Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/35/20230407

# Enhancing medical image classification with convolutional neural networks through transfer learning: A comprehensive review

**Chengkang Gu**

School of Computer Science, University of Southampton, Southampton, UK, SO17 1BJ

cg1y21@soton.ac.uk

**Abstract.** Medical imaging has become crucial in diagnostics, but manual analysis of the vast amount of data is cumbersome. Transfer learning, utilizing techniques like data augmentation, has addressed the challenges faced by deep learning methods like CNNs in this domain, improving classification accuracy. This study offers a comprehensive review of the applications and potential benefits of deep learning and transfer learning in medical image classification while also shedding light on prospective challenges and avenues for future research. After a comprehensive review and analysis of the literature, it can be inferred that transfer learning addresses the primary challenges deep learning faces in the medical field, enhancing its applicability. While deep learning has substantial potential, it grapples with issues of model "black-box" interpretability, where decisions made by the model are hard for clinicians to understand, and potential breaches of individual data security given the sensitive nature of medical records.

**Keywords:** deep learning, CNN, transfer learning, medical image.

## 1. Introduction

As the most prominent subfield in contemporary artificial intelligence, machine learning continues to garner extensive attention from the academic and medical communities. In particular, deep learning and neural networks have brought about revolutionary breakthroughs [1]. CNNs and deep learning have demonstrated exceptional performance in the visual domain, especially in the globally renowned computer vision competition, ImageNet Classification, where they significantly outperformed with an error rate far below the second place's 26% [2]. This achievement can undoubtedly be recognized as a milestone advancement. The formidable capability of CNNs in image classification has also been validated, with its performance in ILSVRC surpassing human capabilities and reducing the error rate to near the Bayesian limit, marking a significant breakthrough in image classification tasks [1]. From the 1970s to 2016, deep learning experienced rapid growth in the medical field, receiving widespread attention and adoption in the medical imaging community [3].

However, medical imaging has always faced challenges related to small datasets and data scarcity [4]. Transfer learning emerges as an effective solution, offering immense potential for addressing these issues and enhancing classification accuracy [5]. This article primarily investigates the application and performance enhancement of transfer learning in medical image classification, while also analyzing various techniques and domains of transfer learning.

## 2. Fundamentals of deep learning and transfer learning

With the incorporation of CNN frameworks into medical image classification, these networks have begun addressing challenges specific to this field [6]. As a specialized structure in deep learning, owing to its unique structure and parameter-sharing mechanism, CNNs have demonstrated exceptional performance in image recognition and classification tasks. A typical CNN primarily consists of convolutional layers, pooling layers, and activation layers [1]. In addition, an activation function needs to be provided, usually a RELU function.

In traditional machine learning and data mining, specific data models are often employed, leveraging a limited set of labeled data and a vast pool of unlabeled data to achieve predictive outcomes [8]. However, it can be difficult to preserve a fundamental assumption underlying conventional machine learning in real-world applications: the test and training data should come from comparable sources and display parallel feature distributions. [7, 8]. To address the high data demands and associated costs, incorporating data from various domains can be beneficial [7]. Transfer learning emerged as a solution to this problem, offering a way to reduce data collection costs while integrating data across domains to alleviate distribution discrepancies [9].

Given the definitions, when there's a disparity between the source domain (DS) and target domain (DT), as well as their respective learning tasks (TS and TT), transfer learning aims to apply knowledge from the source domain (DS) and its associated task (TS) to a predictive function, denoted as 'f', enhancing learning performance in the target domain [8]. For an in-depth understanding, readers are directed to the article "A Survey on Transfer Learning" [8].

## 3. Application of transfer learning in the medical field

The analysis of medical images demands a significant amount of effort, and relying solely on manual methods is inevitably inefficient and less accurate, especially with the growing volume of medical data. As previously mentioned, the introduction of CNNs as an assistant for image analysis presents a promising solution. Yet, medical data often suffers from scarcity, making the construction and training of a model from scratch challenging. Therefore, we need to incorporate transfer learning to utilize models pretrained on big data (from other domains) and transfer their knowledge to image classification tasks, reducing the requirement for labeled samples and enhancing the training accuracy [10].

Transfer learning, as a crucial strategy in deep learning, allows researchers to leverage models trained on large datasets, typically on tasks related to but distinct from our target objective, and apply these pretrained models to our specific goal. Popular models include VGG, ResNet, and Inception, all of which have been trained on extensive datasets and have acquired a wealth of features. By extracting features and knowledge from these pretrained models and fine-tuning them, the models can adapt to the unique characteristics of medical image classification and can be employed to address the challenge of medical image classification. For instance, the GoogLeNet model, trained on ImageNet, has been fine-tuned to cater to the specific domain of brain tumor classification [11]. In medical image classification, based on the nature of the task, different layers may need adjustments. For instance, research introduced the TruncatedTL strategy, which reuses and fine-tunes appropriate lower layers while discarding the rest. This not only yields superior medical image classification performance but also results in a compact model for efficient inference [12].

There have been preliminary attempts at implementing transfer learning in the clinical medical field. During the outbreak of COVID-19, the integration of transfer learning with medical techniques such as CT scans facilitated rapid detection, offering physicians a high-accuracy COVID-19 diagnostic tool and playing a significant role in pandemic control [13]. Other applications include oncological imaging, medical ultrasonography, brain segmentation, and so on [1]. However, how to further harness the attributes of transfer learning to enhance medical image classification remains an intriguing question worth exploring.

## 4. Enhancement of medical image classification through transfer learning

In the realm of medical imaging, transfer learning primarily offers enhancement through the previously mentioned models, fine-tuning, and another technique known as data augmentation. These three methodologies collectively contribute to the precision and efficiency of medical image classification.

While data augmentation, in essence, doesn't strictly fall under the category of transfer learning, reinforcement learning is often employed to supplement transfer learning by generating a larger volume of samples. This not only expands the dataset size but also enhances the model's generalization capabilities, prevents overfitting, and optimizes the model's performance. Data augmentation techniques can be broadly categorized into basic augmentation, deformable augmentation, deep learning-based augmentation, and other miscellaneous augmentation techniques [14]. Basic augmentation techniques encompass methods like mirror flipping, rotation, scaling, and translation. These primarily target the structural, angular, dimensional, and resolution aspects of images, thereby diversifying the dataset. Alternatively, they might introduce certain noise to the images to bolster their classification robustness. Deformable augmentation employs elastic or non-rigid transformations to produce deformed images. Within the realm of deep learning augmentation, a technique reminiscent of game theory, known as GAN-based augmentation, plays a pivotal role in image classification. It primarily consists of a generator network and a discriminator network, which, through adversarial training, produce realistic data samples. The underlying principle involves a competitive interplay between the generator and discriminator, where the discriminator evaluates the authenticity of images. This competitive training paradigm enhances the realism of the images produced by the generator. In the medical domain, there are some innovative applications and use cases. For instance, Pix2Pix GAN is a variant designed for image-to-image transformation. Without compromising the semantic structure and integrity, it facilitates domain transitions and can modify and supplement the original images. CycleGAN aims to learn the mapping between two distinct domains, ensuring that transformations from one domain to another are feasible. A key concept here is the introduction of cycle consistency loss, ensuring that post-transformation, the image bears resemblance to the original. For instance, Xu et al. proposed a semi-supervised attention-guided CycleGAN for enhancing MRI images for brain tumor classification [15]. Additionally, techniques like Deep Convolutional GAN (DCGAN), which introduces noise, convolutional layers, and deconvolutional layers, and the Auxiliary Classifier GAN (ACGAN) for image classification are noteworthy [4]. Researchers have employed these techniques to generate realistic 2D CT images for liver lesions. Radiologists only need to annotate the original real images and confirm the diagnostic results, while the synthesis and classification tasks are accomplished by the GAN.

From another perspective, GANs can also be categorized under Deep Domain Adaptation techniques. Domain adaptation is a crucial sub-technique of transfer learning that enhances image allocation. It primarily addresses scenarios where the target task is similar, but the data distribution varies, such as texts in different languages potentially leading to a decline in classification efficiency. There are numerous domain adaptation techniques tailored for different scenarios, including Importance Weighting, subspace Alignment, Single-Source DA, and the aforementioned Deep Domain Adaptation. Domain adaptation is extensively employed in medicine. For instance, for different modalities like X-rays and CTs, cross-modality adaptation techniques are commonly used. For medical segmentation tasks, image-level adaptation is required, while for classification tasks, feature-level adaptation is necessary. For medical images from different domains, multi-source domain adaptation might be employed. Li et al. [16] proposed a subspace alignment domain adaptation method for Alzheimer's disease classification. They first devised a strategy to align the features extracted from the source and target domains in a common subspace and then trained a discriminator. Yao et al. [17] introduced an innovative approach to address the challenges of cross-modality medical image segmentation. By leveraging multi-style image transformation techniques combined with 3D segmentation methods, they aimed to enhance segmentation accuracy in unlabeled target domains (like MRI) using labeled source domains (like CT). A key component of their method is the quad self-attention module, designed to effectively amplify the relationships between widely separated features in spatial regions, significantly improving segmentation accuracy. This innovative framework has proven especially effective in segmenting complex brain structures, including tumors and cochleae, showcasing its potential in advancing medical imaging practices.

## 5. Future and challenge

From the preceding discussions, it is unequivocally evident that medical imaging, characterized as a high-dimensional data type, confronts substantial challenges in acquiring adequate samples, especially given the stringent requirements for annotated data. Another significant impediment in the realm of medical imaging is the inherent variability among patients and the diversity of imaging protocols. This leads to the presence of heterogeneity in multi-center datasets. In light of these challenges, the prospects for transfer learning in medical imaging appear immensely promising. By synergizing with data augmentation techniques, transfer learning can furnish medical imaging with a sufficient volume of samples, thereby leveraging pre-trained models to address the high costs associated with sample acquisition. As a pivotal sub-branch of transfer learning, domain adaptation techniques offer solutions to the aforementioned heterogeneity issues. Furthermore, recent advancements in the research on transfer learning and Convolutional Neural Networks (CNNs) have witnessed a series of breakthroughs in terms of novel techniques and methodologies, suggesting immeasurable potential for their application in the medical domain.

However, the deployment of deep learning in medicine is not without its formidable challenges. Primarily, neural networks, often analogized as "black-box" models [18], suffer from a relative lack of interpretability. Yet, in the medical sphere, the rationale and processes underpinning a model's decisions are of paramount importance for diagnostic purposes [19]. Consequently, enhancing the interpretability of deep learning models and CNNs might emerge as a salient challenge in the foreseeable future. Additionally, issues related to data imbalance and the generalization capabilities of models still necessitate comprehensive solutions [20].

## 6. Conclusion

With the continual advancements in both the medical field and deep learning, transfer learning and neural networks have been progressively expanding their applicability within the medical domain. This monumental paradigm shift not only augments the accuracy of medical imaging classification but also furnishes clinicians with effortless, in-depth, and precise insights, thereby ensuring more accurate diagnostic outcomes for each patient.

As alluded to earlier, transfer learning, characterized as a technique with vast applicability in the medical sector, not only harbors boundless developmental potential but also grapples with significant challenges. To foster more efficacious innovations and implementations of artificial intelligence in medicine, a collaborative approach is imperative, encompassing medical professionals, computer science researchers, biotechnologists, and other domain-specific experts. In conclusion, it is anticipated that with the relentless propulsion of technology, the realm of medical imaging classification will witness further refinement.

## References

[1] Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik, 29(2), 102-127.

[2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[3] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[4] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 321, 321-331.

[5] Han, D., Liu, Q., & Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. Expert Systems with Applications, 95, 43-56.

[6] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5), 1285-1298.

[7] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big data, 3(1), 1-40.

[8] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.

[9] Niu, S., Liu, Y., Wang, J., & Song, H. (2020). A decade survey of transfer learning (2010–2020). IEEE Transactions on Artificial Intelligence, 1(2), 151-166.

[10] Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. BMC medical imaging, 22(1), 69.

[11] Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. Computers in biology and medicine, 111, 103345.

[12] Peng, L., Liang, H., Luo, G., Li, T., & Sun, J. (2022). Rethinking Transfer Learning for Medical Image Classification. medRxiv, 2022-11.

[13] Horry, M. J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., & Shukla, N. (2020). COVID-19 detection through transfer learning using multimodal imaging data. Ieee Access, 8, 149808-149824.

[14] Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology, 65(5), 545-563.

[15] Xu, Z., Qi, C., & Xu, G. (2019, November). Semi-supervised attention-guided cyclegan for data augmentation on medical images. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 563-568). IEEE.

[16] Li, W., Zhao, Y., Chen, X., Xiao, Y., & Qin, Y. (2018). Detecting Alzheimer's disease on small dataset: A knowledge transfer perspective. IEEE journal of biomedical and health informatics, 23(3), 1234-1242.

[17] Yao, K., Su, Z., Huang, K., Yang, X., Sun, J., Hussain, A., & Coenen, F. (2022). A novel 3D unsupervised domain adaptation framework for cross-modality medical image segmentation. IEEE Journal of Biomedical and Health Informatics, 26(10), 4976-4986.

[18] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-42.

[19] Banegas-Luna, A. J., Peña-García, J., Iftene, A., Guadagni, F., Ferroni, P., Scarpato, N., ... & Pérez-Sánchez, H. (2021). Towards the interpretability of machine learning predictions for medical applications targeting personalised therapies: A cancer case survey. International Journal of Molecular Sciences, 22(9), 4394.

[20] Hartono, A. P., Luhur, C. R., Indriyani, C. A., Wijaya, C. R., Qomariyah, N. N., & Purwita, A. A. (2021, August). Evaluating Deep Learning for CT Scan COVID-19 Automatic Detection. In 2021 International Conference on ICT for Smart Society (ICISS) (pp. 1-7). IEEE.