# A Perspective on Generalizing Learning Models for Out-of-Domain Images

Margarida Gouveia[1,2]
margarida.g.gouveia@inesctec.pt
Hélder P. Oliveira[1,3]
helder.f.oliveira@inesctec.pt
Tania Pereira[1]
tania.pereira@inesctec.pt

[1] INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal
[2] FEUP - Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal
[3] FCUP - Faculty of Science, University of Porto, 4169-007 Porto, Portugal

## Abstract

Deep learning methods show a high performance in different problems. However, when the target domain has a distribution shift from the source domain, a significant reduction in the performance of the models occurs. Learning models applied to medical imaging analysis are limited by characteristics of the medical datasets, which are typically small, low representative and not completely annotated, resulting in problems of domain shift. Typically, these problems are addressed by strategies for domain generalization or adaptation. The literature shows the need to create learning models with stronger generalization capability to deal with two problems: lack of generalization (P1) for cohort populations and (P2) for distinct imaging modalities. This paper gives a perspective on how to address these two problems, mainly centred on the importance of strategies to force the learning of domain invariant feature representations.

## 1 Introduction

Deep learning has surpassed human performance in different tasks while learning directly from data, depending on the high quality and quantity of the training data [6]. The assumption that the source/reference domain (used during training) presents the same data distribution as the target domain (used for testing and deployment) is the base of the application of learning methods [3]. However, domain shifts are a prevalent problem in different applications, i.e., a target domain out-of-distribution causes a decrease in the performance of the models and affects a large-scale deployment [3, 10]. A large distribution shift to the target domain occurs when the source domain does not capture the data distribution complexity due to data scarcity or low sample variability [10]. This is tackled by techniques of domain generalization (DG), under the assumption that the target domain is not accessed during train, and domain adaptation (DA), where the target data (sparsely labelled or unlabelled) is available for model adaptation [7].

In medical image analysis, the problems of domain shift are particularly prevalent due to the low variability of small, imbalanced and not thoroughly annotated datasets [3, 6]. Patient privacy, institutional policies and costs limit data acquisition [3]. The annotating process is expensive, time-consuming, and demands the expertise of healthcare professionals [3]. Small datasets also are more prone to a lack of demographic representation [3] and do not capture heterogeneity in lesions and anatomical structures phenotype [8]. Different hospitals and healthcare professionals follow distinct acquisition and annotation protocols and use equipment with various specifications from multiple vendors, resulting in heterogeneous datasets [3].

In summary, the prevalence of data scarcity and domain shift problems deeply impacts the clinical deployment of learning models. In healthcare, the quality of diagnostic decisions can be a matter of life and death, which demands a high accuracy of computer-aided diagnosis systems and a trust that the system is capable of processing new patient cases with high performance. For all these reasons, addressing domain shift problems is a critical step towards the safe and robust deployment of learning models capable of generalizing well to unseen domains and being applied to the clinical environment [3, 7, 10].

## 2 Literature Review

The most common strategy to increase the training set variability is data augmentation [7] by applying image transformations [9]. An alternative is to generate new synthetic data using Generative Adversarial Networks (GANs) [7] or a Mixup strategy [5]. For the cases where the target domain is identified, GANs can translate images from the target to the source domain [3]. The translation can be between identical and distinct imaging modalities [2].

For a generalization between domains, representation learning is used to force learning models to learn feature representations invariant to domain changes [7], using adversarial learning [5]. It is also possible to train models to adapt their parameters according to the domain of the image to process [7]. Finally, some methods use domain knowledge incorporation for enforcing invariance: one example is symmetry-based regularization [1]. Features without domain bias can also be learned using pre-train models using self-supervised learning [10]. This learning strategy uses unlabelled data with free labels learned from distinct pretext tasks [2] or using contrastive learning [7, 10]. In contrastive learning, the model learns transformation-invariant representations using pairs of original and transform images [4]. Some works suggest the benefit of pre-train the models using large image datasets like ImageNet [1].

The data scarcity and domain shift problems are identified and addressed in the literature with solutions based on DA and DG, for problems in-and-out of the medical domain. However, the solutions are case-based and fail to work with different sources of domain shift. In addition, it is not possible to find an objective analysis of which methods are more suitable for dealing with a specific cause of domain shift.

## 3 Prospective on Learning Models Generalization

The literature analysis shows the need to explore different methods to surpass the consequences of low data variability during training and the consequent domain shift when learning models process new data in medical imaging analysis problems. In the future, two problems of out-of-domain generalization should be addressed: (1) lack of generalization between cohort populations with different demographics and various anatomic/pathological phenotypes, usually represented in datasets acquired and annotated using distinct protocols; and (2) lack of generalization for images of different medical modalities but from the same anatomical structure and pathological manifestation. Figure 1 shows a graphical representation of the causes and consequences of the lack of generalization for each specific problem.

### 3.1 Generalization for cohort populations

Deep learning can be applied in several medical image analysis problems of classification, detection and segmentation, reaching high performance. However, the application of the model in clinical settings is limited by the decrease in performance of the models when applied to new data. This is a direct result of different datasets representing cohort populations with images acquired and annotated using distinct protocols. The learning models must be invariant to domain changes in the datasets.

The construction of models with generalization capability should start with a characterization of the sources of domain shift, such as changes in the acquisition and annotation protocol and the changes in the population represented, namely the analysis of demographic and clinical factors that can impact the anatomical structure and the pathological manifestation. Understanding how an alteration of these factors can impact the image will allow the creation of data augmentation techniques or even the use of generative models to create new samples capable of increasing the diversity of the training set.

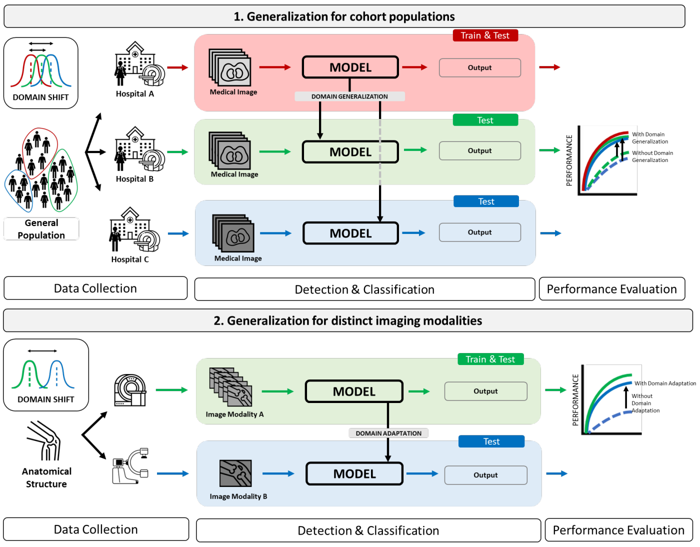


Figure 1: Graphical representation of the problems of (1) lack of generalization between cohort populations; and (2) lack of generalization for images of different medical modalities.

In addition to increasing the training set variability to be representative of different cohort populations, the construction of the learning models must ensure the learning of domain-invariant features, using adversarial learning or regularization methods that penalize the learning of domain-specific features. However, the application of these methods demands annotated data from distinct domains, which is challenging and not always possible.

### 3.2 Generalization for distinct imaging modalities

Constructing a learning model capable of diagnosing the same pathological manifestation using different image modalities offers a versatile approach for clinical deployment. This optimizes clinical workflows and can potentially reduce patient exposure to radiation and healthcare costs. In addition, these models enable optimal data utilization and resilience to data availability challenges, such as the temporarily unavailable of a specific imaging modality.

One of the possible strategies to adopt is to include a pre-processing step in which a generative model is used to generate synthetic images in one modality based on information from another modality. This can allow the training of the main model in just one imaging modality. An alternative is to use just one model trained to learn domain-invariant features present in the different imaging modalities. The training can be done using adversarial learning and regularization methods.

## 4 Conclusion

In summary, the problem of domain shift impacts the performance of learning models when applied to new data domains. This is a problem that especially impacts the field of medical image analysis due to the characteristics of the medical image datasets, which are usually small, imbalanced and not completely annotated. The problem of domain shift is addressed in the literature with strategies of DA and DG for applications in-and-out of the medical image analysis domain.

In the future, to address the problems of lack of generalization for cohort populations and distinct imaging modalities should be use strategies based on implementing methods that force models to learn domain-invariant features during train. The strategies must account for the domain knowledge obtained in the study of the problem at hand, mainly the characterization of the domain-shift problem. One of the main challenges that the application of these methods could encounter is the need for annotated data from distinct domains.

## Acknowledgements

This work is financed by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within the project LUCCA, with reference 2022.03488.PTDC, and a PhD Grant Number: 2023.02607.BD.

## References

[1] Eduardo Castro, Jose Costa Pereira, and Jaime S. Cardoso. Symmetry-based regularization in deep breast cancer screening. *Medical Image Analysis*, 83:102690, January 2023. ISSN 1361-8415. doi: 10.1016/j.media.2022.102690. URL https://www.sciencedirect.com/science/article/pii/S1361841522003188.

[2] Banafshe Felfeliyan, Abhilash Hareendranathan, Gregor Kuntze, Jacob Jaremko, and Janet Ronsky. MRI Knee Domain Translation for Unsupervised Segmentation By CycleGAN (data from Osteoarthritis initiative (OAI)). In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4052–4055, November 2021. doi: 10.1109/EMBC46164.2021.9629705. ISSN: 2694-0604.

[3] Hao Guan and Mingxia Liu. Domain Adaptation for Medical Image Analysis: A Survey. *IEEE transactions on bio-medical engineering*, 69(3):1173–1185, March 2022. ISSN 0018-9294. doi: 10.1109/TBME.2021.3117407. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9011180/.

[4] Zheren Li, Zhiming Cui, Sheng Wang, Yuji Qi, Xi Ouyang, Qitian Chen, Yuezhi Yang, Zhong Xue, Dinggang Shen, and Jie-Zhi Cheng. Domain Generalization for Mammography Detection via Multi-style and Multi-view Contrastive Learning. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Lecture Notes in Computer Science, pages 98–108, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87234-2. doi: 10.1007/978-3-030-87234-2_10.

[5] Egor Panfilov, Aleksei Tiulpin, Stefan Klein, Miika T. Nieminen, and Simo Saarakkala. Improving Robustness of Deep Learning Based Knee MRI Segmentation: Mixup and Adversarial Domain Adaptation. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 450–459, October 2019. doi: 10.1109/ICCVW.2019.00057. URL https://ieeexplore.ieee.org/document/9022164/. Conference Name: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) ISBN: 9781728150239 Place: Seoul, Korea (South) Publisher: IEEE.

[6] Masayuki Tsuneki. Deep learning models in medical image analysis. *Journal of Oral Biosciences*, 64(3):312–320, September 2022. ISSN 1880-3865. doi: 10.1016/j.job.2022.03.003.

[7] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip S. Yu. Generalizing to Unseen Domains: A Survey on Domain Generalization, May 2022. URL http://arxiv.org/abs/2103.03097. arXiv:2103.03097 [cs].

[8] Xiaoqin Wang, Gongbo Liang, Yu Zhang, Hunter Blanton, Zachary Bessinger, and Nathan Jacobs. Inconsistent Performance of Deep Learning Models on Mammogram Classification. *Journal of the American College of Radiology*, 17(6):796–803, June 2020. ISSN 1546-1440. doi: 10.1016/j.jacr.2020.01.006. URL https://www.sciencedirect.com/science/article/pii/S1546144020300284.

[9] Dong Zhang, Yi Lin, Hao Chen, Zhuotao Tian, Xin Yang, Jinhui Tang, and Kwang Ting Cheng. Understanding the Tricks of Deep Learning in Medical Image Segmentation: Challenges and Future Directions, May 2023. URL http://arxiv.org/abs/2209.10307. arXiv:2209.10307 [cs].

[10] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain Generalization: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4396–4415, April 2023. ISSN 1939-3539. doi: 10.1109/TPAMI.2022.3195549. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.