VGG16-Based Analysis of Lymph Node Biopsy Classification

1

1st Miaojin Hu
Department of Statistics
University of Michigan
Ann Arbor, United States
miaojinh@umich.edu

***Abstract*—This study uses a modified VGG16 model to classify lymph node biopsy images as benign or malignant. The dataset of 400 images was preprocessed and split for training and validation. The model achieved a validation accuracy of 91.25%, demonstrating its effectiveness in automated diagnosis.**

***Index Terms*—Lymph node biopsy, VGG16 Model, Image classification, Deep learning**

I. Introduction

Many cancers, such as breast cancer and lung cancer, are often associated with extensive lymph node metastases and the presence or absence of lymph node metastases plays a critical role in tumor staging, treatment planning, and predicting patient prognosis [3]. Digital pathology, which involves digitizing tissue images and slides, has transformed the field by enabling more efficient storage, visualization, and analysis of tissue samples. This digitization has the potential to streamline diagnostic workflows and improve the overall efficiency of routine pathological analysis [7].

Deep learning models have shown promise in achieving high accuracy in classification of cancer images [11]. Recent studies indicate deep learning models such as VGG16, VGG19, ResNet50, MobileNet, and Xception are commonly used methods in medical image analysis [11]. These models have been applied extensively across various domains, especially the VGG16 model [5]. For instance, Jiang et al. (2023) utilized VGG16 to evaluate tumor microenvironments and predict treatment responses in gastric cancer [5]. Bakasa and Viriri (2023) employed VGG16 as a feature extractor combined with an Extreme Gradient Boost Classifier for pancreas cancer prediction [9]. Similarly, Ye et al. (2022) incorporated VGG16 model to classify genomic pan-cancer [10]. These applications underscore VGG16's reliability and efficiency in handling complex medical imaging datasets.

The motivation for this project stems from the critical role lymph node metastases play in cancer diagnostics and the availability of pathology image datasets. I plan to utilize one of the convolutional neural network (CNN) [6], VGG16 and the Lymphnode Cancer Biopsy Dataset [2] [1] from Hugging Face to explore the potential CNNs in medical imaging. The Lymphnode Cancer Biopsy Dataset contains biopsy images of lymphnode cancer tissues, divided into two classes: benign and malignant. Each sample is stored in a separate image file, organized into respective class folders.

II. Method

A. Data Description

The problem is formulated as a supervised learning task where the input is an RGB image and the output is a binary classification label indicating whether the image is benign or malignant. The dataset contains 400 images in total, with 90% labeled as benign and 10% labeled as malignant. The first sample is shown in Fig. 1. To prepare the dataset, I resized all

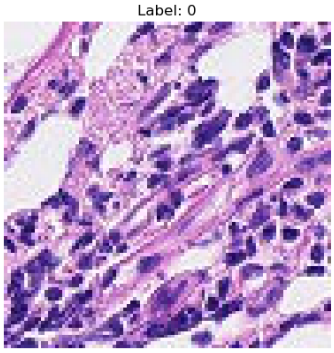

Fig. 1. First sample

images to a fixed size of 224×224 pixels. Each image was converted into a PyTorch tensor and normalized using the mean and standard deviation values of the ImageNet dataset. After preprocessing, the dataset was split into two parts: 80% was used for training, and the remaining 20% was used for validation. This split ensured that the model was trained on most of the data while keeping a portion aside to evaluate its performance.

B. VGG16 Model

To solve the classification problem, I used a modified version of the VGG16 model. The model's pre-trained weights

from ImageNet were used to take advantage of its strong feature extraction abilities. The first convolutional layer of the model was kept the same to handle 3-channel RGB images. The last fully connected layer was replaced by a new fully connected layer with two output neurons. These neurons represent the two classes: benign and malignant. This change ensures that the model outputs probabilities for each class, making it easier to interpret the results.

The model was trained using the Cross-Entropy Loss function, which measures how well the predicted probabilities match the true labels. To optimize the model, I used the Adam optimizer with an initial learning rate of 0.001 and a weight decay of 0.01. The weight decay helps prevent overfitting by regularizing the model parameters. To further avoid overfitting, I implemented early stopping. This technique stops the training process if the validation loss does not improve for five consecutive epochs.

During training, the model processed the data in two main steps: forward and backward passes. In the forward pass, the images were passed through the model to compute predictions. In the backward pass, the loss was used to compute gradients, which were then used to update the model's parameters. At the end of each epoch, the model's performance was evaluated on the validation set by calculating the validation loss and accuracy. The best-performing version of the model, based on validation loss, was saved for further evaluation.

Once the training was complete, the final model was tested on the validation set. The performance was measured by calculating the accuracy and analyzing the predictions to ensure the model's reliability in distinguishing between benign and malignant images.

III. Results

The training and validation loss curves over epochs demonstrated a consistent trend toward convergence. As shown in the training log, the training loss started at 0.7403 and decreased steadily to 0.4216 by epoch 13. Similarly, the validation loss began at 0.4079 and achieved a minimum of 0.3110 at epoch 8, after which it remained stable (Fig.2). This indicates that the model has gradually learned the features of the data and has effectively fitted the training data. Early stopping was triggered at epoch 13, as there was no significant improvement in validation loss for five consecutive epochs, ensuring the model did not overfit.

The validation accuracy remained stable throughout the training process, achieving a consistent value of 91.25% from the first epoch onward, which is a high value. This indicates that the model was able to reliably distinguish between benign and malignant images early in the training process, despite the class imbalance in the dataset (90% benign and 10% malignant). In addition, early stopping was triggered at epoch 13, ensuring the model did not overfit the training data while retaining its ability to generalize.

Therefore, The model saved at epoch 8 should be used as the final model since its stable accuracy and minimal validation loss.

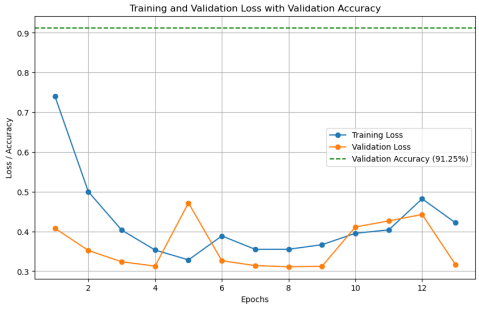

Fig. 2. Training and Validation Loss with Validation Accuracy

IV. Conclusion

This project showed that a modified VGG16 model can classify lymph node biopsy images as benign or malignant with high accuracy. Using pre-trained ImageNet weights and early stopping rule, the model achieved 91.25% validation accuracy, proving it to be reliable for medical image analysis.

References

[1] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling, "Rotation Equivariant CNNs for Digital Pathology," *arXiv preprint arXiv:1806.03962*, 2018.

[2] B. E. Bejnordi, M. Veta, P. J. van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, and the CAMELYON16 Consortium, "Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017, doi: 10.1001/jama.2017.14585.

[3] J. Ping, W. Liu, Z. Chen, and C. Li, "Lymph node metastases in breast cancer: Mechanisms and molecular imaging," *Clinical Imaging*, vol. 103, p. 109985, Nov. 2023, doi: 10.1016/j.clinimag.2023.109985.

[4] R. J. Chen, M. Y. Lu, D. F. K. Williamson, T. Y. Chen, J. Lipkova, Z. Noor, M. Shaban, M. Shady, M. Williams, B. Joo, and F. Mahmood, "Pan-cancer integrative histology-genomic analysis via multimodal deep learning," *Cancer Cell*, vol. 40, no. 8, pp. 865–878.e6, Aug. 2022, doi: 10.1016/j.ccell.2022.07.004.

[5] Y. Jiang, K. Zhou, Z. Sun, H. Wang, J. Xie, T. Zhang, S. Sang, M. T. Islam, J. Y. Wang, C. Chen, Q. Yuan, S. Xi, T. Li, Y. Xu, W. Xiong, W. Wang, G. Li, and R. Li, "Non-invasive tumor microenvironment evaluation and treatment response prediction in gastric cancer using deep learning radiomics," *Cell Reports Medicine*, vol. 4, no. 8, p. 101146, Aug. 2023, doi: 10.1016/j.xcrm.2023.101146.

[6] P. K. Kushwaha and M. Kumaresan, "Machine learning algorithm in healthcare system: A Review," in *Proceedings of the 2021 International Conference on Technological Advancements and Innovations (ICTAI)*, Tashkent, Uzbekistan, Nov. 2021, pp. 478–481.

[7] H. Zhang, W. Li, and H. Zhang, "An Image Recognition Framework for Oral Cancer Cells," *Journal of Healthcare Engineering*, vol. 2021, p. 2449128, 2021, doi: 10.1155/2021/2449128. [Retraction in: J. Healthc. Eng., Sep. 2023, doi: 10.1155/2023/9834098].

[8] T. Ye, S. Li, and Y. Zhang, "Genomic pan-cancer classification using image-based deep learning," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 835–846, Jan. 2021, doi: 10.1016/j.csbj.2021.01.010.

[9] W. Bakasa and S. Viriri, "VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction," *Journal of Imaging*, vol. 9, no. 7, p. 138, Jul. 2023, doi: 10.3390/jimaging9070138.

[10] L. Y. Ye, X. Y. Miao, W. S. Cai, and W. J. Xu, "Medical image diagnosis of prostate tumor based on PSP-Net+VGG16 deep learning network," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106770, Jun. 2022, doi: 10.1016/j.cmpb.2022.106770.

[11] N. Madusanka, P. Jayalath, D. Fernando, L. Yasakethu, and B. I. Lee, "Impact of H&E Stain Normalization on Deep Learning Models in Cancer Image Classification: Performance, Complexity, and Trade-Offs," *Cancers (Basel)*, vol. 15, no. 16, p. 4144, Aug. 2023, doi: 10.3390/cancers15164144.