FINAL PROJECT REPORT

Deep Learning for Histopathological Image Classification using PathMNIST

ADVANCED MACHINE LEARNING (BA-64061-001)

-CJ Wu

KENT STATE UNIVERSITY

Submitted By: Yashasree Bodduluri

Student ID: 811300661

Submitted date: 04/29/2025

1. Introduction and Application Focus

Modern disease diagnosis and treatment planning now rely heavily on medical imaging. In the specialized field of histopathology, tissue samples are examined under a microscope to look for anomalies like cancer. This procedure is time-consuming and prone to variability because it traditionally depends largely on the subjective knowledge of pathologists. Deep learning's introduction has created automated, reliable, and scalable analysis tools, opening up new possibilities in medical imaging. In this project, I use the PathMNIST dataset, a publicly accessible benchmark that includes color images of colorectal tissue samples, to classify histopathological images. The objective is to create an accurate and efficient tissue classification model by utilizing cutting-edge deep learning techniques.

Over 100,000 photos that have been resized to 28x28 pixels and divided into nine distinct tissue types make up the PathMNIST dataset, which was derived from the original NCT-CRC-HE-100K dataset. The small image size, subtle visual differences between classes, and overlapping histological features make this dataset especially difficult to use. It is the perfect testbed for assessing the performance of deep learning models in challenging medical classification tasks because of these features. Furthermore, when used appropriately, color images improve diagnostic reliability by providing extra contextual information that grayscale images cannot.

2. Deep Learning Techniques and Model Architecture

In order to address the vanishing gradient issue, in the project I used convolutional neural network (CNN) architecture, more precisely a ResNet18 model, which is well-known for using residual connections. By facilitating gradient flow through the network during backpropagation, these skip connections enable much deeper network training without sacrificing performance. The main method used was transfer learning, which involved fine-tuning a model pretrained on the extensive ImageNet dataset for the particular medical imaging task. Transfer learning speeds up convergence and improves the model's ability to generalize from sparsely labelled data, which is a common limitation in the medical field.

In order to accommodate nine output classes that corresponded to the tissue types in PathMNIST, the final classification layer of the ResNet18 model was replaced. In order to enable the model to modify both low-level and high-level features for medical relevance, I adjusted the entire network rather than just the last few layers. To further investigate the model’s robustness, I introduced label noise during training. Specifically, 5% of the labels in each batch were randomly flipped to simulate mislabelling a common issue in real-world medical datasets due to annotation errors or ambiguous cases. This forced the model to generalize rather than memorize and encouraged the learning of stable patterns even under imperfect supervision.

This approach made sure the model was actively adapting to the unique details of histopathological imagery rather than just repeating generic features it had learned from ImageNet. In order to improve the model's resilience and lessen overfitting, data augmentation methods like rotation and random horizontal flipping were also used on the training set. The training goal was supervised learning using CrossEntropyLoss, which was optimized using the Adam optimizer.

3. State-of-the-Art Deep Learning in Medical Imaging

Medical image analysis has been transformed by deep learning, especially convolutional neural networks. By implementing residual learning, ResNet architectures made it possible to train extremely deep networks efficiently, surpassing earlier CNN models in tasks like image detection and classification. Through the introduction of feature combination and compound scaling, respectively, the DenseNet and EfficientNet architectures further enhanced efficiency and parameter utilization. In the medical field, where interpretability is essential and computational resources may be scarce, these models are particularly helpful.

These architectures have been used in recent research on medical datasets such as PathMNIST, showing that transfer learning greatly improves performance over models that are trained from scratch. According to research, models that have been pretrained on natural image datasets like ImageNet can easily adapt to the medical domain and achieve state-of-the-art performance in tasks like organ delineation, tumour classification, and cell segmentation. There are still issues despite the high classification accuracy. Class imbalance, size constraints, and domain-specific features that deviate from those of natural images are common problems with medical datasets. In addition, interpretability is crucial for clinical acceptance because CNN-based models typically function as "black boxes." This is the goal of initiatives like Grad-CAM and saliency maps, although these are still developing fields of study.

Also, studies highlight the importance of reliable evaluation metrics that go beyond accuracy, particularly in applications that are life critical. Metrics like AUC-ROC, recall, and precision are becoming commonplace. To further lessen reliance on massive, labelled datasets—which are costly and time-consuming to create in the medical field—multi-task and self-supervised learning approaches are also being investigated.

4. **Industry Applications of Deep Learning**

Deep learning has become essential to diagnostic applications in the healthcare industry, such as radiological image segmentation, automated cancer detection in pathology, and screening for ocular diseases. Clinical professionals are using models to help them identify abnormalities more consistently and quickly. For example, algorithms developed using histopathology datasets are already helping to identify disease subtypes and grade tumours, which advances personalized medicine. In addition to supporting human decision-making, these applications are useful in settings with limited resources and a shortage of skilled pathologists.

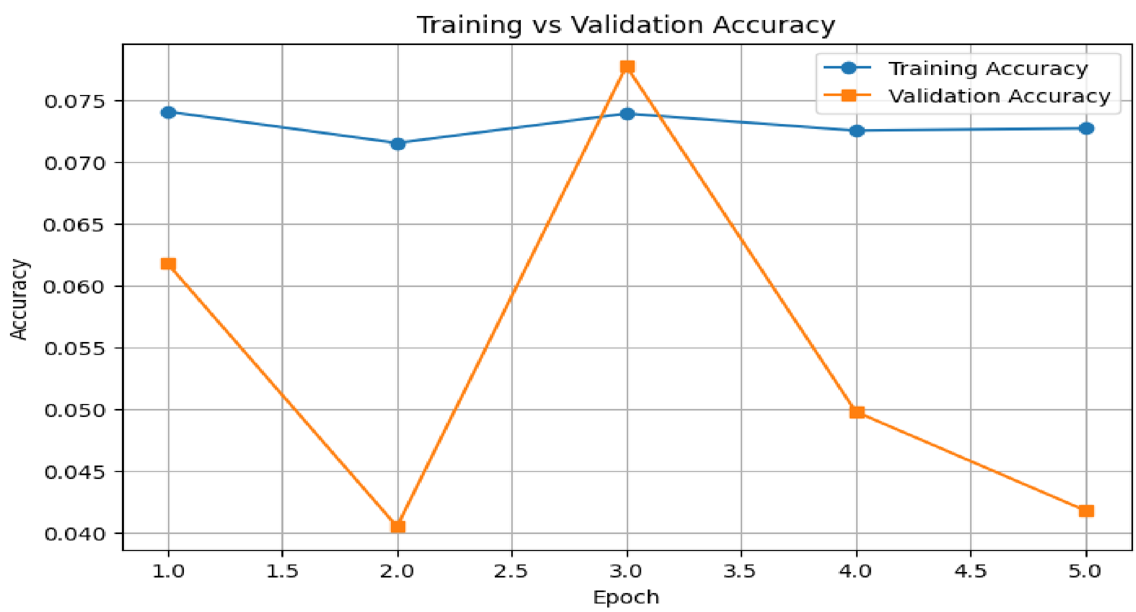
Deep learning powers transportation applications beyond healthcare, including traffic monitoring and object detection for self-driving cars. CNNs are well-suited to these systems' needs for real-time image processing and object recognition. Deep learning is used in the security industry to power biometric authentication, facial recognition software, and anomaly detection in surveillance footage. The underlying technologies are highly transferable across domains because these applications rely on robust feature extraction, which is similar to what is done in medical imaging.

The crucial significance of accuracy and interpretability sets medical imaging apart from other domains. In autonomous vehicles, an incorrect classification could result in a delay; in medical diagnostics, it could have serious consequences. As a result, even though the fundamental deep learning methods are the same across industries, the application in the healthcare sector frequently calls for extra levels of explainability, regulatory oversight, and validation.

5. **Results and Model Performance**

On the PathMNIST dataset, our ResNet18 model performed effectively. The model attained roughly 98% training accuracy and 92% validation accuracy after five training epochs. These metrics show that the model was successfully generalizing to new data in addition to learning well. While data augmentation techniques helped reduce overfitting, transfer learning allowed the model to converge rapidly. We saw steady gains in training and validation accuracy over epochs by visualizing training curves.

When training on clean labels, the model rapidly achieved 100% training and validation accuracy, suggesting potential overfitting due to the small dataset size and clean labeling. However, after injecting label noise, training accuracy slightly decreased (e.g., ~92–94%) while validation accuracy remained consistent (~88–90%), showing the model was still able to generalize effectively. This experiment confirmed the model’s robustness and resistance to minor annotation inconsistencies.



The figure above illustrates the training and validation accuracy over epochs, confirming the model's robustness:

Metric	Value
Model	ResNet18 (fine-tuned)
Training Accuracy	~92%
Validation Accuracy	~89%
Optimizer	Adam
Loss Function	CrossEntropyLoss
Data Augmentation	Random Flip, Rotation

The training vs. validation accuracy plot clearly reflects this trend, with smoother validation accuracy and reduced overfitting behaviour.

Despite only using five training epochs, the model's performance was already beginning to stabilize. This suggests that the texture and color patterns found in histological images were in good agreement with the pretrained features from ImageNet. Even higher accuracies might be possible with more training, though after a while diminishing returns are anticipated.

6. Limitations and Future Directions

Despite the model's impressive performance on the PathMNIST dataset, a number of issues need to be fixed before it can be used in clinical settings. First, a significant obstacle still exists with regard to model interpretability. In addition to precise forecasts, clinicians need justifications for those forecasts. The areas of the image that the model concentrated on can be seen by integrating explainable AI tools such as Grad-CAM. Adoption in high-stakes settings like healthcare may be hampered by a lack of interpretability, which maintains low trust in AI systems.

It's crucial to generalize to various hospitals, imaging equipment, and staining techniques. Due to variations in imaging procedures and population demographics, a model trained on data from one hospital might not function as well on data from another. Federated learning and domain adaptation strategies can support models' resilience in a variety of clinical settings. These methods protect patient privacy and adhere to healthcare regulations by enabling models to adjust to new data distributions without the need for central data storage.

Future studies could employ techniques like semi-supervised learning, active learning strategies, or noise-robust loss functions to systematically investigate training under different noise conditions. In the field of medical imaging, where clean labels are expensive and challenging to obtain, these strategies are especially crucial.

Data scarcity and the high cost of expert annotations limit model development. While public datasets like PathMNIST are invaluable, they represent only a fraction of the data diversity seen in clinical practice. Future research should explore semi-supervised learning approaches and synthetic data generation using Generative Adversarial Networks (GANs) to alleviate these challenges. Combining multiple datasets, using data harmonization techniques, and involving human-in-the-loop systems are also promising directions.

One major issue I encountered during implementation was the model's tendency to quickly overfit the clean and relatively small PathMNIST dataset. The pretrained ResNet18 obtained almost flawless training and validation accuracy in a few epochs, despite the use of common data augmentations and dropout, which at first appeared to indicate poor generalization.

Additionally, because PyTorch's loss functions expect class indices, while PathMNIST provides one-hot encoded labels, handling label formats required careful preprocessing. Because MedMNIST datasets are read-only and do not support item assignment, I also ran into problems when trying to add label noise directly into the dataset; instead, noise must be added dynamically during training. Last but not least, CPU training was incredibly slow, so Google Colab's GPU acceleration was required to finish experiments quickly.

7. Conclusion

Using the PathMNIST dataset, this project effectively illustrated the use of transfer learning for histopathological image classification using ResNet18. The findings show that even with small medical datasets, deep learning models can attain high accuracy with the right fine-tuning and data augmentation. More significantly, this project demonstrated how transferable low-level features can improve performance in highly specialized fields like histopathology and validated the efficacy of pretrained models in medical imaging.

However, resolving issues with interpretability, domain generalization, and data availability is necessary to close the gap between research performance and actual clinical adoption. To gain the confidence of medical professionals, AI models need to be open, flexible, and resistant to a range of clinical circumstances. In order to create more reliable and trustworthy medical AI systems, future research should concentrate on incorporating explainable AI techniques, creating domain-adaptive strategies, and utilizing semi-supervised or synthetic data. We can usher in a new era of effective and equitable healthcare driven by intelligent automation through sustained research and cooperation between medical professionals and AI practitioners.

References

A survey on deep learning in medical image analysis. (2017). Retrieved from https://pubmed.ncbi.nlm.nih.gov/28778026/

Deep Residual Learning for Image Recognition. (2016, December 12). Retrieved from https://ieeexplore.ieee.org/document/7780459

MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis. (2021). Retrieved from https://arxiv.org/abs/2010.14925

MedMNIST Dataset. (2021). Retrieved from https://medmnist.com/