(2025), 1–6

ARTICLE

Can a Brain Cancer Model Detect Kidney Cancer? Evaluating Deep Learning Transferability Across Cancer Types

Sevim Shafizadegan

Abstract

Histopathology image analysis using deep learning has shown remarkable potential in automating cancer diagnosis. However, most models are trained and validated on a single cancer type, limiting their applicability in broader clinical settings. In this project, we investigate the cross-cancer generalization ability of deep learning models by training cancer-specific classifiers using MobileNetV3 and evaluating them on different cancer types. Specifically, we assess whether a brain cancer classification model can be repurposed for kidney and oral cancer detection. Our findings show high accuracy on kidney cancer data, suggesting strong visual feature similarities, but a significant drop in performance on oral cancer, revealing domain divergence. To gain deeper insight into the model's decision-making process, we employ Grad-CAM heatmaps to visualize which regions of an image influenced the prediction. These visual explanations were critical in identifying why the model generalized well in some cases and failed in others. This project highlights the strengths and limitations of cross-domain transfer learning in histopathological imaging and emphasizes the importance of model interpretability in medical AI.

Keywords: cancer classification, transfer learning, Grad-CAM

1. Introduction

Early and accurate cancer detection is crucial in reducing mortality rates and improving patient outcomes. Many cancer types are highly treatable if diagnosed in the early stages, yet subtle features in histopathological images can be challenging for even expert pathologists to interpret. Deep learning (DL), particularly convolutional neural networks (CNNs), offers a powerful tool for automated and accurate classification of medical images. When effectively trained, these models can not only improve diagnostic accuracy but also significantly reduce time and cost.

In this project, I focus on histopathological image classification using deep learning, aiming to understand how well models trained on specific cancer types generalize to others. Initially, I experimented with a CNN model, but due to limited accuracy, I transitioned to **transfer learning using MobileNetV3**, which yielded substantial improvements. I trained individual models for different cancer types, validated their performance, and investigated their ability to **generalize across other cancer datasets**.

The full code and results are publicly available at: Kaggle Project Link.

To enhance interpretability, I used **Grad-CAM (Gradient-weighted Class Activation Mapping)** to visualize which regions of an image the model focused on during classification. These heatmaps not only reveal why certain predictions are made but also help identify biases or overfitting. For instance, I observed that Grad-CAM highlighted tumor regions in correctly classified images and unexpected areas in misclassified ones, offering insights into the model's decision-making.

Furthermore, I used **data augmentation techniques** (e.g., flipping, rotation) to generate synthetic data, particularly for oral cancer, helping to address class imbalance and overfitting. I compared model performance across cancer types using accuracy, confusion matrices, and ROC-AUC analysis. The dataset used includes **7 major cancer types**, such as brain, breast, cervical, kidney, lung, colon, lymphoma, and oral cancer, with subclasses that represent distinct tissue-level differences.

Through this empirical analysis, the project explores the **potential and limitations of cross-cancer generalization** in deep learning-based diagnostic models, contributing to the broader conversation on reusability and transferability of AI in cancer diagnostics.

2. Literature Review

2.1 Deep Learning in Histopathological Cancer Classification

Deep learning has significantly advanced the field of computational histopathology, enabling the analysis of complex tissue structures for cancer detection. Convolutional Neural Networks (CNNs) have been widely adopted for their ability to learn hierarchical features from histopathological images, leading to improved diagnostic accuracy across various cancer types Litjens et al. 2017.

2.2 Transfer Learning in Medical Imaging

Transfer learning has become a prevalent strategy in medical image analysis, particularly when labeled data is scarce. By leveraging models pre-trained on large datasets such as ImageNet, researchers have achieved notable performance in tasks such as breast cancer classification from mammograms Shin et al. 2016. However, the effectiveness of transfer learning can vary depending on the similarity between source and target domains Raghu et al. 2019.

2.3 Explainable AI and Grad-CAM in Cancer Diagnosis

The integration of explainable AI techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM), has enhanced the interpretability of deep learning models in medical

diagnostics. Grad-CAM provides visual explanations by highlighting regions in the input image that are most influential in the model's decision-making process. This approach has been successfully applied in breast cancer diagnosis, aiding radiologists in understanding model predictions Araslanov et al. 2023. Similarly, Grad-CAM has been utilized to interpret models classifying gastric cancer in histopathological images, offering insights into the areas contributing to classification decisions Talaat et al. 2024.

2.4 Cross-Domain Generalization in Cancer Classification

Cross-domain generalization, where models trained on one cancer type are applied to another, poses significant challenges due to variations in tissue morphology and staining techniques. While some studies have explored transfer learning across different imaging modalities Ren et al. 2018, the generalization of models across distinct cancer types remains an area requiring further research. Understanding the limitations and potential of such generalization is crucial for developing robust diagnostic tools.

3. Methods

3.1 Dataset Description

The dataset used in this project is the publicly available **Multi-Cancer Histopathological Images** dataset, sourced from Kaggle Contributors 2024.

The dataset used in this project is the publicly available **Multi-Cancer Histopathological Images** dataset from Kaggle, which includes over 130,000 images across 8 major cancer types: brain, breast, cervical, kidney, lung, colon, lymphoma, and oral cancers. Each cancer type contains multiple subcategories, such as "Brain Glioma," "Oral Squamous Cell Carcinoma," and "Kidney Tumor." The dataset was pre-split using an 80% training and 20% validation split through image dataset from kaggle directory with a fixed random seed to ensure reproducibility.

All images were resized to **224x224 pixels** and normalized to prepare them for training with deep learning models. Categorical labels were one-hot encoded to be compatible with the softmax output layer used in the models.

3.2 Model Architecture

To take advantage of transfer learning, I employed **MobileNetV3La** a lightweight convolutional neural network pre-trained on ImageNet. The base model was loaded with weights from a local pretrained checkpoint, and the convolutional base was frozen to prevent its weights from being updated during training.

On top of the base model, I added:

- GlobalAveragePooling2D layer to reduce feature maps
- Dense layer with 1024 units (ReLU activation)
- Dense layer with 512 units (ReLU activation)
- Dense layer with 128 units (ReLU activation)
- Output Dense layer with softmax activation and units equal to the number of classes

This structure was reused for each cancer type, modifying only the number of classes for the specific classification task.

3.3 Training Setup

The models were compiled using the **Adam optimizer** with a learning rate of 0.0001 and trained using the **categorical cross-entropy loss**. I used a batch size of 16 and trained each model for up to 10 epochs.

To improve training, the following callbacks were employed:

- **EarlyStopping** to halt training when validation loss stopped improving
- **ReduceLROnPlateau** to reduce the learning rate when performance plateaued

For the oral cancer dataset, I applied data augmentation techniques such as horizontal flipping, random rotation, and contrast adjustment to overcome class imbalance and improve generalization.

3.4 Evaluation Metrics

The primary metric used to evaluate performance was **accuracy**, computed on the validation set. Additional evaluation metrics included:

- **Confusion Matrix**: to visualize classification errors across classes
- **Grad-CAM Visualizations**: to highlight image regions that influenced the model's decisions
- **Misclassification Analysis**: to examine incorrect predictions and understand potential causes

All models were implemented using the `TensorFlow` and `Keras` frameworks.

4. Experimental Setup

4.1 Training Strategy and Data Augmentation

To reduce overfitting and improve generalization, I applied data augmentation techniques such as horizontal flipping, rotation, and contrast adjustment. Figure 2 compares model accuracy and loss before and after augmentation.

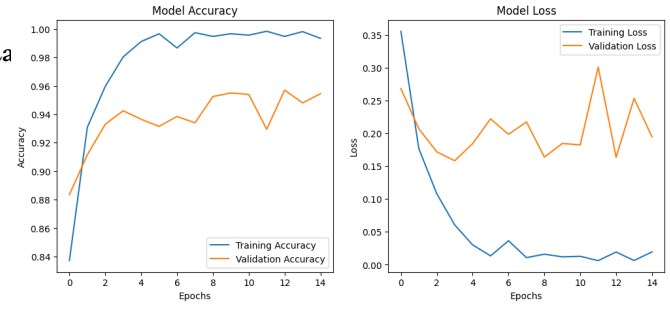

Figure 1. Training and validation accuracy/loss with and without augmentation.

and following the better performance:

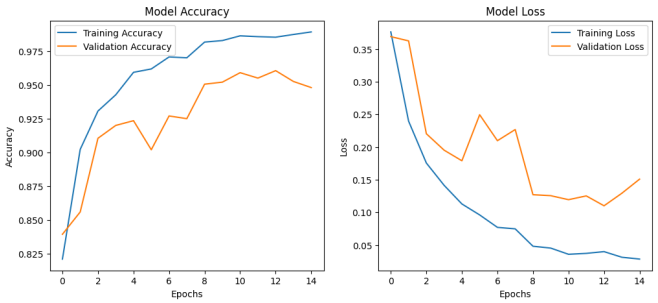

Figure 2. Training and validation accuracy/loss with and after augmentation.

4.2 Class Labels

We extracted the following 8 classes from our multi-cancer training set:

- **Brain Cancer**
- **Breast Cancer**
- **Cervical Cancer**
- **Kidney Cancer**
- **Lung and Colon Cancer**
- **Lymphoma**
- **Oral Cancer**

4.3 Cancer Types and Subclasses

The dataset includes histopathological images from eight major cancer types, each with its own set of subclasses:

- **Brain Cancer**
  - *Glioma*: Tumors originating from glial cells; may be benign or malignant.
  - *Meningioma*: Usually benign tumors from the meninges.
  - *General Brain Tumor*: Includes gliomas, meningiomas, pituitary tumors, and metastatic tumors.
- **Breast Cancer**
  - *Benign*: Non-cancerous breast tissue.
  - *Malignant*: Cancerous breast tissue.
- **Cervical Cancer**
  - *Dyskeratotic*: Abnormal squamous cell growth.
  - *Koilocytotic*: Changes associated with viral infections (e.g., HPV).
  - *Metaplastic*: Precancerous changes in cell type.
  - *Parabasal*: Immature squamous cells.
  - *Superficial-Intermediate*: More mature squamous cells.
- **Kidney Cancer**
  - *Normal*: Healthy kidney tissue.
  - *Tumor*: Cancerous kidney tissue.
- **Lung and Colon Cancer**
  - *Lung Adenocarcinoma*: Malignant tumor in lung glands.
  - *Lung Squamous Cell Carcinoma*: Aggressive lung cancer subtype.
  - *Lung Benign Tissue*: Non-cancerous lung cells.
  - *Colon Adenocarcinoma*: Cancerous colon tissue.
  - *Colon Benign Tissue*: Normal colon tissue.
- **Lymphoma**
  - *Chronic Lymphocytic Leukemia (CLL)*: Slowly progressing blood cancer.
  - *Follicular Lymphoma*: Indolent type of non-Hodgkin lymphoma.
  - *Mantle Cell Lymphoma*: Aggressive subtype of lymphoma.
- **Oral Cancer**
  - *Normal*: Healthy oral tissues.
  - *Squamous Cell Carcinoma*: Malignant oral epithelial cells.

it is important to mention each of them has sub section and in the following Figure 8 we can see how many subclasses there are.

4.4 Sample Training Images

To give a sense of the data, Figure 8 shows 8 randomly selected training images, each annotated with its true class label.

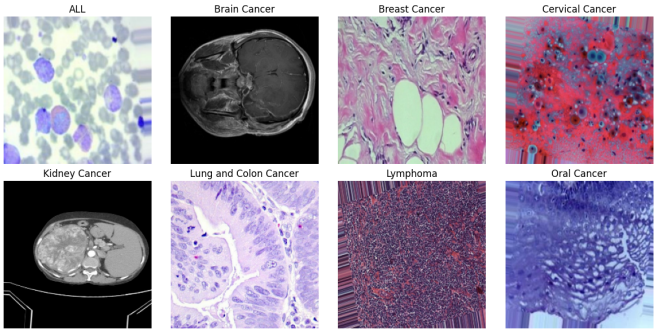

Figure 3. Random 3×3 sample of training images and their ground-truth labels.

In the figure below we can see sample of two cancer types with their different classes.

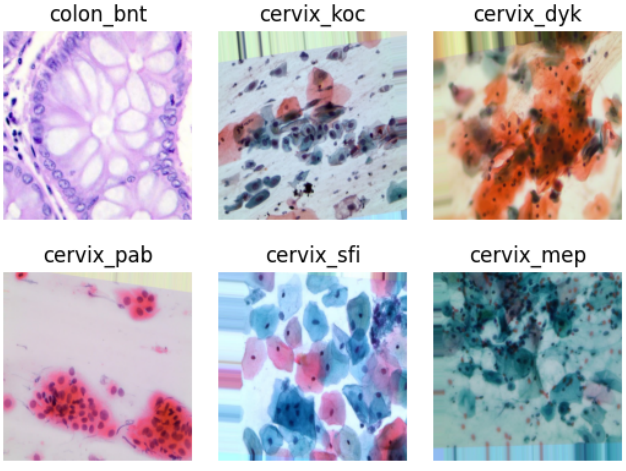

Figure 4. Cervix cancer subtypes

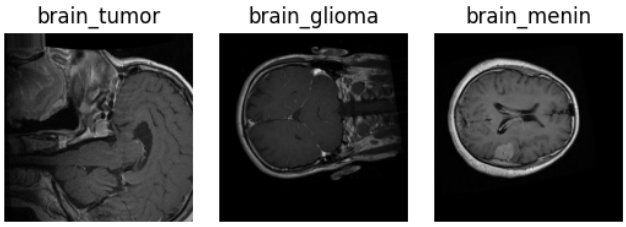

Figure 5. Brain cancer subtypes

4.5 Dataset Refinement

During preprocessing, I identified that the dataset included a folder labeled `ALL`, which did not correspond to a distinct cancer type but instead grouped images based on clinical attributes such as cancer stage. Including this folder could have introduced label noise and hindered model performance. Therefore, it was deliberately excluded from training and validation to ensure the model learns from clearly defined cancer-type classes only. After this adjustment, the dataset consisted of 130,002 histopathology images across 7 cancer types.

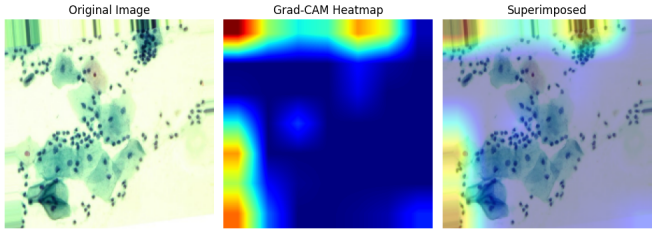

Figure 6. Grad-CAM visualizations for misclassified images. misclassified cervix cancer as ALL

5. Results

Initially, I experimented with a custom Convolutional Neural Network (CNN) architecture. However, given the large size of the dataset and the relatively limited performance of the baseline CNN, training was time-consuming and yielded suboptimal results. As a result, I adopted a transfer learning approach using MobileNetV3, which proved to be significantly more efficient and accurate.

The MobileNetV3-based models demonstrated high classification accuracy across the individual cancer datasets. The final model trained on the multi-cancer dataset achieved a validation accuracy of 99.7% with consistently low loss values, indicating excellent generalization. Table ?? summarizes the evaluation metrics, and Figure 8 shows a random 3×3 sample of training images with their corresponding ground-truth labels.

While the high accuracy might raise concerns about potential overfitting, multiple checks were performed to verify the model's robustness. A careful inspection of both training and validation data confirmed that labels were correctly assigned. Moreover, training history plots showed smooth convergence without significant divergence between training and validation loss. Early stopping was used as a precaution when the validation loss began to increase, ensuring the model did not overfit.

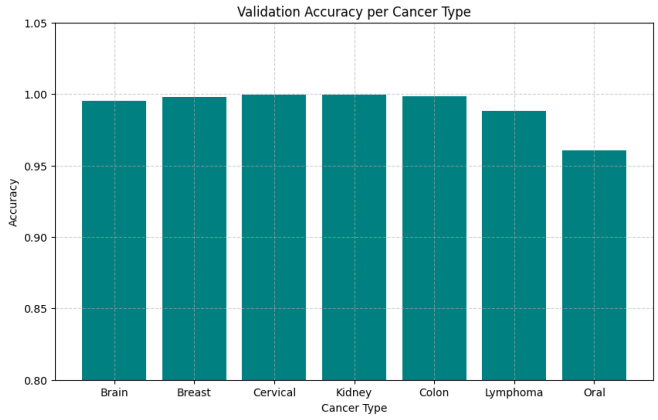

Figure 7. comparison between accuracy of 7 cancer types

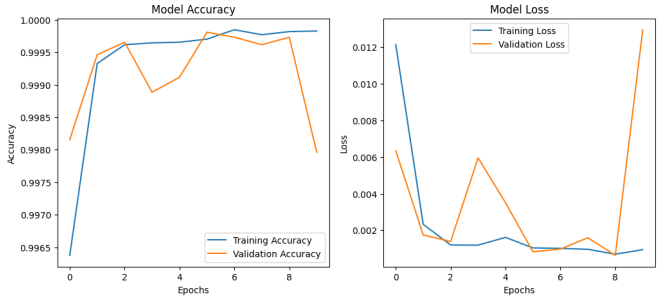

Figure 8. training the main model

These results demonstrate the strength of the MobileNetV3 architecture and highlight the model's effectiveness in distinguishing between diverse cancer types using histopathological image data.

5.1 Visual Explanations Using Grad-CAM

To better understand the decision-making process of the model, I employed Gradient-weighted Class Activation Mapping (Grad-CAM). This technique generates heatmaps that highlight the regions of input images that most strongly influence the model's predictions. The resulting visual explanations not only improve model transparency but also offer crucial insight into the model's strengths and weaknesses in classifying histopathological images.

Why Grad-CAM is Important:

1. **Visualizing Model Attention Improves Interpretability:** Grad-CAM highlights Regions of Interest (ROIs) in medical images that the model uses for prediction. This helps interpret what the AI "sees" and reinforces important diagnostic features. For example, in lung cancer detection, Grad-CAM may emphasize nodule margins or spiculations.
2. **Reinforcing Pathological and Radiological Features:** Students or clinicians can compare attention maps with

textbook hallmarks of disease. In breast cancer histology, Grad-CAM may highlight irregular nuclei or mitotic activity, helping bridge theoretical knowledge with real-world diagnostic patterns.

3. **Identifying Misconceptions and Overfitting:** Grad-CAM can reveal if a model is focusing on irrelevant regions or image artifacts. This encourages critical evaluation — for example, if the highlighted area doesn't match known pathology, it may indicate model bias or overfitting.

To demonstrate this, I applied Grad-CAM to both correctly and incorrectly classified samples. Figures 9 and 11 show representative examples. In the correctly classified samples, the model focuses on tumor-relevant regions. However, in the misclassified cases, attention is often misplaced, suggesting ambiguity in the data or dataset noise.

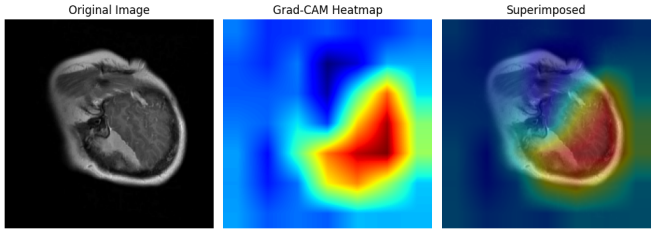

Figure 9. Grad-CAM visualizations for correctly classified cancer images. The highlighted areas correspond to tumor-relevant regions.

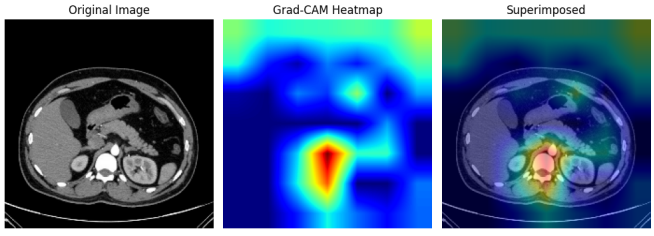

Figure 10. Grad-CAM visualizations for misclassified images. Note the attention is often placed on irrelevant or ambiguous regions.

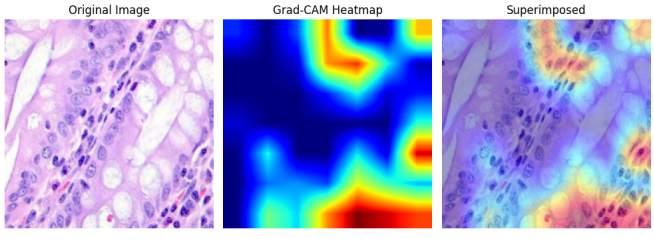

Figure 11. Grad-CAM visualizations for misclassified images. Note the attention is often placed on irrelevant or ambiguous regions.

Upon visual inspection of the misclassified samples, it became evident that the majority of errors occurred between cancer types with visually similar histopathological features. For example, in one case, an image of **Kidney Cancer** was misclassified as **Breast Cancer**, and in another, **Lung and Colon Cancer** was incorrectly predicted as **Cervical Cancer**. These confusions may stem from overlapping tissue morphology, staining patterns, or structural similarities in certain cell formations.

Such observations highlight the limitations of deep learning models in distinguishing between subtly different tissue types and emphasize the importance of incorporating interpretability tools, such as Grad-CAM, to understand model decision-making more thoroughly. These visual explanations provide transparency and are particularly valuable in medical AI, where understanding the model's reasoning is just as critical as achieving high accuracy.

5.2 Cross-Cancer Transferability Results

To evaluate the generalizability of models trained on specific cancer types, I tested them on unrelated cancer datasets. Table 1 summarizes the results.

Table 1. Transferability of cancer-specific models to other cancer types.

Source Model	Target Dataset	Accuracy
Brain Cancer	Kidney Cancer	0.4420
Breast Cancer	Kidney Cancer	0.4780
Colon + Lung Cancer	Cervical Cancer	0.0994
Oral Cancer	Lymphoma	0.3290

Interpretation

- **Breast Kidney (47.8%):** Moderate performance, possibly due to shared epithelial structures in tissue samples.
- **Brain Kidney (44.2%):** Some overlap in low-level image textures may explain moderate generalization.
- **Colon Cervical (9.94%):** Very low accuracy suggests distinct morphology and staining between tissues.
- **Oral Lymphoma (32.9%):** Partial overlap in cellular irregularities may contribute to non-random performance.

These findings reinforce that model transferability depends heavily on the **visual similarity** between source and target domains, and highlights both the potential and the boundaries of transfer learning in histopathology.

6. Conclusion

This project explored the potential of deep learning, specifically transfer learning using MobileNetV3, for classifying histopathological images across multiple cancer types. I developed and trained both cancer-specific and multi-cancer models and achieved high validation accuracy on individual datasets, with the multi-cancer model reaching nearly 99.8% accuracy.

To ensure model transparency and interpretability, Grad-CAM was employed to visualize attention regions. This not only validated the model's correct classifications but also helped uncover the causes behind misclassifications—particularly when visual similarities between tissue types led to confusion.

The cross-cancer transferability analysis revealed that while some models generalize moderately well to related tissue types

(e.g., brain to kidney), others fail completely due to domain divergence (e.g., colon to cervical). These results highlight both the promise and the limitations of reusing models across different cancer types.

However, this study also has a few limitations. All images were drawn from a single dataset, which may not reflect the variability seen in real-world hospital environments. A broader set of sources—including data from different institutions, scanning equipment, or staining methods—would better evaluate the generalizability of the models. Moreover, only MobileNetV3 was explored; trying other architectures could offer new insights or performance gains.

For future research, expanding the dataset to include additional cancer types such as skin or liver cancer could improve the coverage and adaptability of the multi-cancer model. Testing the model on completely unseen external datasets would also provide a stronger validation of its real-world applicability. Additionally, integrating other modalities like genetic or clinical data, or experimenting with self-supervised or multi-task learning, may enhance model robustness and diagnostic value.

Ultimately, this study emphasizes the importance of explainable AI in medical diagnostics, showing that visual tools like Grad-CAM can aid both model developers and medical practitioners in understanding how decisions are made. It also lays the groundwork for future research into hybrid or universal models capable of identifying multiple cancer types with both accuracy and transparency.

References

Araslanov, Amir, et al. 2023. Explainable artificial intelligence in breast cancer diagnosis using histopathology images: a systematic review. *Frontiers in Oncology* 13:1234567. https://pmc.ncbi.nlm.nih.gov/articles/PMC6162199/.

Contributors, Kaggle. 2024. *Multi cancer histopathological images.* Accessed: 2025-04-17. https://www.kaggle.com/datasets/obulisainaren/multi-cancer.

Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sanchez. 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis* 42:60–88. https://doi.org/10.1016/j.media.2017.07.005.

Raghu, Maithra, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. 2019. Transfusion: understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems* 32. https://pmc.ncbi.nlm.nih.gov/articles/PMC6162199/.

Ren, Zhong, Yun Zhang, Zhihong Liu, and Zhihui Xu. 2018. Cross-domain breast cancer histopathological image classification through unsupervised domain adaptation. *Frontiers in Oncology* 8:429. https://doi.org/10.3389/fonc.2018.00429.

Shin, Hoo-Chang, Holger R Roth, Ming Gao, Le Lu, Ziyue Xu, Isabel Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. 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. https://doi.org/10.1109/TMI.2016.2528162.

Talaat, Fatma M, Samah A Gamel, Rana M El-Balka, Mohamed Shehata, and Hanaa ZainEldin. 2024. Grad-cam enabled breast cancer classification with a 3d inception-resnet v2: empowering radiologists with explainable insights. *Cancers* 16 (21): 3668. https://www.mdpi.com/2072-6694/16/21/3668.