LaTeX Guidelines for Author Response

A. Introduction and Problem Statement

Applying CNNs and transfer models across different datasets enhances efficiency by leveraging learned features and promoting model generalization in diverse image classification tasks. This project is centered on optimizing CNN models as a powerful deep learning model optimized for accurate image recognition and classification to overcome challenges associated with large datasets and subtle variations. The optimization process fine-tunes hyperparameters, prioritizing simpler architectures like ResNet-50. The iterative training involves processing the dataset through the CNN, adjusting weights based on loss, and assessing convergence. Detailed documentation is essential for a comprehensive understanding of performance. Additionally, t-distributed stochastic neighbor embedding (t-SNE) is used for dimensionality reduction and visualizing CNN encoder outputs. Challenges include managing large datasets and subtle image variations, while expected results involve optimized CNN models for accurate image patch classification. The project aims to gain insights into features, comprehend architecture impacts, and apply findings to broader image classification challenges.

B. Methodologies and Datasets

The utilization of ResNet-50, a deep convolutional neural network (CNN) model, is proposed for the classification of histopathological images of colorectal cancer, prostate cancer, and animal faces. ResNet-50 is a state-of-the-art CNN model known for its high performance on various image recognition tasks, exemplified by its success on ImageNet. Comprising 50 layers of residual blocks, ResNet-15 employs skip connections to mitigate the issue of vanishing gradients and enhance the network's learning capacity.

Three datasets, with reduced number of samples and classes, used for training and testing the model: Dataset 1, Colorectal Cancer - it includes patches with three classes out of eight in original dataset. Techniques such as rotation, flipping, cropping, and normalization enhance diversity and balance.[1] Dataset 2, Prostate Cancer - it mirrors Dataset 1's three classes. Similar data augmentation and normalization methods are applied.[2] Dataset 3, Animal Faces - this dataset features three classes: cats, dogs, and wildlife animals. Employing the same augmentation and normalization techniques as in previous Datasets.[3]

Performance will be evaluated using the following metrics on each dataset: accuracy, precision, recall, F1-score, and confusion matrices.

Research Questions and Hypotheses:

Research Question 1 (RQ1): The study evaluates the efficacy of ResNet-50 in classifying histopathological images of colorectal and prostate cancer. Hypothesis 1 (H1): ResNet-50 is expected to outperform other CNN models, demonstrating higher accuracy, precision, recall, and F1-score on both cancer datasets.

Table 1. Datasets Statistics

Dataset	Field	Original Images	Original Classes	Project Images	Project Classes
1	Colorectal Cancer	100k	8	6k	3
2	Prostate Cancer	120k	3	6k	3
3	Animal Faces	16k	3	6k	3

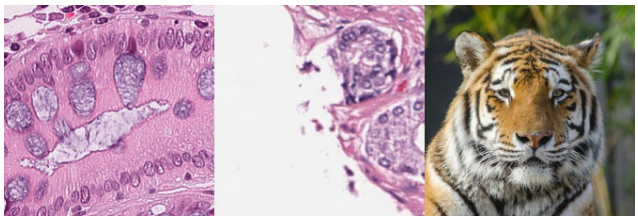

Figure 1. sample image in dataset 1 (left), 2(middle) and 3 (right)

Research Question 2 (RQ2): The investigation assesses the performance of ResNet-50 in classifying images of animal faces. Hypothesis 2 (H2): ResNet-50 is anticipated to exhibit superior performance (accuracy, precision, recall, and F1-score) on the animal faces dataset compared to other CNN models.

Research Question 3 (RQ3): The study explores variations in ResNet-50's performance across different datasets and classes. Hypothesis 3 (H3): ResNet-50 is expected to perform better on the animal faces dataset due to higher contrast and image diversity. Performance variations across different classes are also anticipated based on image similarity and complexity.

These hypotheses will be tested by training and testing ResNet-50 on each dataset and comparing the results with other CNN models, such as VGG-16, Inception-V3, and DenseNet-121. Additionally, confusion matrices will be analyzed to identify sources of errors and misclassifications.

C. Attempts at Solving the Problem: Results

A ResNet-50[4] model was trained from scratch with our dataset. Subsequently, the hyperparameters were tuned using grid search, and the best configuration of the learning rate, batch size, and cost function was found. Next, a pre-trained ResNet-50 model from ImageNet was used and tested on our dataset. Finally, t-SNE[5] was used to understand the data distribution.

For Task 2, the performance of the model we trained

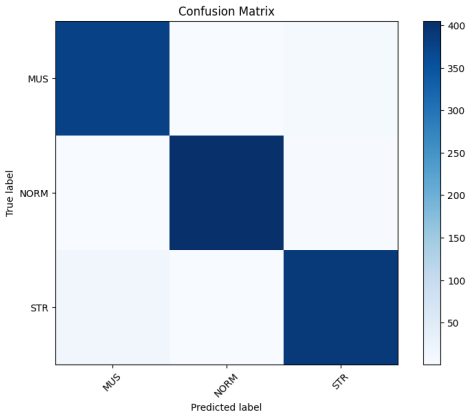

Figure 2. Confusion Matrix for Task 1

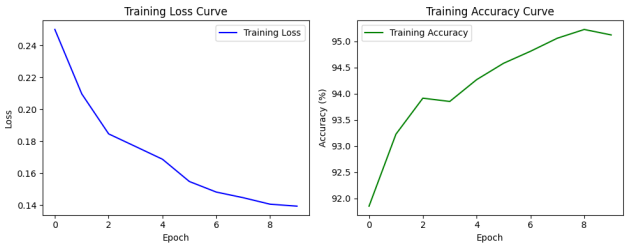

Figure 3. Loss and Accuracy for Task 1

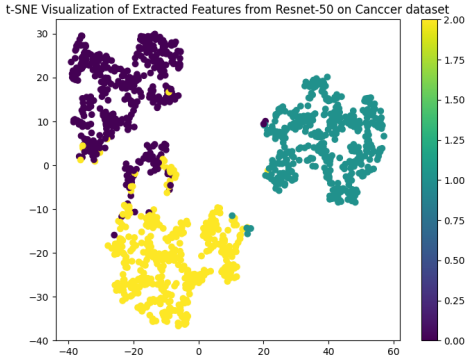

Figure 4. Colorectal Cancer t-SNE

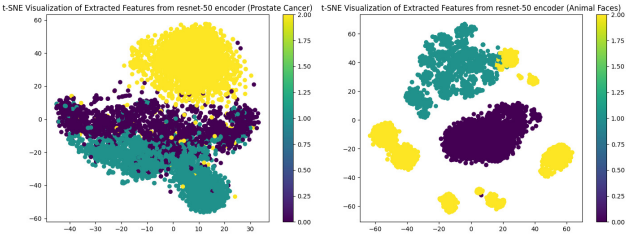

Figure 5. t-SNE of Prostate Cancer (left) and Animal Faces (right)

and the pretrained model was evaluated on datasets 2 and 3. Then, ML models were trained to classify the extracted features by the ResNet encoders. Two ML techniques were used: k-nearest neighbors clustering (unsupervised learning)[6] and random forest (supervised learning)[7]. After the grid search, the best configuration of parameters was found. The loss and accuracy plot for the ResNet-50 model trained from scratch was analyzed, and the confusion matrix, recall, and F1 score were computed.

```
Parameters: {   'batch_size': 64,
                'loss_func':
                'NLLLoss',
                'lr': 0.001,
                'momentum': 0.9}

Training Loss: -835540.9878571428
Training Accuracy: -889961.9041666667
Validation Loss: 0.3354761904761905
Validation Accuracy: 0.33111111111111113

Best parameters: {  'batch_size': 64,
                    'loss_func':
                    'CrossEntropyLoss',
                    'lr': 0.01,
                    'momentum': 0.9}
```

Challenges included attempts with batch sizes of 16 and 32, but limited memory led to a reduction to a batch size of 8, doubling the training time. Using more than 2 workloads caused ”End Of File” and ”worker error” issues, requiring adjustment to 2 workloads. Initially considering CPU utilization proved inadequate due to the high parameter count, leading to a decision to switch to GPU usage.

D. Future Steps

Based on the obtained results, it can be observed that the t-SNE clusters related to prostate cancer are not well separated. Improvements are planned for these clusters in the future. Furthermore, there is a need to enhance our overall performance, such as by implementing measures to save the model, thereby eliminating the need for repetitive training.

References

1. Kather, Jakob Nikolas; Halama, Niels; Marx, Alexander. (2018). 100,000 histological images of human colorectal cancer and healthy tissue (v0.1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1214456

2. Tolkach, Yuri. (2021). Datasets Digital Pathology and Artifacts, Part 1. https://doi.org/10.5281/zenodo.4789576

3. Choi, Yunjey, Uh, Youngjung, Yoo, Jaejun, Ha, Jung-Woo. (2020). StarGAN v2: Diverse Image Synthesis for Multiple Domains. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. https://www.kaggle.com/datasets/andrewmvd/animal-faces

4. He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://arxiv.org/abs/1512.03385

5. Van der Maaten, L., Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11). https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbcl

6. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. (2003). KNN model-based approach in classification. In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings (pp. 986-996). Springer Berlin Heidelberg. https://link.springer.com/chapter/10.1007/978-3-540-39964-3_62

7. Svetnik, Vladimir, et al. "Random forest: a classification and regression tool for compound classification and QSAR modeling." Journal of chemical information and computer sciences 43.6 (2003): 1947-1958. https://pubs.acs.org/doi/abs/10.1021/ci034160g