DEEP LEARNING FOR MEDICAL IMAGING FINAL REPORT

CHIH-HAO LIAO
School of Forestry and Resource Conservation
Graduate Institute of Biomedical Electronics and Bioinformatics
National Taiwan University
Taipei, Taiwan
R11625015@ntu.edu.tw

Richard LEE LAI
Graduate Institute of Biomedical Electronics and Bioinformatics
National Taiwan University
Taipei, Taiwan
D12945021@ntu.edu.tw

OPENAI GPT-3.5*
OPENAI
San Francisco, CA, USA
contact@openai.com

***Contributed on grammar correction and translation**

June 6, 2024

ABSTRACT

In this study, we employ swin transformers to classify and determine the extent and existance of tumors on the chest, providing diagnostic feedback durign health inspections. The models are established by self trained and pretrained models, and the latter then further finetuned based on chest x-ray images provided better results. We show that for chest images, swin transformers offers excellent results, with an average accuracy of more than 85%, fine-tuned model reaches 94%. Concerning chest X-ray images, we attribute potential limitations primarily to interference stemming from the lung cavity, alongside the prominent arrangement of ribs. We aspire to further enhance our model by combine object detection before classification, enabling the automatic classificaion and labeling of the medical images and the doctors can easily figure out the target tumor cells for further process.

Keywords Swim Transformer · Pneumonia · Chest X-Ray

1 Introduction

On average, most patients need to wait for around two to three weeks for health inspection results. For those suspected of having cancer, such uncertainty can result in immense fear and anxiety. Real-time cancer detection during imaging procedures can provide immediate feedback, alleviating such negative emotions. However, the relatively high resolution of most diagnostic images makes this a difficult task. For example, digital X-ray images of breast cancer screenings often contain up to 100 MBs of information, which would require more an hour for image consolidation, image processing and tumor detection on a personal computer. This would interfere with standard operating procedures for health inspections, which are often pipelined to reduce potential bottlenecks. Therefore, we propose using the Swin transformer, a transformer-based deep learning model for inference.

2 Methods

2.1 Dataset

The dataset utilized in this study was obtained from Daniel [1], who collected and labeled the chest x-ray dataset. In brief, a total of 5,232 chest X-ray images from pediatric patients were collected, comprising 3,883 images depicting pneumonia images (2,538 bacterial and 1,345 viral) and 1,349 normal images 1. The images were sourced from from 5,856 patients for training purposes. These images were sourced from 5,856 patients for training purposes. Additionally, for testing, 624 patients contributed 234 normal images and 390 pneumonia images (242 bacterial and 148 viral). By merging with the original dataset [2], our dataset consisted of 5863 chest X-ray images with different resolutions, For the training, testing, and validation datasets, there were 1341, 234, and 8 images of normal patients and 3875, 390, and 8 images of pneumonia patients, respectively.

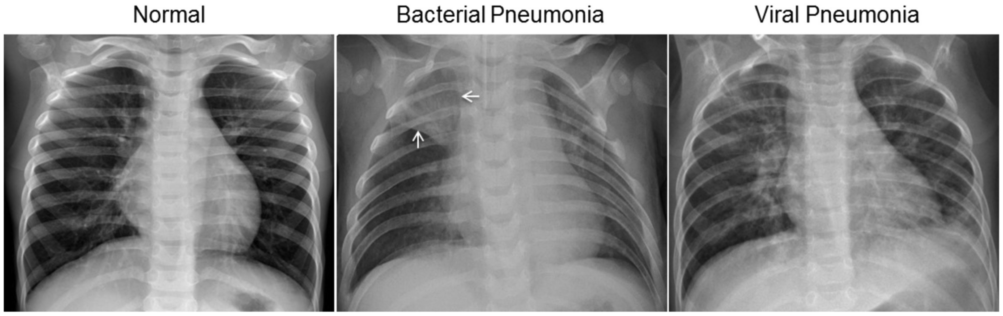


Figure 1: Illustrative Examples of Chest X-Rays in Patients with Pneumonia.

2.2 Model

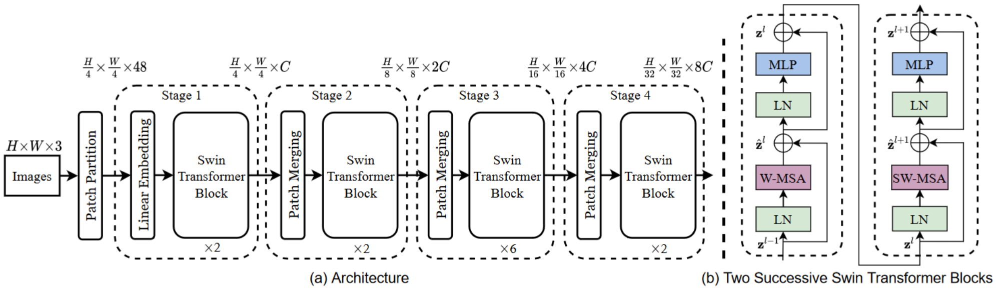
The Swin Transformer is a type of Vision Transformer designed for image classification and other vision tasks. It introduces a hierarchical architecture that processes images at multiple scales using shifted windows for efficient computation. This approach allows the model to handle high-resolution images and capture local context effectively. Compared with the original Vision Transformer, the Swin Transformer enables the model to capture more detailed information and achieve better performance on various vision benchmarks.

The Swin Transformer consists of four modules, each with a varying number of layers. The first module processes the input image at the original resolution, while the subsequent modules process the image at a reduced resolution. This hierarchical architecture enables the model to capture information at multiple scales and achieve state-of-the-art performance on various vision benchmarks. The Swin Transformer balances computational efficiency with high accuracy, making it suitable for a range of computer vision applications.

The hierarchical nature also enables the model to achieve state-of-the-art performance on various vision benchmarks, including image classification, object detection, and semantic segmentation. The Swin Transformer balances computational efficiency with high accuracy, making it suitable for a range of computer vision applications. In our study, the baseline model that we used is from Microsoft swin transformer v1 [3], and the model structure that we used from pytorch is swin transformer v2 base [4], and the pre-trained model that we used on Hugging Face is inherited from the Microsoft swinv2-tiny-patch4-window8-256.[4].

2.3 Preprocessing

For the training data, a series of transformations are applied to augment the dataset and improve model generalization. Firstly, images are resized to match the dimensions expected by the feature extractor, which is 256256. Then, a random horizontal flip is applied, horizontally flipping images randomly to augment the dataset with mirrored versions. Following this, images are randomly rotated up to 15 degrees to further diversify the dataset. Subsequently, the transformed images are converted into PyTorch tensors. Finally, normalization is performed using the mean and standard deviation of the image dataset provided by the feature extractor. For the validation and testing dataset, only resize and normalization are applied to them. The self-trained model is normalized with pre-trained weights of ImageNet, and the fine-tuned model is applied using the mean and standard deviation of the image dataset provided by the feature extractor.



(a) Architecture

(b) Two Successive Swin Transformer Blocks

Figure 2: Architecture of Swin (shifted window) transformer. The Swin transformer consists of four modules, each with a varying number of layers.

3 Results

The models undergo testing on the test dataset to assess their performance, with the outcomes revealing key classification metrics for each class beforehand. The findings are presented in Table 1 and Table 2 below.

3.1 Self trained model

Based on the evaluation metrics presented in Table 1, our self-trained model demonstrates promising performance in classifying chest X-ray images into two classes: NORMAL and PNEUMONIA. The precision, recall, and F1 score for each class provide insights into the model's ability to correctly identify instances of each class.

For the NORMAL class, the model achieves a precision of 99.3%. However, the recall for this class is lower at 61.5%, suggesting that the model misses identifying a considerable portion of actual NORMAL instances. Consequently, the F1-score, which balances precision and recall, is 76.0% for this class.

In contrast, for the PNEUMONIA class, the model exhibits a precision 81.2%, which means less instances predicted as PNEUMONIA are indeed NORMAL cases. The high recall of 99.7% indicates that the model effectively captures almost all true PNEUMONIA instances. Consequently, the F1-score for this class is notably higher at 89.5%, reflecting the model's balanced performance in identifying PNEUMONIA cases.

The overall accuracy indicating that it correctly classifies approximately 85.4% of the instances in the dataset. The macro average F1-score, which considers the balanced performance across both classes, is 82.8%, while the weighted average F1-score, which accounts for class imbalance, is 84.5%.

Table 1: Classification report for the self-trained model.

	Precision	Recall	F1-Score	Support
NORMAL	99.3%	61.5%	76.0%	234.0
PNEUMONIA	812%	99.7%	89.5%	390.0
Accuracy			85.4%	624.0
Macro avg	90.3%	80.6%	82.8%	624.0
Weighted avg	88.0%	85.4%	84.5%	624.0

3.2 Fine-tuned model from pre-trained model

Based on the results obtained from our fine-tuned model, we present a comprehensive evaluation of its performance in classifying pneumonia cases from chest X-ray images. The Table 2 illustrates the result.

Our model achieves an impressive precision of 97.1% for NORMAL cases, indicating its ability to correctly identify true negative instances. Moreover, with a recall of 85.9%, our model demonstrates a strong capacity to capture the majority of true positive pneumonia cases. The corresponding F1-score of 91.2% further solidifies the model's balanced performance in terms of precision and recall for NORMAL class prediction.

For the PNEUMONIA class, our model maintains a high precision of 92.1%, signifying its proficiency in correctly labeling positive instances. Additionally, a remarkable recall of 98.5% suggests the model's capability to effectively identify the vast majority of pneumonia cases present in the dataset. This is reflected in the F1-score of 95.2%, indicating a harmonious balance between precision and recall for PNEUMONIA class prediction.

The overall accuracy of our fine-tuned model is reported at 93.7%, underscoring its ability to make correct classifications across both classes. The macro-average F1-score of 93.2% and the weighted-average F1-score of 93.7% further validate the model's robust performance across all classes, considering their respective support values.

These results demonstrate the effectiveness of our fine-tuned model in accurately classifying pneumonia cases from chest X-ray images. The high precision and recall values attained for both classes underscore the model's potential for clinical application in assisting radiologists with accurate diagnosis and patient management. Moreover, the accuracy is better than the previous studies [1], which is 92.8%.

Table 2: Classification report for fine-tuned model based on pre-trained model.

	Precision	Recall	F1-Score	Support
NORMAL	97.1%	85.9%	91.2%	234.0
PNEUMONIA	92.1%	98.5%	95.2%	390.0
Accuracy			93.7%	624.0
Macro avg	94.6%	92.2%	93.2%	624.0
Weighted avg	94.0%	93.8%	93.7%	624.0

3.3 Confusion matrix

From these results below, including Figure 3 and 4, it's evident that the model performs well in terms of true positives and true negatives, indicating a good ability to distinguish between normal and pneumonia cases. However, the presence of false positives and false negatives suggests that there is room for improvement, particularly in reducing false positives to avoid unnecessary concern or treatment.

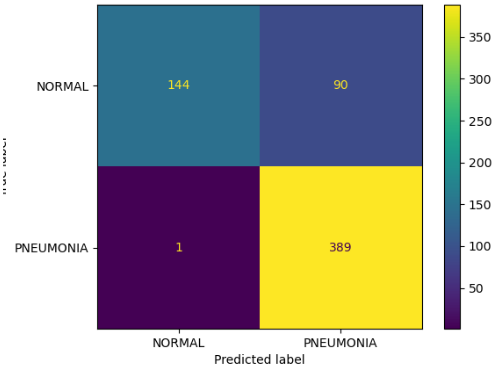


Figure 3: The confusion matrix of self trained model.

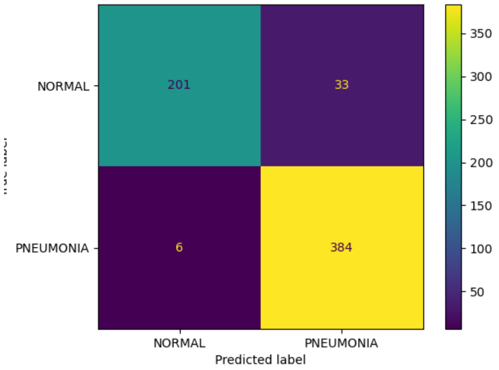


Figure 4: The confusion matrix of the fine-tuned model based on pre-trained model.

To assess the model's performance, we showcase several sample images from the test set in Figure 5. From the illustrations, it's evident that the model adeptly identifies pneumonia in the X-ray images with remarkable precision. The instances of misclassification are minimal, indicating a commendable outcome for the model's predictive capabilities.

4 Conclusion

In this study, we not only established a Swim Transformer model for detecting Pneumonia in Chest X-Ray images but also enhanced its accuracy to up to 93.7% through fine-tuning a pretrained model available on Hugging Face. This not only assists future physicians in identifying pneumonia in X-ray images but also enables the precise localization of bacterial and viral pneumonia using the object detection method of Swim Transformer in the future.

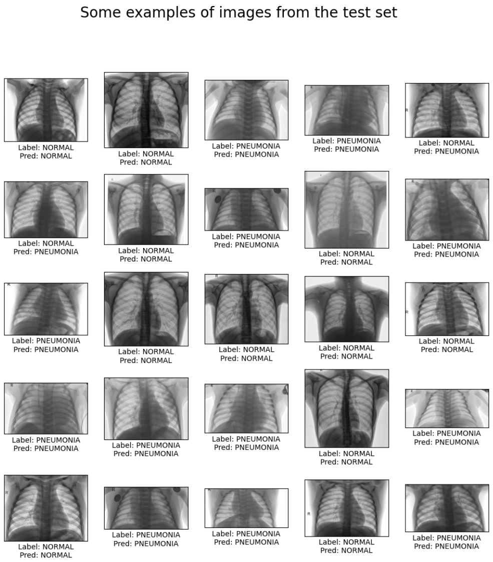


Figure 5: Some examples of images from the test set.

References

[1] Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C.S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, Justin Dong, Made K. Prasadha, Jacqueline Pei, Magdalene Y.L. Ting, Jie Zhu, Christina Li, Sierra Hewett, Jason Dong, Ian Ziyar, Alexander Shi, Runze Zhang, Lianghong Zheng, Rui Hou, William Shi, Xin Fu, Yaou Duan, Viet A.N. Huu, Cindy Wen, Edward D. Zhang, Charlotte L. Zhang, Oulan Li, Xiaobo Wang, Michael A. Singer, Xiaodong Sun, Jie Xu, Ali Tafreshi, M. Anthony Lewis, Huimin Xia, and Kang Zhang. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131.e9, 2018.

[2] Daniel Kermany. Labeled optical coherence tomography (oct) and chest x-ray images for classification, 2018.

[3] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows, 2021.

[4] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, Furu Wei, and Baining Guo. Swin transformer v2: Scaling up capacity and resolution, 2022.

5 Supplementary Material

5.1 Data availability

All the data used in this analysis are publicly available, as indicated in the data tables and figures in the main text and the supplementary material. All the data used in this analysis is openly available at the following link in the GitHub repository: `https://github.com/CodeamonCat/2024_spring_deep_learning_for_medical_imaging_final`.

5.2 Appendix

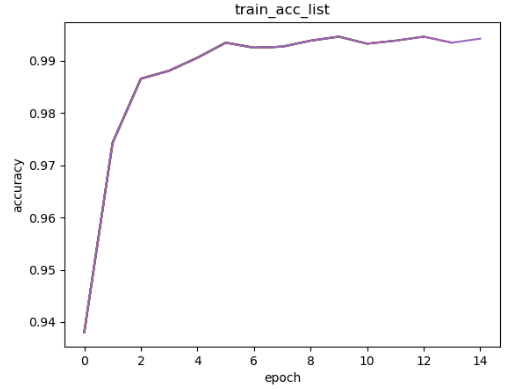


Figure 6: Accuracy curve of self-trained model during training.

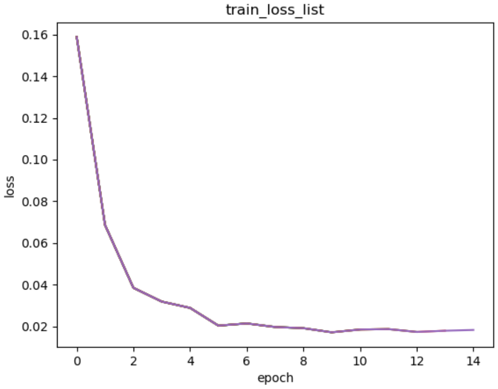


Figure 7: Loss curve of self-trained model during training.

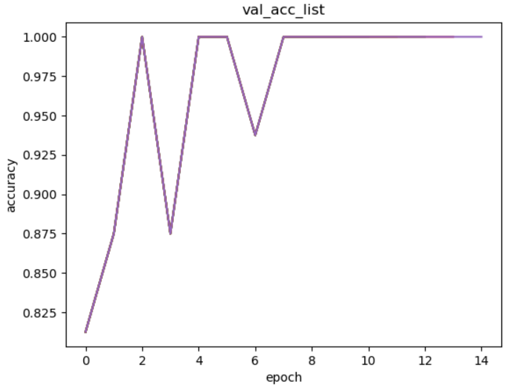


Figure 8: Accuracy curve of self-trained model during validation.

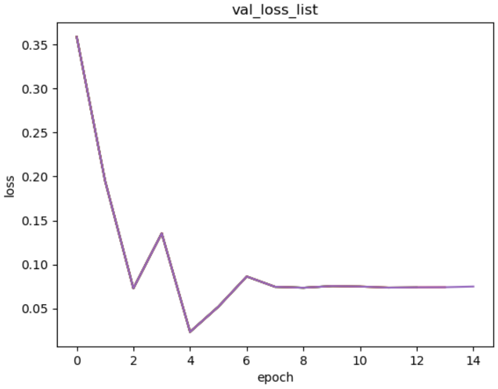


Figure 9: Loss curve of self-trained model during validation.