Transfer Learning for Image Classification: Exploring Neural and Attention Networks using Pre-Trained Models

Link to article: IMPact | Transfer Learning for Image Classification | Web
Digital Object Identifier (DOI): https://doi.org/10.82406/y1r0-a693

Celia Aspin

Year 4 MComp Undergraduate Student Researcher
School of Engineering and Physical Sciences
University of Lincoln

Abstract

An accurate diagnosis is a critical challenge in oncology. Errors remain an issue, as inaccuracies contribute to avoidable harm, increased costs and delays in diagnoses. Over the past decade, several deep learning solutions have offered promising solutions in improving cancer diagnosis with some having transitioned from research to clinical deployment. This project explored the effectiveness of transfer learning using two modern architectures: ResNet50, a Convolutional Neural Network, and DeiT, a transformer-based attention model. The project was conducted in two stages: First a binary classification was performed using four types of cancer to establish a baseline, Second, a multi-class classification was performed using one of the previous cancer types to test greater recognition and generalisation. Both models are initially trained on ImageNet, fine-tuned on the dataset and evaluated using accuracy and loss, f1-score and confusion matrices. Results show that both models performed comparably across all tasks, only experiencing minor fluctuations typical of training variance. Despite architectural differences, both CNN and transformer-based models are capable of producing reliable results in cancer image classification.

Keywords: Cancer, Neural Networks, Attention Networks, Deep Learning, Image Classification

Introduction

This Project was financed and supported by the Undergraduate Research Opportunity Scheme (UROS). A competitive bursary scheme which allows collaborative research between students and academics.

Early and accurate cancer diagnosis is crucial as it allows for more effective, less invasive treatments that significantly increase survival rates and improve quality of life. Deep learning has proved to be a powerful tool for this task, holding a remarkable performance in pattern recognition as well as anomalies in diverse datasets (Kumar et al., 2024). ResNet50, a Convolution Neural Network (CNN) has been widely used in medical image analysis showing a strong performance by extracting features in a hierarchical structure (Xu, Fu and Zhu, 2023). Meanwhile, attention-based architectures like Data-efficient image Transformers (DeiT) analyse images as patches and use attention mechanisms to relate information covering the entire image (He et al., 2023). This project investigates whether these fundamentally different deep learning architectures perform comparably in the task of cancer image classification, and how effectively each approach supports accurate cancer diagnosis when applied to histology and MRI images.

Project Background

Cancer is one of the leading causes of death worldwide, and having accurate interpretations of histology images is essential for treatment. Traditionally, the process has relied on the expertise of pathologists, but manual analysis is time-consuming, prone to variability and challenging when identifying similar tissue types. In recent years, deep learning has emerged as a powerful tool to support medical image classification, offering improvements to speed and consistency.

Convolution Neural Networks have been the architecture used most significantly in classification tasks for the past decade, with models like ResNet50 achieving high scores across multiple disciplines. CNNs specialise in capturing local features, such as textures and edges, combining higher-level patterns needed in classification. However, since CNNs are a hierarchical structure, it makes long-distance relationships within images hard to capture, which maybe important in a medical context.

Transformer-based architectures are a recent adaptation of natural language processing to computer vision. The Data-efficient Image Transformer processes images by splitting the image into patches and applying self-attention. This is to learn the relationship between all the patches simultaneously, letting global context be considered throughout training (Touvron et al., 2021). This shift in architecture has raised interested in transformers, and whether they are superior to CNNs with special focus on the medical field where fine distinctions are required.

This project investigates the application of both ResNet50 and DeiT to cancer image classification, comparing their performance in binary and multi-class tasks. The study aims to evaluate their strengths, limitations and suitability for image analysis.

Literature Review

Convolutional Neural Networks (CNNs) have seen a rapid rise in the field of medical imaging, presenting significant opportunities for conserving healthcare resources and decreasing wait time while maintaining a high level of accuracy (Jia et al., 2024). This is due to the CNNs hierarchical feature extraction and ability to capture local image patterns, allowing for high accuracy rates in detection and classification. ResNet50 has been widely applied in cancer detection tasks, demonstrating a strong performance on multiple types of tissue across datasets (Xu, Fu and Zhu, 2023).

Transformers were originally developed for natural language processing, recently having been adapted for computer vision purposes. Vision Transformers (ViTs) and Data-efficient Image Transformer (DeiT) employ self-attention mechanisms to analyse the relationship across an entire image by splitting it into patches. This allows transformers to capture local and global dependencies, potentially improving classification where the spatial pattern is relevant. Recent studies have applied transformers to medical imaging, showing a comparable or superior performance to CNNs in tasks such as tumour classification and segmentation (Touvron et al., 2021).

Approaches combining CNNs and transformers into a hybrid model has also been explored, leveraging CNNs local feature extraction with transformers' global modelling. These studies indicate that while transformers require more data to achieve an optimal performance, utilising transfer learning from large pre-trained models can mitigate the limitations (Djoumessi, Mensah and Berens, 2025).

Overall, literature suggests that both CNNs and transformers are viable architectures for medical image classification. However, many existing studies are conducted on limited or highly curated datasets, focus on a single imaging modality or lack direct comparisons across classification. As a result, comparative analyses of CNN and transformer-based models applied to cancer histology and MRI scans remain limited. This project builds on these findings by evaluating the performance of ResNet50 and DeiT in both binary and multi-class classification for cancer identification, providing insight into their practical applications and relative advantages.

Methodology

This project evaluated the performance of two machine learning architectures, ResNet50 and Data-efficient Image Transformer (DeiT), for a cancer image classification task using transfer learning. Both models were pretrained on ImageNet, a large, publicly available database and fine-tuned on histology and MRI images.

The dataset used in this project was acquired from Kaggle, a source of free, publicly available datasets (Naren 2024).

The dataset consisted of images from four types of cancer for binary classification for whether an image was malignant or benign. The classes acute lymphoblastic leukaemia, breast and cervical were all histology images whereas the kidney images were MRI scans. All the images underwent preprocessing, including resizing to match input dimensions required by the models (224 x 224 pixels), normalisation and augmentation in the form of random rotations, flips and colour jittering to improve generalisation. Following the completion of the binary classification task, a multi-class classification task was performed on the acute lymphoblastic leukaemia, for either benign, early or advanced stages. The dataset was split into training, validation and testing to enable model evaluation on unseen data.

Both models had the classification head replaced with a custom fully connected network with ReLU activations and dropout for regularization. Training was performed using the Adam optimiser and a ReduceLROnPlateu scheduler. Label smooth cross-entropy loss was used to improve robustness against overconfidence. Additionally, early stopping was applied with a patience of three epochs.

The model’s performance was evaluated using accuracy, precision, recall, f1-score and confusion matrices. Sample predictions were also visualised to provide a qualitative insight into the model’s outputs.

All tasks were implemented using PyTorch, with ResNet50 from TorchVision and DeiT from the timm library.

Results

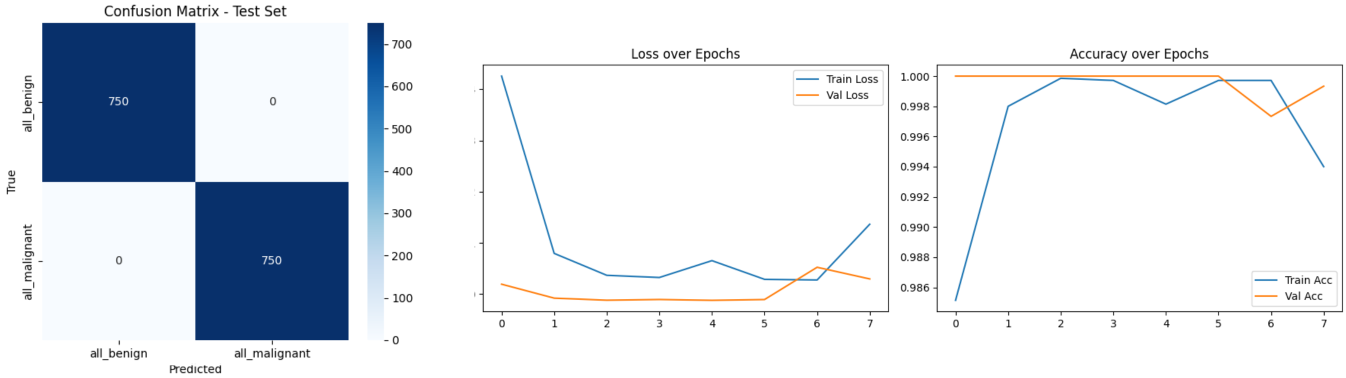
The performance of ResNet50 and DeiT were evaluated on two tasks: binary classification across 4 cancer types and multi-class classification with a single cancer type. Model performance was assessed using accuracy, precision, recall, f1-score and confusion matrices. Training and validation learning cures were also analysed to compare convergence and generalization.

Binary Classification Results

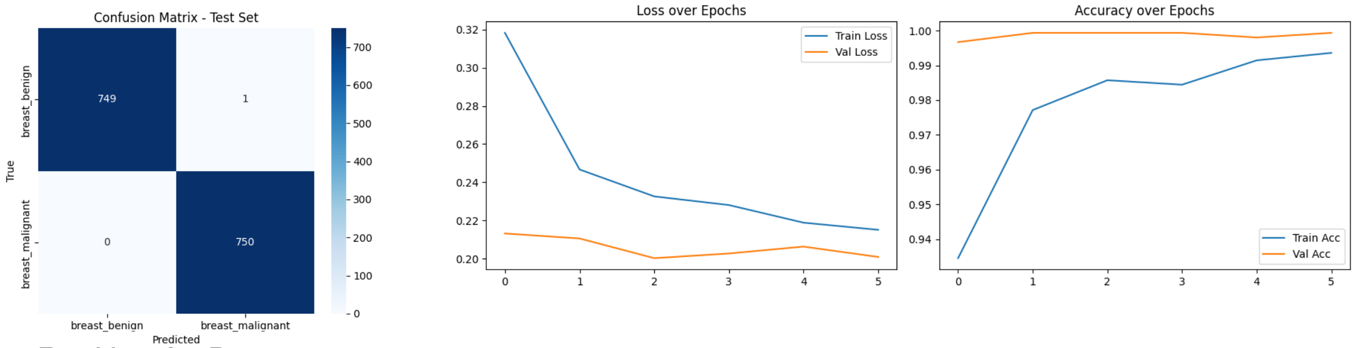
Model	Cancer Type	Accuracy	Precision (Malignant)	Recall (Malignant)	F1-Score (Malignant)	Precision (Benign	Recall (Benign)	F1-Score (Benign
ResNet50	ALL	100%	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
ResNet50	Breast	99.93 %	0.9987	1.0000	0.9993	1.0000	0.9987	0.9993

Model	Cancer Type	Accuracy	Precision (Malignant)	Recall (Malignant)	F1-Score (Malignant)	Precision (Benign	Recall (Benign)	F1-Score (Benign
ResNet50	Cervical	100%	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
ResNet50	Kidney	99.93%	1.0000	0.9987	0.9993	0.9987	1.0000	0.9993
DeiT	ALL	100%	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
DeiT	Breast	99.87%	0.9987	0.9987	0.9987	0.9987	0.9987	0.9987
DeiT	Cervical	100%	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
DeiT	Kidney	99.40%	0.9987	0.9893	0.9940	0.9894	0.9987	0.9940

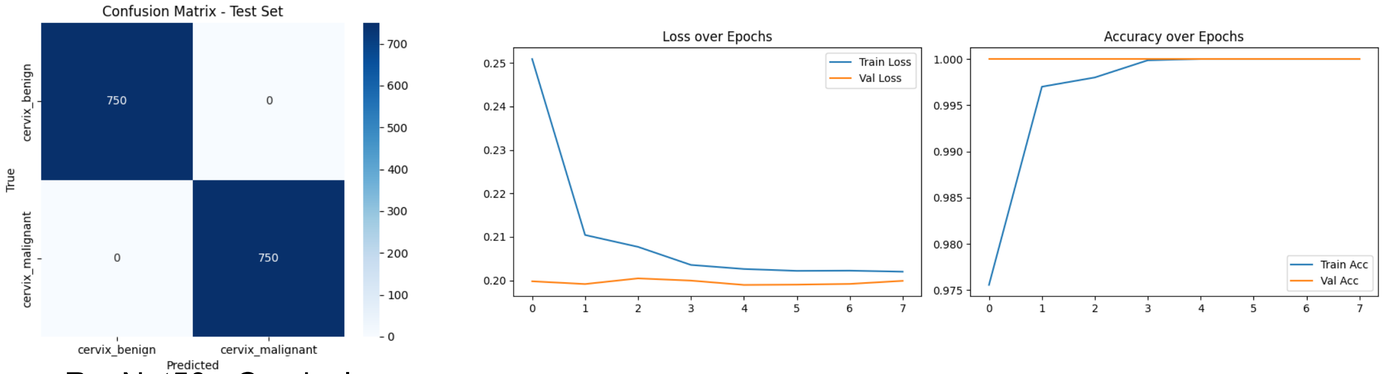
Table 1; Binary Classification Report



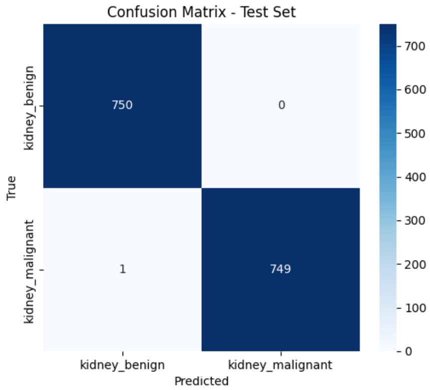
ResNet50 - ALL



ResNet50 - Breast



ResNet50 - Cervical



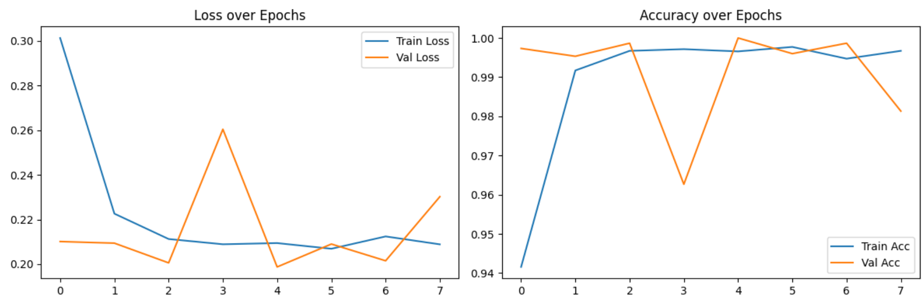
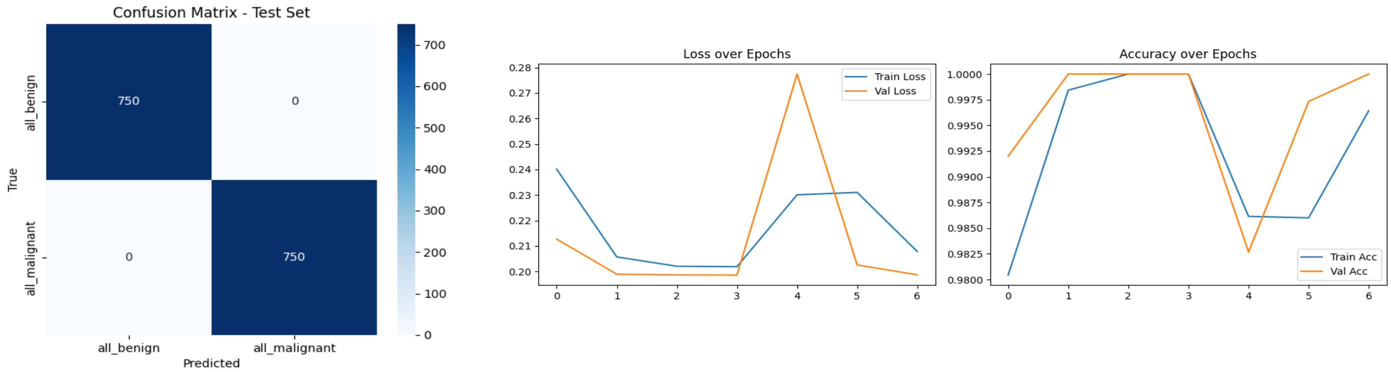
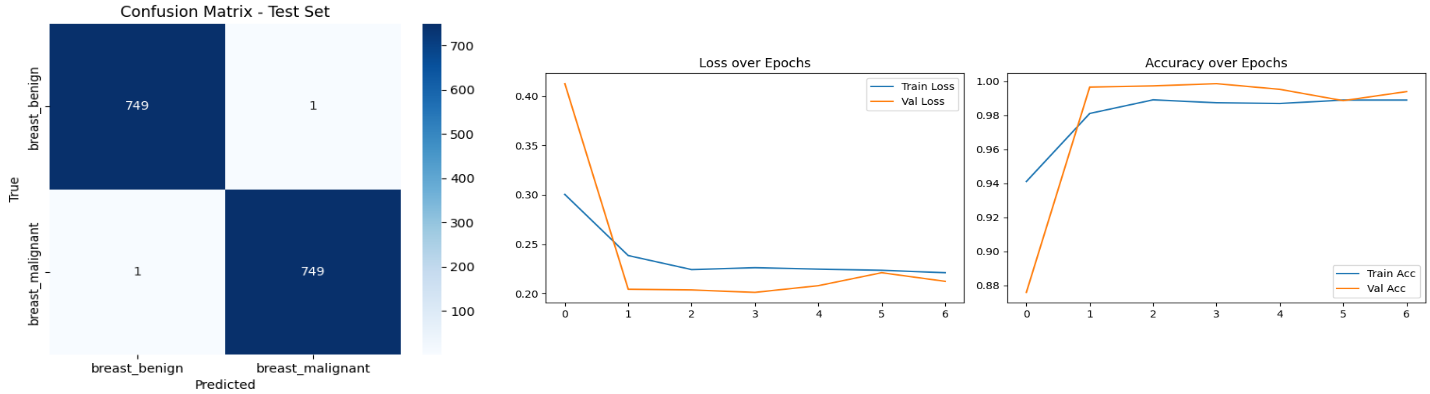


Figure 1; Binary Classification ResNet50

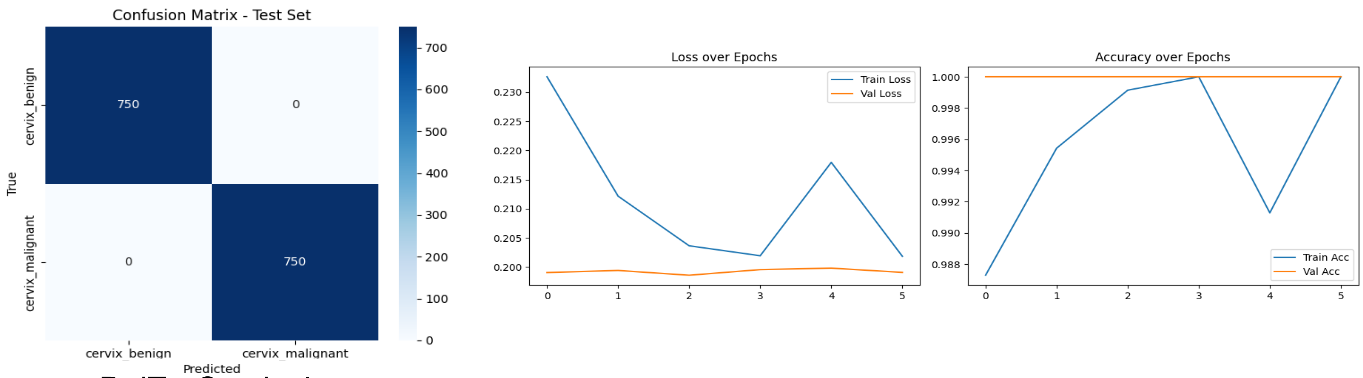
ResNet50 - Kidney



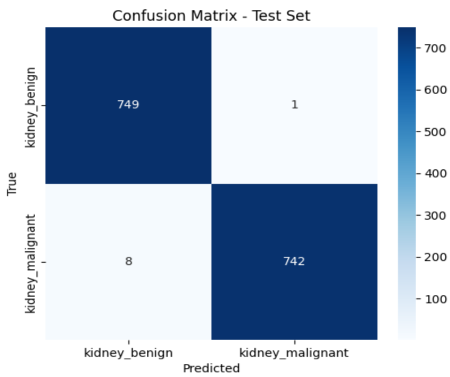
DeiT - ALL



DeiT - Breast



DeiT - Cervical



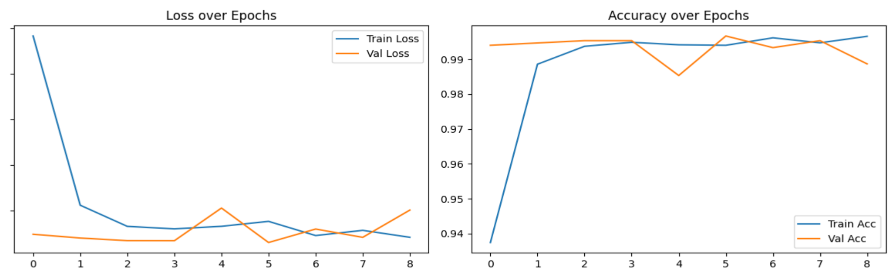


Figure 2; Binary Classification DeiT

DeiT - Kidney

Table 1 displays the binary classification tasks, both models demonstrated consistently high performance across all cancer types. ResNet50 achieved near-perfect results, never straying below 99%. Precision, recall and F1-scores remained at or close to 1.0000 for both malignant and benign classes, showing a strong balance between sensitivity and specificity. Similar, DeiT achieved equivalent results on the ALL and cervical datasets, with slightly lower performance on the breast and kidney cancer images. In the kidney dataset, performance dropped in malignant recall and benign precision, reflecting the small increase in misclassification seen in Figure 3.

Overall, both models demonstrated robust generalisation and class separation, with only minor fluctuations that can be due to dataset complexity or training variability.

Multi-Class Classification

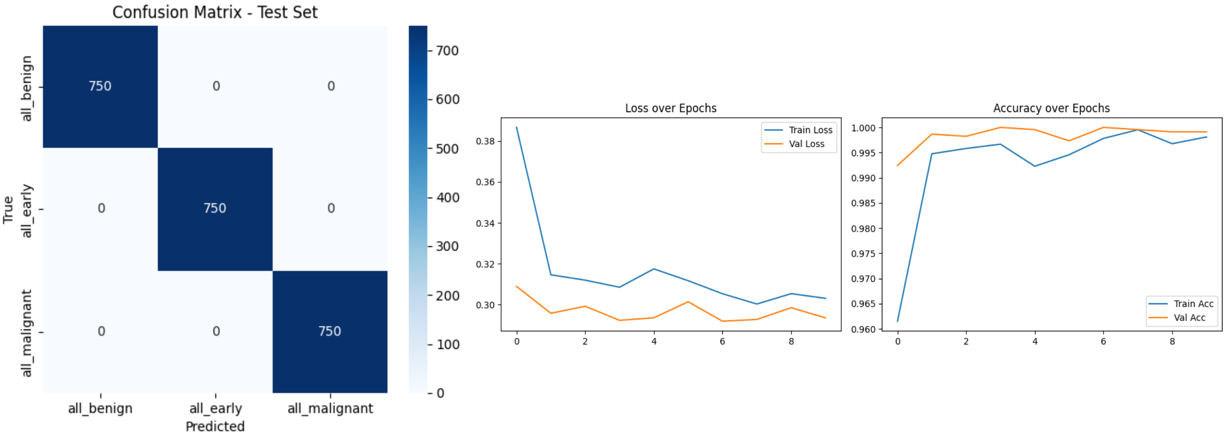
Model	Cancer Stage	Accuracy	Performance	Recall	F1-Score
ResNet50	Benign	100%	1.0000	1.0000	1.0000
	Early		1.0000	1.0000	1.0000
	Advanced		1.0000	1.0000	1.0000
DeiT	Benign	99.91%	1.0000	0.9973	0.9987
	Early		0.9973	1.0000	0.9987
	Advanced		1.0000	1.0000	1.0000

Table 2; Multi-Class Classification Report

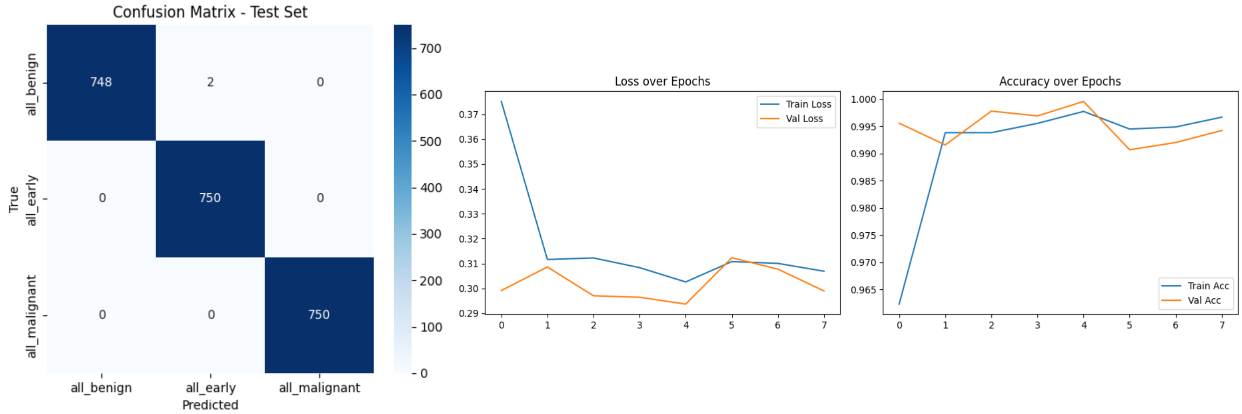
The multi-class classification tasks maintained a similarly high performance as seen in the binary classification across both ResNet50 and DeiT architectures. ResNet50 achieved a perfect score across all cancer stages, demonstrating an exceptional ability to distinguish not only benign and malignant cases but different stages of cancer, indicating robust feature extraction.

DeiT also performed exceptionally well, achieving a near-perfect accuracy. While benign, earl and advanced stages all exhibited near-perfect precision, recall and f1-scores, minor differences were observed. The benign class with recall dipping to 0.9973 with early-stage precision dropping to the same level. These deviations suggest a small number of misclassifications, as seen in Figure 3, this likely arose from subtle inter-class similarities or training variance.

Like with binary classification, both models displayed good generalisation to multiple cancer stages. These results indicate that both architectures are highly effective for multi-class cancer stage identification, extending the strong performance seen in the binary classification tasks to a more challenging multi-class setup.



ResNet50



DeiT

Figure 3; Multi-Class Metrics

Conclusion

The results demonstrate the strong capability of both ResNet50 and DeiT in cancer classification tasks across binary and multi-class settings. In the binary classification tasks, both architectures achieved near perfect accuracy, precision, recall and F1-scores, highlighting their ability to reliably distinguish malignant from benign cases. ResNet50 consistently reached 100% on ALL and Cervical datasets, while DeiT showed marginally lower performance on Breast and Kidney, indicating minor sensitivity to dataset complexity. The multi-class classification had both models maintain an exceptional performance when distinguishing between benign, early and advanced cancer stages. These findings indicate that both models generalise effectively for fine-grained detail found in cancer classification. Overall, the results confirm the effectiveness of deep learning for reliable and precise cancer diagnosis. However, it is difficult to make a definitive decision on which architecture is better given the scope of this project. ResNet50, and CNNs have proven robustness and reliability within medical imaging compared to DeiT's slight variability seen in this project. Though the dataset used was not particularly large, nor truly representative of a diverse, real-world scenario. DeiT models have greater potential for scaling when faced with real clinical data, providing better interpretability with its attention-based mechanisms.

UROS Experience

Participating in UROS has deepened my understanding of machine learning, particularly in the application of deep learning architectures to complex imaging problems but also modern techniques within the field. By implementing ResNet50 and DeiT across binary and multi-class classification, I have gained a much stronger understanding of the strengths and limitations of different model types. Working through the tasks highlighted not only technical aspects of model training and evaluation but the importance of transfer learning. Leveraging pretrained models allowed me to reach greater levels of performance even with limited data.

Beyond the technical implementation, this project refined my research skills. I learned how to critically evaluate results, not only in terms of raw metrics but also consider robustness, generalisability and potential clinical applications. Thinking of tasks to compare binary and multi-class settings taught me the importance of structuring a research question, defining clear evaluation criteria and interpreting outcomes in context.

Overall, this experience has been invaluable in strengthening my technical skills in machine learning but my broader research abilities as well. It has equipped me with the skills to approach future projects more critically and effectively.

References

Djoumessi, K., Mensah, S.O. and Berens, P. (2025). *A Hybrid Fully Convolutional CNN-Transformer Model for Inherently Interpretable Medical Image Classification*. [online] arXiv.org. Available at: https://arxiv.org/abs/2504.08481.

He, K., Gan, C., Li, Z., Rekik, I., Yin, Z., Ji, W., Gao, Y., Wang, Q., Zhang, J. and Shen, D. (2023). Transformers in medical image analysis. *Intelligent Medicine*, [online] 3(1), pp.59–78. doi: https://doi.org/10.1016/j.imed.2022.07.002.

Jia, H., Zhang, J., Ma, K., Qiao, X., Ren, L. and Shi, X. (2024). Application of convolutional neural networks in medical images: a bibliometric analysis. *Quantitative*

Imaging in Medicine and Surgery, 14(5), pp.3501–3518. doi: https://doi.org/10.21037/qims-23-1600.

Kumar, Y., Shrivastav, S., Garg, K., Modi, N., Wiltos, K., Woźniak, M. and Ijaz, M.F. (2024). Automating cancer diagnosis using advanced deep learning techniques for multi-cancer image classification. *Scientific Reports*, [online] 14(1), pp.1–25. doi: https://doi.org/10.1038/s41598-024-75876-2.

Obuli Sai Naren. (2024). Multi Cancer Dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/9537604.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A. and Jegou, H. (2021). *Training data-efficient image transformers & distillation through attention*. [online] proceedings.mlr.press. Available at: https://proceedings.mlr.press/v139/touvron21a.html.

Xu, W., Fu, Y.-L. and Zhu, D. (2023). ResNet and its application to medical image processing: Research progress and challenges. *Computer Methods and Programs in Biomedicine*, [online] 240, p.107660. doi: https://doi.org/10.1016/j.cmpb.2023.107660.