using CycleGAN and Deep PreTrained Models. *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications.* IWINAC 2022. Lecture Notes in Computer Science, vol 13258. DOI: https://doi.org/10.1007/978-3-031-06242-1_15

[17] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, Samy Bengio. 2019. Transfusion: Understanding Transfer Learning for Medical Imaging. arXiv:1902.07208.
Retrievd from https://arxiv.org/abs/1902.07208

[18] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural net- works with noisy labels. In Advances in neural information processing systems, pages 8778–8788, 2018.

[19] Gamaleldin Elsayed, Dilip Krishnan, Hossein Mobahi, Kevin Regan, and Samy Bengio. Large margin deep networks for classification. 2018. arXiv:1803.05598.
Retrievd from https://arxiv.org/abs/1803.05598

[20] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[21] Rafael Müller, Simon Kornblith, Geoffrey E. Hinton. 2019. When does label smoothing help? arXiv:1906.02629. Retrieved from https://arxiv.org/abs/1906.02629

[22] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. 2017. mixup: Beyond empirical risk minimization. arXiv:1710.09412. Retrieved from https://arxiv.org/abs/1710.09412

[23] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. 2015. arXiv:1503.02531. Retrieved from https://arxiv.org/abs/1503.02531

[24] Carl Doersch, Abhinav Gupta, and Alexei A. Efros. 2015. Unsupervised Visual Representation Learning by Context Prediction. 2015 IEEE Interna- tional Conference on Computer

Vision (ICCV), (May 2015). DOI:https://doi.org/10.1109/iccv.2015.167

[25] Jiuwen Zhu, Yuexiang Li, Yifan Hu, Kai Ma, S. Kevin Zhou, and Yefeng Zheng. 2020. Rubik's Cube+: A self-supervised feature learning framework for 3D medical image analysis. Medical Image Analysis 64, (Aug. 2020), 101746. DOI:https://doi.org/10.1016/j.media.2020.101746

[26] Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality Re- duction by Learning an Invariant Mapping. 2006. Dimensionality Reduction by Learning an Invariant Mapping. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR 2), (Jun. 2006), 1735-1742. DOI:https://doi.org/10.1109/cvpr.2006.100

[27] Aaron van den Oord, Yazhe Li, Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. arXiv:1807.03748. Retrieved from https://arxiv.org/abs/1807.03748

[28] Philip Bachman, R Devon Hjelm, William Buchwalter. 2019. Learning representations by maximizing mutual information across views. arXiv:1906.00910. Retrieved from https://arxiv.org/abs/1906.00910

[29] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Mar. 2020).
DOI: https://doi.org/10.1109/cvpr42600.2020.00975

[30] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. *In Proceedings of the 37th International Conference on Machine Learning (ICML'20)*,Article 149, (Jul. 2020), 1597–1607.
DOI: https://dl.acm.org/doi/10.5555/3524938.3525087

[31] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, Ender Konukoglu. 2020. Contrastive learning of global and local features for medical image segmentation with limited annotations. arXiv:2006.10511. Retrieved from https://arxiv.org/abs/2006.10511.

[32] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, Vivek Natarajan, Mohammad Norouzi. 2021. Big Self-Supervised Models Advance Medical Image Classification. arXiv:2101.05224. Retrieved from https://arxiv.org/abs/2101.05224