



Medical Image Classification using Transfer Learning Using Machine Learning

VISHVA TRIVEDI¹, SWAR BHAVSAR², KRISH PARIKH³, SAHIL PATEL⁴, PROF. JANKI TEJAS PATEL⁵

¹: 231133107073

²: 231133107009

³: 231133107039

⁴: 231133107051

SAL COLLEGE OF ENGINEERING,
DEPARTMENT OF ENGINEERING, AHMEDABAD, GUJARAT, INDIA
DEPARTMENT OF ENGINEERING

ABSTRACT :

Transfer learning (TL) has emerged as an essential technique for medical image classification, addressing key challenges such as limited labelled data, domain shift, and high annotation costs. This report reviews the principles of transfer learning, examines widely-used convolutional neural network architectures (e.g., VGG, ResNet, EfficientNet), and discusses methodology aspects including data preprocessing, augmentation, fine-tuning strategies, and evaluation metrics. Case studies and datasets commonly used in the field (e.g., CheXpert, ChestX-ray14, ISIC, HAM10000) are summarised. Empirical evidence and recent reviews show that TL — especially domain-adaptive pretraining and careful fine-tuning — consistently improves classification performance in clinical imaging tasks (Kim, 2022; Morid et al., 2021). The report concludes with limitations, ethical considerations, and future directions such as self-supervised pretraining, multimodal models, and federated learning.

Keywords; Transfer learning; Medical image classification; Convolutional Neural Networks; Fine-tuning; CheXpert; EfficientNet

1. Introduction

Medical imaging is central to modern diagnostics and treatment planning, spanning modalities such as radiography (X-ray), computed tomography (CT), magnetic resonance imaging (MRI), and dermatoscopic photographs. Automated image analysis using machine learning (ML) promises faster, more consistent, and scalable interpretation, but practical deployment faces two recurring challenges: limited labelled clinical data and distribution shifts between photographic datasets used for pretraining and clinical imaging data. Transfer learning (TL) — the practice of reusing model representations learned from one task/domain and adapting them to another — offers a pragmatic solution by leveraging large-scale pretraining on natural images (e.g., ImageNet) or domain-specific corpora (e.g., chest radiographs). This report provides an academic overview of TL techniques applied to medical image classification, integrating insights from recent reviews and empirical studies (Kim, 2022; Morid et al., 2021; Alzubaidi et al., 2021).

2. Literature Review

Early deep learning breakthroughs in computer vision depended on large labelled datasets such as ImageNet. Directly training deep convolutional neural networks (CNNs) from scratch for medical tasks is often infeasible due to smaller dataset sizes. Consequently, TL — particularly fine-tuning ImageNet-pretrained networks — became a widespread practice. Comprehensive reviews (Morid et al., 2021; Kim, 2022) synthesise the literature, noting that popular backbone models include VGG, ResNet, DenseNet, Inception, and, more recently, EfficientNet and vision transformers. Studies emphasise two important findings:

- (1) naive transfer from natural images sometimes underperforms domain-adaptive pretraining, and
- (2) The choice of fine-tuning strategy (linear probing, partial fine-tuning, full fine-tuning) substantially affects downstream performance (Rethinking Transfer Learning, 2023; Gu et al., 2024).

3. Why transfer learning?

- **Limited labelled data:** annotating medical images requires expert clinicians and is expensive.
- **Faster convergence and better generalisation:** pretrained models provide sensible initial feature detectors; fine-tuning requires fewer labelled

examples.

- **Computational efficiency:** avoids training very deep networks from scratch.
- **Feature reuse:** low-level filters (edges, textures) learned from natural images still provide useful representations for many medical images.

However, the domain gap between natural photos and medical images can limit direct transfer. Strategies such as modality-specific pretraining, multi-stage transfer, or self-supervised pretraining help bridge this gap.

4. Transfer learning strategies

4.1. Feature extraction (fixed backbone)

Freeze pretrained convolutional layers and train only a new classifier head. Useful when labelled data is scarce and the risk of overfitting is high.

4.2. Fine-tuning

Unfreeze some top layers or the entire network and train at a reduced learning rate. More flexible and often yields better performance when more labelled data is available.

4.3. Multi-stage transfer (progressive fine-tuning)

Start from an ImageNet-pretrained model, fine-tune on a large intermediate medical dataset (if available), then fine-tune on the target dataset. This reduces the domain gap.

4.4. Domain-specific pretraining

Pretrain on medical image datasets or use weights from models trained for similar medical tasks. Examples: pathology-pretrained encoders for histology tasks.

5. Popular model architectures

· **ResNet family (ResNet-18/34/50/101):** widely used baseline for transfer learning due to stability. ResNet models are widely used for transfer learning due to their residual connections, which prevent vanishing gradients in deep networks. They provide stable and accurate performance across many medical imaging tasks. Variants like ResNet-50 and ResNet-101 capture both low- and high-level features efficiently. These models are often used as strong baselines for classification. Their pretrained ImageNet weights make fine-tuning easier and more effective.

· **DenseNet:** popular for chest x-ray tasks because of parameter efficiency and gradient flow. DenseNet connects each layer to all previous layers, improving feature reuse and gradient flow. This design makes the model more efficient with fewer parameters. It performs exceptionally well in chest X-ray analysis and other diagnostic tasks. DenseNet captures fine details in medical images, aiding disease detection. Its compactness and accuracy make it popular in healthcare applications.

· **EfficientNet / EfficientNetV2:** strong performance per parameter; often used in competitions. EfficientNet scales depth, width, and resolution together for balanced accuracy and efficiency. It delivers strong performance with fewer parameters compared to traditional CNNs. The model is widely used in medical imaging, where computational resources are limited. EfficientNetV2 improves training speed and generalisation. These models often achieve top results in image classification competitions.

· **Vision Transformers (ViT) and hybrid CNN-ViT models:** gaining traction, often require more data or good pretraining. Vision Transformers treat images as sequences of patches and use attention mechanisms to learn global relationships. They can capture long-range dependencies that CNNs often miss. ViTs need large datasets or pretrained weights to perform well. Hybrid CNN-ViT models combine CNN's local feature extraction with ViT's global context understanding. They are gaining popularity in complex medical imaging tasks.

· **Specialised pathology encoders:** models trained on gigapixel data and patch-level encoders for WSIs. These models are designed for large Whole Slide Images (WSIs) used in pathology. They process image patches and learn meaningful representations for disease detection. Pathology encoders capture cellular and tissue-level details effectively. Many are pretrained on large pathology datasets for higher accuracy. They are crucial for diagnosing cancers and other microscopic abnormalities.

6. Detailed Explanation of Models

6.1 Convolutional Neural Networks (CNNs) CNNs extract hierarchical features via learned convolutional filters. Early architectures such as AlexNet and VGG demonstrated the value of deep stacks of convolutions. VGG networks use uniform small (3×3) kernels and deep layers; they are conceptually simple but parameter-heavy.

6.2 ResNet and Residual Learning ResNet introduced residual connections to mitigate the vanishing gradient problem, enabling much deeper networks and improved transferability of features.

6.3 DenseNet and Feature Reuse Dense Net connects layers densely, promoting feature reuse and efficient parameter usage; Dense Net variants (e.g.,

DenseNet-121) are common in chest X-ray tasks.

6.4 Efficient Net and Scaling Efficient Net uses compound model scaling (width, depth, resolution) guided by neural architecture search, achieving Strong ImageNet performance with fewer parameters. Multiple studies adapted Efficient Net to medical imaging tasks with competitive results (Ebenezer et al., 2021).

3.5 Vision Transformers (ViT) Transformers adapted to images use self-attention to model long-range dependencies; their utility increases with larger pretraining datasets. Domain-specific pretraining or hybrid CNN-Transformer models are emerging in medical imaging research.

6.6 Transfer Learning Strategies Common TL

Approaches include: feature extraction (freeze backbone, train classifier), linear probing (train a new head and optionally partial fine-tuning), and full fine-tuning (update all weights). The optimal approach depends on target dataset size and domain gap (Kim, 2022; Gu et al., 2024).

7. Preprocessing, augmentation, and handling medical images

1. Normalisation:

Normalisation ensures consistent intensity or colour ranges across medical images. For CT or MRI scans, intensity scaling is used to standardise pixel values between specific ranges (e.g., 0–1). In histopathology images, colour normalisation helps reduce stain variation across samples. This step improves model stability and convergence during training. Proper modality-specific normalisation enhances feature extraction and overall accuracy.

2. Resizing:

Medical images often vary greatly in size, especially Whole Slide Images (WSIs) that can be extremely large. Resizing helps fit these images into the model's fixed input size while preserving important details. Tile-based extraction is commonly used, where large images are divided into smaller patches. Careful sampling ensures that each tile represents meaningful medical regions. This approach maintains both computational efficiency and diagnostic relevance.

3. Data Augmentation:

Data augmentation increases dataset diversity and reduces overfitting by creating varied training samples. Common techniques include rotations, flips, elastic transformations, and random cropping. Advanced methods like cutout and mix-up help models generalise better. In medical imaging, domain-specific augmentations are used—for example, stain normalisation or colour jittering in histopathology. These augmentations simulate real-world variations, improving model robustness.

4. Class Imbalance Handling:

Medical datasets often suffer from class imbalance, where diseased samples are much fewer than normal ones. Techniques like oversampling minority classes and using class-weighted or focal loss functions help balance training. Synthetic data generation using GANs (Generative Adversarial Networks) can also increase minority samples. Handling imbalance ensures fair learning and prevents bias toward common classes. This leads to improved diagnostic performance and reliability.

8. Methodology

8.1 Datasets Medical imaging research uses modality-specific public datasets: CheXpert and ChestX- for chest radiographs; ISIC and HAM10000 for dermatoscopic skin lesion images; BraTS and TCIA collections for brain MRI; and various CT datasets for COVID-19 and pulmonary disease studies. CheXpert is notable for its size (224,316 radiographs) and label uncertainty modelling (Stanford ML Group, CheXpert).

8.2 Preprocessing and Augmentation Preprocessing steps typically include resizing, intensity normalisation (e.g., pixel scaling to [0,1] or z-score), and modality-aware transforms (windowing for CT). Data augmentation (rotation, flipping, colour jitter, elastic transforms) reduces overfitting and simulates realistic variabilities. For dermatoscopic images, colour-space augmentation and hair-removal preprocessing are helpful.

8.3 Transfer Learning Workflow A typical TL pipeline: select pretrained backbone (ImageNet or domain-pretrained), replace the classifier head with task-specific layers, choose an optimisation strategy (Adam/SGD with weight decay), and select a fine-tuning schedule. Linear probing for several epochs followed by gradual unfreezing often stabilises learning. Cross-validation and test-splits must respect patient-level separation to avoid information leakage.

8.4 Evaluation Metrics Clinical tasks use accuracy, precision, recall (sensitivity), specificity, F1-score, area under ROC (AUC), and calibration metrics. For multi-label chest X-ray tasks, per-label AUC and micro/macro averages are reported.

Clinical utility also depends on positive predictive value at operating points relevant to disease prevalence.

9. Results and Discussion

Numerous empirical studies show that TL improves sample efficiency and performance compared to training from scratch, especially for small datasets (Morid et al., 2021; Kim, 2022). Domain-adaptive pretraining — i.e., pretraining on large in-domain collections (e.g., chest radiographs) before fine-tuning — often outperforms vanilla ImageNet transfer (Vendrow et al., 2022). Efficient Net and ResNet variants remain competitive backbones; however, the optimal architecture depends on image resolution, task complexity, and available compute. Fine-tuning choices matter: linear probing. Followed by selective unfreezing tends to be robust. Recent work (Gu et al., 2024) highlights that mixing supervised pretraining with self-supervised objectives or using multistage TL further improves downstream classification. Careful attention to data splits, augmentation pipelines, and reporting per-class performance is crucial to avoid overoptimistic claims. Limitations observed across the literature include dataset bias, label noise (report-derived labels), and domain shifts between institutions. Reproducibility remains a concern; authors should publish code, weights, and exact preprocessing. Finally, external validation on geographically diverse cohorts is necessary for Clinical translation.

10. Conclusion and Future Scope

Transfer learning has become a foundational strategy for medical image classification, offering practical benefits in data efficiency and performance. Empirical and review studies (Kim, 2022; Morid et al., 2021; Alzubaidi et al., 2021) indicate that domain-aware pretraining, appropriate fine-tuning schedules, and rigorous evaluation practices are key to achieving clinically useful models. Future research directions include:

- Self-supervised and contrastive pretraining on large unlabeled medical datasets to reduce dependency on curated labels.
- Federated and privacy-preserving TL to leverage multi-institutional data without centralising sensitive records.
- Multimodal models that combine imaging with clinical text and structured EHR data for improved diagnostic context.
- Explainability and uncertainty quantification to make model outputs actionable for clinicians.

With careful methodological rigour and ethical safeguards, TL-driven medical image classification can augment clinical workflows and improve patient outcomes.

REFERENCES

1. Kim, H. E., et al. (2022). Transfer learning for medical image classification: A literature review. *BMC Medical Imaging*, 22, Article 93. <https://doi.org/10.1186/s12880-022-00793-7>
2. Morid, M. A., et al. (2021). A scoping review of transfer learning research on medical imaging. *Artificial Intelligence in Medicine*, 117, 102077.
3. Alzubaidi, L., et al. (2021). Novel transfer learning approach for medical imaging with limited labelled data. *Journal of Medical Systems*, 45, 97.
4. Gu, C., et al. (2024). Deep transfer learning using real-world image features for medical image classification. *Scientific Reports*, 14, 1234.
5. Ebenezer, A. S., et al. (2021). Effect of image transformation on EfficientNet model for COVID-19 detection using CT images. *Applied Sciences*, 11(22), 1-20.
6. Stanford ML Group. (2019). CheXpert: A large chest radiograph dataset. <https://stanfordmlgroup.github.io/competitions/chexpert/>
7. Salehi, A. W., et al. (2023). A study of CNN and transfer learning in medical imaging. *Sustainability*, 15(7), 5930.
8. Vendrow, E., et al. (2022). Understanding transfer learning for chest radiograph tasks. *arXiv:2205.02841*.
9. Rethinking Transfer Learning for Medical Image Classification. (2023). *medRxiv*. <https://www.medrxiv.org/content/10.1101/2022.11.26.22282782v2>