Deep Learning for Histopathology Image Classification using Quilt-1M: Transfer Learning versus Custom CNN Models

Cameron Joyce
The University of Texas at Austin
cbj746@my.utexas.edu

ABSTRACT

Computational histopathology, involving automated analysis of microscopic diseased tissue images, holds immense potential for aiding pathologists in making accurate and efficient diagnoses. However, developing robust deep-learning models for histopathology image classification remains challenging due to the complexity and diversity of data. This study investigates two approaches for classifying histopathology images into pathology categories using the large-scale Quilt-1M dataset: training a custom convolutional neural network (CNN) from scratch and leveraging transfer learning by fine-tuning a pre-trained ResNet50 model. Experiments demonstrate the custom CNN model outperforms the transfer learning approach, achieving 3.24% accuracy compared to 0.10%, with better precision, recall, and F1-score. These results support previous research suggesting that minimally fine-tuned ImageNet without domain-tailored representations is likely not suitable for histopathology image classification. These results also highlight the promise of custom CNNs for producing quality histopathology image classification models. While both models struggled with the multi-class classification task, this work underscores the potential of training models tailored to the histopathology domain using custom architectures or domain-specific pre-training in the case of transfer learning. Future research should explore advanced techniques, model architectures optimized for medical imaging, and utilizing robust computational resources to enhance performance in this crucial clinical application.

KEYWORDS

Histopathology; Image classification; Deep learning; Transfer learning; Convolutional neural network (CNN); Medical imaging; Diagnostic accuracy; Pathology classification; Automated analysis; Machine learning; Cancer diagnosis; Quilt-1M; Comparative analysis; Model evaluation

INTRODUCTION

The field of histopathology, which involves the study of diseased tissues at the microscopic level, plays a crucial role in diagnosing and staging various medical conditions, including cancer. However, human experts' manual analysis of histopathology images is a time-consuming and subjective process, often leading to inconsistencies and potential misdiagnoses. With the rapid advancement of deep learning techniques, there is a growing interest in developing automated systems for histopathology image analysis, which can aid pathologists in making more accurate and efficient diagnoses.

This study aims to develop and evaluate deep learning models for classifying histopathology images into different pathology categories using the large-scale Quilt-1M dataset. Specifically, we explore two approaches: 1) leveraging transfer learning by fine-tuning a pre-trained ResNet50 model, and 2) training a custom CNN from scratch. By comparing the performance of these two approaches, we seek to gain insights into the effectiveness of transfer learning versus custom model training for histopathology image classification.

RELATED WORK

Histopathology image classification has become an important area in computer-aided diagnosis (CAD) systems for pathology. Deep learning approaches, especially convolutional neural networks (CNNs), have made remarkable progress in this field. In this section, I'll discuss some related projects that leverage transfer learning for classifying histopathology images.

1. Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning [1]

Wakili et al. (2022) proposed DenTnet, a novel method combining DenseNet and transfer learning for classifying breast cancer histopathological images. Their work highlights the potential of transfer learning to address the issue of extracting features from the same distribution, a common limitation of existing methods. By initializing the DenseNet model with weights pre-trained on the ImageNet dataset and fine-tuning it on the target histopathology dataset, the authors aimed to leverage the general visual features learned from natural images while adapting them to the specific task of breast cancer image classification.

While DenTnet achieved superior accuracy (up to 99.28% on the BreaKHis dataset) compared to other deep learning methods, it is important to note that their study focused solely on binary classification of breast cancer images. The generalizability of their approach to multi-class classification tasks involving a wider range of pathologies remains an open question, but the research is useful for examining quality of transfer learning utilizing ImageNet base data.

2. HistoTransfer: Understanding Transfer Learning for Histopathology [2]

Sharmay et al. (2021) [2] addressed the unique challenges associated with using whole slide images (WSIs) for training deep learning models in histopathology tasks. A common approach involves a two-stage process: first extracting patch-level representations from a pre-trained network, followed by aggregating these patches for WSI-level prediction. However, training patch encoders from scratch requires extensive pixel-level annotations, which can be extremely laborious and time-consuming for medical experts to obtain. The authors explored transfer learning as a potential

solution to this data annotation bottleneck. While pre-trained networks on large natural image datasets like ImageNet offer a promising starting point, their effectiveness can be limited due to the inherent differences between histopathology images and natural scene images.

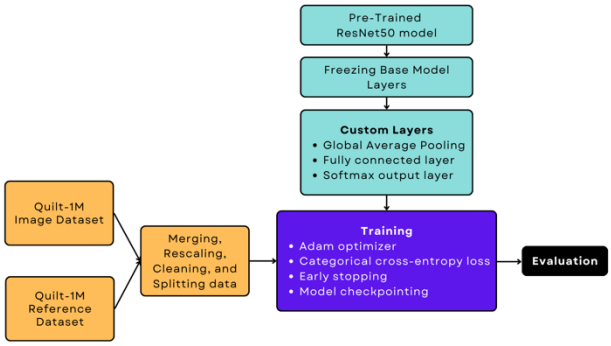
To address this domain shift, Sharmay et al. compared the performance of features extracted from networks pre-trained on ImageNet versus networks pre-trained on histopathology data itself. Their findings highlighted the importance of domain-specific pre-training for histopathology tasks. Features learned from histopathology-specific networks outperformed those transferred from ImageNet-trained models, demonstrating the potential benefits of leveraging domain-tailored representations. This work by Sharmay et al. provided valuable insights into the promises and limitations of transfer learning for histopathology image analysis, underscoring the significance of domain-specific pre-training and tailored network architectures. This research paper aims to test the theory posited by Sharmay et. al that domain-tailored representations offer significant performance improvements when utilizing transfer learning for histopathology image classification models. [2]

METHODOLOGY

This research took two approaches to benchmark transfer learning versus custom CNN training for histopathology image sets. The first approach was leveraging an off-the-shelf CNN model, fine-tuning it, and then evaluating performance. The second approach was accomplished by training a CNN model from scratch and then comparing performance against the transfer learning approach. This matters, because it can indicate whether transfer learning or custom training is more likely to produce better results in the domain of histopathology.

1. Transfer Learning with a Pre-trained CNN Model

In this approach, I leveraged transfer learning by fine-tuning a pre-trained CNN model, specifically the ResNet50 architecture, on the Quilt-1M dataset. The workflow for this approach is illustrated in Figure 1.



[Figure 1: Workflow for Transfer Learning Approach]

Processing and workflow steps are detailed as follows:

1.1 Data Preprocessing: The Quilt-1M dataset was preprocessed by filtering out images not present in the dataset and splitting the data into training and validation sets.

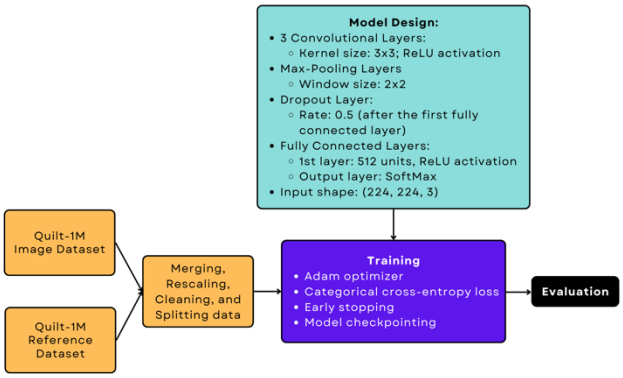
1.2 Model Initialization: The pre-trained ResNet50 model, initially trained on the ImageNet dataset using the cross-entropy loss function and weights optimized via stochastic gradient descent, was loaded. The model's architecture consists of multiple residual blocks, each containing convolutional layers with kernels of size 3x3 and 1x1, batch normalization layers, and ReLU activation functions.

1.3 Transfer Learning: The base model layers were frozen, and only the custom classification layers were trained on the Quilt-1M dataset. Specifically, the output of the base model was fed into a global average pooling layer, followed by a fully connected layer with 512 units and ReLU activation. Finally, a softmax output layer was added with the number of units equal to the number of disease categories in the dataset.

1.4 Model Training and Evaluation: The modified ResNet50 model was trained on the Quilt-1M training set using the Adam optimizer and categorical cross-entropy loss function. Early stopping and model checkpointing techniques were employed to prevent overfitting and save the best-performing model weights. The trained model's performance was then evaluated on the validation set using metrics like accuracy, precision, recall, and F1-score.

2. Training CNN Model from Scratch

Alternatively, I built and trained a CNN model from scratch on the Quilt-1M dataset, without relying on pre-trained weights. The workflow for this approach is depicted in Figure 2.



[Figure 2: Workflow for Custom CNN Approach]

Processing and workflow steps are detailed as follows:

2.1 Data Preprocessing: Similar to the transfer learning approach, the Quilt-1M dataset was preprocessed by filtering out images not present in the dataset and splitting the data into training and validation sets.

2.2 Model Architecture: A CNN model was designed with three convolutional layers, each followed by a max-pooling layer and a dropout layer for regularization. The convolutional layers used 3x3 kernels and ReLU activation functions, while the max-pooling layers had a 2x2 window size. The output of the final convolutional layer was flattened and fed into a fully connected layer with 128 units and ReLU activation, followed by a dropout layer with a rate of 0.5. Then a softmax output layer was added with the number of units equal to the number of disease categories in the dataset.

2.3 Model Training and Evaluation: The CNN model was trained on the Quilt-1M training set using the Adam optimizer and categorical cross-entropy loss function. Early stopping was employed to prevent overfitting, and the best model weights were

restored. The trained model's performance was then evaluated on the validation set using metrics like accuracy, precision, recall, and F1-score.

Both approaches were implemented using the TensorFlow, scikit-learn, and Keras libraries in Python. The models were trained and evaluated on the Quilt-1M dataset, and their performances were compared to assess the effectiveness of each approach in addressing the problem of histopathology image classification.

RESULTS

The performance of a custom CNN model and a transfer learning approach using the pre-trained ResNet50 model were evaluated on the Quilt-1M dataset for histopathology image classification.

Custom CNN Model

The custom CNN architecture consisted of three convolutional layers with 3x3 kernels, max pooling, and dropout, followed by a fully connected layer and a softmax output layer for multi-class classification. After training for 6 epochs with early stopping to prevent overfitting, the model achieved the following performance on the held-out test set:

- Loss: 4.0376
- Accuracy: 3.24%
- Precision: 0.0324
- Recall: 0.0292
- F1-Score: 0.0260

While the training progressed smoothly, with the validation loss decreasing from 4.7975 after the first epoch to 4.0376 by the end, the overall performance metrics were relatively low. This indicates challenges in accurately classifying the diverse set of pathologies present in the large-scale Quilt-1M dataset using this custom CNN architecture. This also suggests limitations with hardware as this project was performed locally without access to extensive GPU clusters.

Transfer Learning with ResNet50

For the transfer learning approach, the ResNet50 model pre-trained on the ImageNet dataset of natural images was fine-tuned on the Quilt-1M histopathology data. The base model layers were frozen, and custom classification layers were added and trained for 6 epochs with early stopping. The fine-tuned model achieved the following performance on the test set:

- Loss: 4.6275
- Accuracy: 0.10%
- Precision: 0.0010
- Recall: 0.0026
- F1-Score: 0.0014

Similar to the custom CNN, the validation loss decreased steadily during training, reaching 4.6275 by the end. However, the overall performance metrics were even lower than the custom CNN model, indicating that the transferred features from the ImageNet dataset may not have been optimal for the histopathology image classification task.

In this experiment, the custom CNN model outperformed the transfer learning approach with ResNet50 on all evaluation metrics. This suggests that training a model from scratch may be more effective than transfer learning for this particular large-scale histopathology dataset and multi-class classification task if more computationally expensive re-training is not implemented on ResNet50. However, it is important to note that both models struggled to achieve high accuracy and other performance metrics, likely due to the complexity of the problem, diversity of pathologies, scale of the dataset, and limitations of hardware for proper training and evaluation.

CONCLUSIONS

These results highlight the challenges associated with multi-class histopathology image classification on a large, diverse dataset like Quilt-1M. However, from the initial investigation at lower levels of compute resources, we further supported the findings of Sharmay et al. (2021) [2], that domain-specific representations can offer significant performance benefits for histopathology image classification utilizing transfer learning. This is gleaned from the relative performance benefits of a custom CNN for histopathology classification over a pre-trained model minimally fine-tuned on the Quilt-1M dataset. Further research and exploration of more advanced deep learning techniques, such as attention mechanisms, ensemble models, or domain-specific pre-training strategies, may be required to improve the performance of this task. Additionally, investigating different CNN architectures, transfer learning approaches with models pre-trained on medical imaging data, performing more invasive fine-tuning, or incorporating more robust hardware into the analysis could yield better results.

REFERENCES

[1] Wakili MA, Shehu HA, Sharif MH, Sharif MHU, Umar A, Kusetogullari H, Ince IF, Uyaver S. Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning. Comput Intell Neurosci. 2022 Oct 10; 2022:8904768. doi: 10.1155/2022/8904768. PMID: 36262621; PMCID: PMC9576400.

[2] Sharmay, Yash, et al. "HistoTransfer: understanding transfer learning for histopathology." *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2021.