Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 3, No. 2, 2025

Automatic Deep Learning-based Histopathologic Image Classification

Hongfei Zhao

Henan University, Kaifeng, China

Abstract: Histopathologic image analysis is a critical component in cancer diagnosis, yet traditional manual inspection methods are often time-consuming, subjective, and error-prone. This study presents a fully automated deep learning framework for the classification of histopathologic images stained with Hematoxylin and Eosin (H&E). Leveraging Convolutional Neural Networks (CNNs), particularly DenseNet and ResNet architectures, the proposed system integrates essential components such as data preprocessing, augmentation, hyperparameter optimization, and training automation using the fastai library. Experiments were conducted on the PCam dataset derived from Camelyon16, comprising over 320,000 labeled image patches. Results demonstrate that DenseNet outperforms ResNet in terms of accuracy and AUC, achieving 84.37% test accuracy and 0.96 AUC. The framework shows high reproducibility, efficiency, and potential for clinical deployment, offering a scalable solution to improve diagnostic accuracy in pathology. Future directions include exploring hybrid models and advanced augmentation techniques to further enhance classification performance.

Keywords: Histopathology; Deep Learning; DenseNet; ResNet; Image Classification; PCam Dataset.

1. Introduction

1.1. Aim of the Study

The main aim of this study is to develop an automated deep learning framework for classifying histopathologic images, focusing on images stained with Hematoxylin and Eosin (H&E). It aims to provide a powerful and easy-to-use system and improve the cancer diagnosis process and tackle the challenges of manual image analysis. It also offers an automated solution to enhance both the accuracy and efficiency of histopathologic image classification.

1.2. Importance of the Study

Histopathologic image analysis is crucial for the clinical evaluation and diagnosis of cancer. It helps determine the cancer stage and how far it has spread. Previous attempts for the histopathologic image analysis are relied on visual observe of the physicians, which can be time-consuming, subjective and error-prone. Using machine learning to automate this process can significantly improve diagnostic accuracy, reduce the workload of pathologists, and provide more consistent results. With the growing availability of large annotated medical image datasets, deep learning has the potential to change medical diagnostics. Therefore, it is important to develop a fully automated framework for histopathologic image classification that can independently construct the machine learning pipeline for training and be easily deployed.

1.3. General Introduction of the Algorithm

This study uses the method based on Convolutional Neural Networks (CNN) and employs widely used architectures as DenseNet and ResNet. CNN is very effective in processing image data, especially for tasks which involve pattern recognition in complex visual inputs. The framework includes key components such as data pre-processing, data augmentation, model training, and hyperparameter selection. All of these processes are automated to simplify deployment. By integrating advanced techniques like random data augmentation and automatic learning rate optimization with the one-cycle policy, the framework aims to maximize classification accuracy and minimizing the need for manual intervention.

1.4. Introduction of the Chosen Dataset

The study uses the PCam dataset [1], which is derived from the Camelyon16 Challenge dataset [2]. This dataset contains 400 whole-slide images of sentinel lymph node sections that were stained with H&E and scanned at 40x magnification [1]. For this research, the dataset was downsampled by a factor of 10, resulting in a pixel resolution of 2.43 microns. The dataset is well-suited for histopathologic image classification due to the large number of labeled images, making it an ideal resource for training deep learning models. It consists of 327,680 color images (96 x 96 pixels) extracted from histopathologic scans of lymph node sections. The dataset is divided into a training set of 262.144 (2^18) examples, and a validation and test set both of 32.768 (2^15) examples. There is no overlap in WSIs between the splits, and all splits have a 50/50 balance between positive and negative examples. A positive label indicates that the center 32x32px region of a patch contains at least one pixel of tumor tissue. Tumor tissue in the outer region of the patch does not influence the label. This outer region is provided to enable the design of fully-convolutional models that do not use any zero-padding, to ensure consistent behavior when applied to a whole-slide image.

2. Techniques

2.1. The Method Used in The Paper

The method presented in the paper is based on automatic deep learning techniques, particularly using Convolutional Neural Networks (CNNs) for the classification of histopathologic images. The framework automates the entire machine learning pipeline, from data preprocessing to model training and evaluation, making it accessible even for non-

experts in machine learning. Below are the key principles of the method used:

The core method for classification relies on CNNs, which are widely used for image recognition tasks due to their ability to automatically learn hierarchical features from raw image data. In this study, two popular CNN architectures, DenseNet and ResNet, are employed:

2.1.1. DenseNet

DenseNet [3] is a CNN architecture known for its dense connections between layers. Instead of the traditional sequential flow of information from one layer to the next, DenseNet introduces connections where each layer receives input from all preceding layers and passes its output to all subsequent layers. This creates a densely connected network where:

Feature Reuse: Each layer in DenseNet can directly access the feature maps learned by all previous layers, promoting feature reuse and making the network more efficient. This leads to fewer parameters since layers don't have to relearn redundant features.

Gradient Flow: The dense connectivity helps gradients flow smoothly during backpropagation, allowing deeper networks to be trained without the problem of vanishing gradients.

Parameter Efficiency: DenseNet achieves high accuracy with fewer parameters compared to traditional CNNs, due to its efficient feature reuse.

Training Efficiency: DenseNet reduces overfitting due to its compact design and the fact that fewer parameters need to be trained, even though the architecture is deeper.

The use of DenseNet in histopathologic image classification benefits from its ability to capture fine details in medical images, where small variations (such as the presence or absence of metastases) are crucial for accurate diagnosis.

2.1.2. ResNet

Resnet [4] is another CNN architecture, introduced to address the problem of vanishing gradients in deep networks. ResNet key innovation is the introduction of residual connections, which allow information to bypass one or more layers:

Residual Learning: Instead of learning the full mapping from input to output, ResNet layers learn the residual (difference) between the input and the desired output. This means that each layer focuses on refining the features learned by earlier layers, leading to more efficient learning.

Skip Connections: These residual connections enable direct paths for the gradient to propagate through the network during backpropagation, even in very deep networks. This allows ResNet to train very deep architectures (such as ResNet-50, ResNet-101, and ResNet-152) without suffering from vanishing gradients.

Deeper Networks: ResNet enables the construction of very deep networks (with hundreds of layers) while maintaining strong performance and convergence, which is particularly useful for complex image classification tasks like histopathologic image analysis.

In the context of this paper, ResNet is advantageous because its depth allows the network to capture very fine-grained features that are necessary for distinguishing subtle differences in histopathologic images, such as cancerous versus non-cancerous tissues.

2.2. The Functionality of Each Building Block

2.2.1. Data loading

The dataset used in this framework is in HDF5 (h5) format, so we utilize the h5py library to manage the data. The h5py library is a Python package that interfaces with Hierarchical Data Format version 5 (HDF5), this format is especially suitable for storing and accessing large amounts of data, such as images and numerical arrays, especially when the dataset is too large to fit into memory all at once. By using h5py, the framework can efficiently read and write parts of the data directly from the disk without loading everything into memory, making it easier to handle large datasets smoothly.

2.2.2. Pre-processing

Before training the model, the images are resized to a fixed size of 96x96 pixels. This is to ensure consistent input dimensions. In addition, the data is normalized by subtracting the mean and dividing by the standard deviation to standardize pixel values. This can bring the data into a similar range, speeding up the training process and improving convergence.

2.2.3. Random Data augmentation

To further improve the robustness of the model and reduce overfitting, random data augmentation techniques are applied during training. These include:

Random Rotation: Rotating the images by a random degree to simulate different orientations.

Random Cropping: Extracting random regions from the images and resizing them to the desired input size.

Random Flipping: Horizontally or vertically flipping the images with a certain probability.

Lighting Adjustments: Altering brightness and contrast to create variations in the lighting conditions.

These augmentations help the model generalize better to unseen data by artificially expanding the training dataset.

2.2.4. Model implementation

The framework implements both DenseNet and ResNet architectures, two popular deep learning models known for their ability to efficiently learn hierarchical features from image data:

DenseNet: Promotes feature reuse by connecting each layer to all previous layers, improving gradient flow and reducing the number of parameters.

ResNet: Utilizes residual connections that help deeper networks avoid vanishing gradient problems and allow them to learn efficiently.

The training process is automated using the fastai library, which facilitates:

Hyperparameter Selection: Automated using the one-cycle policy, which optimizes learning rate and weight decay to improve training speed and convergence.

Early Stopping: Automatically halts training when the validation performance stops improving, preventing overfitting.

Optimizer and Loss Function: The framework uses Stochastic Gradient Descent (SGD) as the optimizer and cross-entropy as the loss function, which is common for classification tasks.

2.2.5. Evaluation and Comparison

To assess the performance of DenseNet and ResNet models, the following steps are taken:

Loss and Accuracy: The performance of both models is evaluated on the test and validation datasets using loss and accuracy metrics.

AUC-ROC Curve: The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is used as a key evaluation metric. This metric assesses how well the model distinguishes between classes (e.g., cancerous vs. non-cancerous tissues).

Visualization: The results, including loss and accuracy, are presented using relevant charts to visually compare the performance of DenseNet and ResNet.

This approach allows for a detailed comparison between the two architectures in terms of their ability to classify histopathologic images accurately.

3. Experiments and Results

3.1. Experimental Setup

3.1.1. Dataset

We use the PCam benchmark dataset in the experiments, which has been divided into three HDF5 files: training set (262144 samples), validation set (32768 samples) and test set (32768 samples). The Figure 1 below shows 3 random negative and positive examples in the dataset.

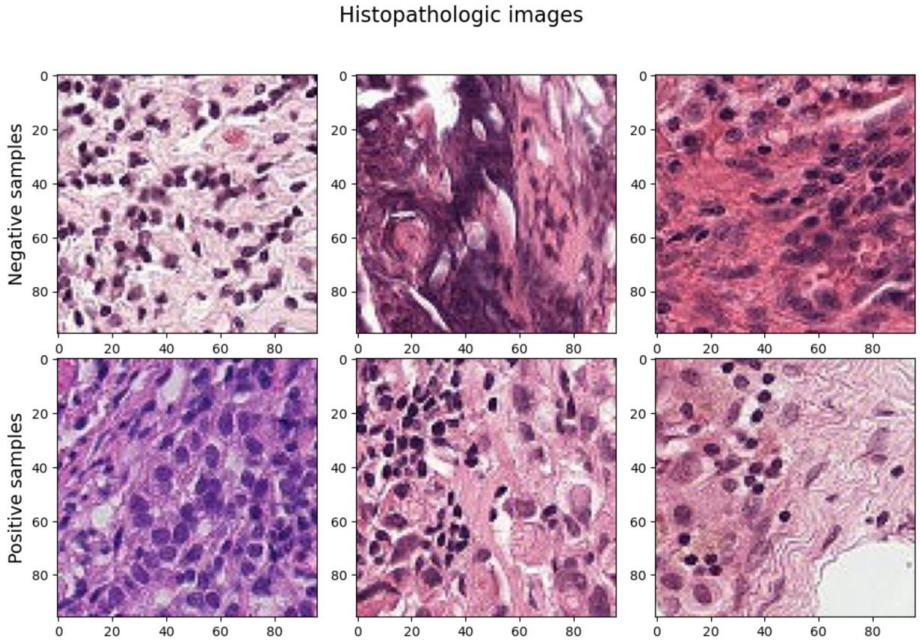

Figure 1. Sample of the histopathologic images

3.1.2. Training configurations

Number of Epochs:

The model training process set up with 10 epochs for both DenseNet and ResNet models.

An early stopping function was considered to prevent overfitting and running unnecessary training, it will be triggered if there is no improvement in validation loss over three consecutive epochs.

Batch Size:

The batch size set to 32 during training with fastai ImageDataLoaders.

During the data loading process with PyTorch's DataLoader, the batch size set to 64 for both the training and validation datasets. This allows for training with decent performance while also maintaining good memory usage.

Learning Rate:

For deep learning model training, setting hyperparameters is always a big challenge. We follow the One-Cycle Policy [5] mentioned in the original paper to find the best learning rate and weight decay. The code uses fastai's lr_find() function to determine the learning rate for the two models, and then uses fine_tune to perform transfer learning and fine-tuning to ensure efficient and dynamic adjustment of the learning rate throughout the training process.

Through One-Cycle Policy, the best learning rate for DenseNet is 00057, and ResNet is 0.0012.

The training process use these learning rates for fine-tuning.

Hardware and Device:

The models were trained using a GPU (cuda), ensuring faster training times and efficient computation.

Optimizer and Loss Function:

SGD as the optimizer was used with a specific momentum of 0.9 and a weight decay of 0.01 for both models.

Cross-Entropy Loss function - a standard loss definition for binary classification tasks.

3.2. Reproducing the Results

The performance of the model was evaluated and compared with the results reported in the original study using the accuracy of the training, validation and test sets as well as key metrics such as loss curves and ROC curves.

3.2.1. Accuracy on Training, Validation, and Test Sets

Table 1 shows that both DenseNet and ResNet models achieved high accuracy across training, validation, and test datasets, with DenseNet consistently outperforming ResNet in each case:

Training Accuracy: DenseNet achieved a training accuracy of 95.68%, indicating effective learning from the training data. ResNet showed a slightly lower training accuracy of 93.77%, means it required more effort to fit the data compared to DenseNet.

Validation Accuracy: For validation accuracy, DenseNet is 88.12%, while ResNet reached 86.86%. The higher validation accuracy for DenseNet indicates better generalization performance during training.

Test Accuracy: On the test dataset, DenseNet with an accuracy of 84.37% is better than ResNet 82.49%. This result suggests that DenseNet's architecture provided a better ability to generalize to new, unseen data.

Table 1. Accuracy for DenseNet and ResNet

Model	Train-accuracy	Valid-accuracy	Test-accuracy
DenseNet	0.9568	0.8812	0.8437
ResNet	0.9377	0.8686	0.8249

3.2.2. Loss Curves

Figure 2 shows the training and validation loss trends over the epochs for both DenseNet and ResNet models:

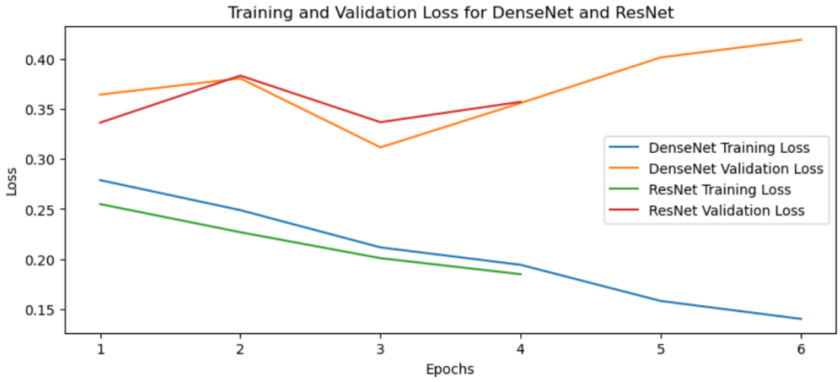

Figure 2. Training and Validation Loss

Training Loss Trends: Both DenseNet and ResNet show a steady decrease in training loss over the epochs. When there is no significant improvement in validation loss over several epochs, an early stopping mechanism is triggered to stop training. This approach helps prevent overfitting because it stops the training process before the model starts to memorize too much training data.

Validation Loss Trends: The validation loss for both models show some wave motion, particularly after the 2nd epoch, where early stopping played an impotent role. For DenseNet, the training was stopped after epoch 2 when there was no improvement in the validation loss, means that further training might not lead to better generalization. As the same,for ResNet, early stopping occurred very early in the process, indicating that the model reached its optimal state quickly.

Impact of Early Stopping: Using early stopping helps stabilize the performance of the model, ensuring that training does not continue once the model starts to overfit. This can be observed in the validation loss curve, where the upward trend is stopped by the early stopping mechanism. Thus, it helps maintain a balance between minimizing the training loss and keeping the validation loss low, thereby improving generalization performance on unseen data.

3.2.3. ROC curves

The Receiver Operating Characteristic (ROC) curves for the ResNet and DenseNet models illustrate their performance in distinguishing between positive and negative samples across various classification thresholds:

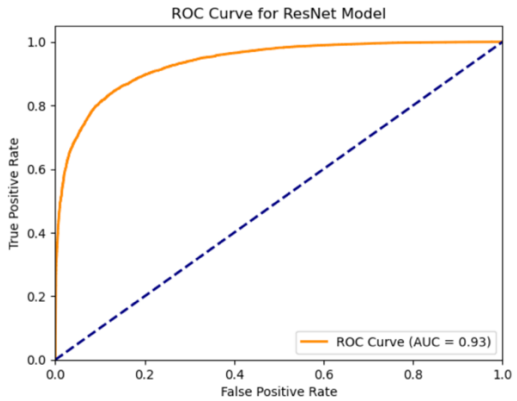

Figure 3. ROC Curve for ResNet Model

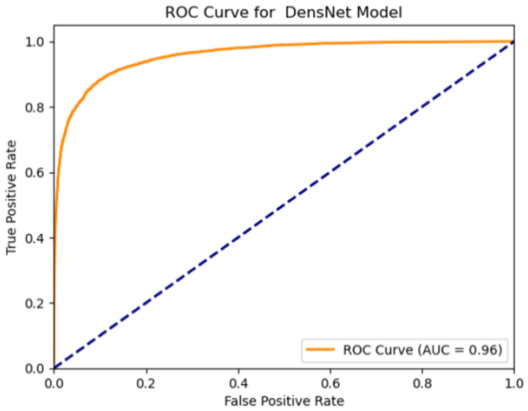

Figure 4. ROC Curve for DensNet Model

ResNet Model:

The ROC curve for the ResNet model shows an AUC of 0.93, indicating strong discriminatory power. The curve stays well above the diagonal line (representing random guessing), demonstrating that the model can effectively differentiate between classes.

The shape of the curve, which rises quickly towards the top-left corner, suggests that the model achieves a high true positive rate with a relatively low false positive rate for most thresholds.

DenseNet Model:

The ROC curve of the DenseNet model has a greater AUC of 0.96, and that means it is more optimal than ResNet. At the same time, it conveys a more powerful way to distinguish between the categories if the curve is very close to the upper left corner.

The higher AUC for DenseNet supports the earlier findings from the accuracy and loss metrics, further confirming its effectiveness in histopathologic image classification.

3.2.4. Discussion of Results

The reproduced results closely match those reported in the original paper, while some differences were observed, which could be attributed to various factors:

Comparison with Original Paper:

In the original paper, the DenseNet model consistently outperformed ResNet across various metrics, which is also reflected in the reproduced results. The reimplementation obtained a test accuracy of 84.37% for DenseNet and 82.49%

for ResNet, which is approximately the same if we compare the performance characteristics in the original study.

In this implementation, DenseNet achieved an AUC of 0.96, while ResNet reached an AUC of 0.93. The same comparison is also conducted for ResNet, with an AUC of 0.93 for the model. While these numbers are close to the 0.99 AUC acquired in the original paper, they confirm that DenseNet outperforms other models in terms of identifying images.

Possible Reasons for Differences:

Data Quality and Preprocessing: Differences in data preprocessing steps, such as the specific stain normalization technique or data augmentation methods used, could impact learning process.

Model Settings and Hyperparameters: The original paper may have used different hyperparameter settings, such as learning rates, batch sizes, or regularization techniques. While this reimplementation followed the principles outlined in the paper, slight variations in hyperparameter tuning or optimization strategies could lead to variations in model performance.

Early Stopping: In this reimplementation, early stopping was employed to prevent overfitting based on the validation loss. The timing of early stopping may have differed from the original study, potentially leading to fewer training epochs and thus slightly different results. For instance, early stopping occurred for DenseNet after epoch 2, which could have limited further improvements in accuracy.

Hardware and Computational Resources: The original study may have used more extensive computational resources, allowing for longer training times, more thorough hyperparameter optimization, or the use of larger batch sizes. These factors can influence the final model's performance, especially in fine-tuning stages.

4. Conclusions and Discussion

The reproduction of the automatic deep learning-based framework for histopathologic image classification yielded results that are consistent with the original paper's findings. Both the DenseNet and the ResNet gave high accuracy in the classification, with a slightly higher score on DenseNet compared with ResNet on most of the metrics. It scored 84.37% on the test accuracy and 0.96 on the AUC, while the ResNet gave 82.49% on the test accuracy and 0.93 on the AUC. As shown by the results, the techniques selected for the classification task, which included the use of transfer learning, data augmentation and early stopping, all contributed to making the models better informed by new data, resulting in increased recognition rates for new testing images. The results validate DenseNet's superior architecture for capturing complex patterns in histopathologic images, making it a more suitable model for this classification task.

Some limitations were encountered during the reproduction process, such as uncertainty about the specific stain normalization and data augmentation methods used in the original paper, as well as the hyperparameter settings, learning rate, batch size or regularization techniques, etc. These may affect model performance.

In order to improve the ability of automatic classification of histopathology images in the future, we can focus on exploring more advanced architectures, such as EfficientNet or Hybrid Models (CNN + Transformer) [6], which can provide better accuracy and generalization. Integrating complex data augmentation techniques such as Mixup or GAN-based methods [7] can further improve the robustness of the model. In addition, using model ensembles to combine outputs from different architectures may improve classification performance.

References

[1] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, M. Welling. "Rotation Equivariant CNNs for Digital Pathology". arXiv: 1806.03962

[2] Ehteshami Bejnordi et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA: The Journal of the American Medical Association, 318(22), 2199–2210. doi:jama.2017.14585

[3] Huang G, Chen D, Li T, et al. Multi-scale dense convolutional networks for efficient prediction [J]. arXiv preprint arXiv: 1703.09844, 2017, 2(2).

[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[5] Leslie N Smith. 2018. A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. arXiv preprint arXiv: 1803.09820 (2018).

[6] Li, X., Wang, W., Hu, X., Yang, J., & Lin, L. (2021). TransPath: Transformer-based self-supervised learning for histopathologic image classification. Medical Image Analysis, 73.

[7] Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F. K., & Mahmood, F. (2021). Synthetic data augmentation using GAN for improved liver histopathology classification. IEEE Transactions on Medical Imaging, 40(7), 1810-1821.