Classification Of Histopathological Images Of Oral Cancer Using Deep Learning Techniques

Shriram Ashok Birajdar
Malaviya National Institute of Technology, Jaipur
Student ID: 202111078
Email: 20211178@diu.iiitvadodara.ac.in

***Abstract*—This report presents the work undertaken during a summer internship focused on the classification of histopathological images of oral cancer using advanced deep learning techniques. The internship involved multiple key tasks: the implementation of various convolutional neural network (CNN) models, the development of vision transformer (ViT) and transformer models, and the integration of attention mechanisms within these models to enhance performance. A comprehensive comparative analysis of all models was conducted, examining metrics such as balanced accuracy, precision, recall, and F1-score. The project utilized the newly created P-NDB-UFES dataset, comprising 3,763 histopathological images categorized into three classes: oral squamous cell carcinoma (29%), dysplasia (51.29%), and without dysplasia (18.79%). This work demonstrates the potential of deep learning models in automating the diagnosis of oral cancer and highlights the importance of model selection and optimization in achieving high classification accuracy.**

I. Introduction

The diagnosis of oral squamous cell carcinoma (OSCC) and oral leukoplakia, both with and without dysplasia, has traditionally relied on the expertise of pathologists examining histopathological images. This manual diagnostic process is inherently time-consuming and can vary significantly between different pathologists, leading to potential discrepancies in diagnosis. With recent advancements in deep learning and computer vision, there is a significant opportunity to automate the detection and classification of these conditions using digital images, thereby improving diagnostic accuracy and efficiency.

Oral cancer, predominantly in the form of OSCC, represents a critical global health concern, ranking among the most prevalent cancers worldwide. Key risk factors contributing to its incidence include tobacco and alcohol use, sun exposure, and HPV infection. Unfortunately, many cases are diagnosed at advanced stages due to the lack of early screening and awareness. Consequently, early and accurate diagnosis is vital for enhancing treatment outcomes and improving survival rates among patients.

The primary objective of this internship project was to leverage deep learning techniques to develop an automated system capable of classifying histopathological images of oral cancer. A significant challenge in this domain is the limited availability of public datasets, which often hinders research and model training. To address this limitation, the project involved the creation and annotation of a novel dataset, the P-NDB-UFES, which comprises 3,763 histopathological images categorized into three distinct classes: OSCC, dysplasia, and images without dysplasia.

The internship comprised distinct project tasks, each addressing a specific challenge:

- Implementation of a variety of CNN models for classification of histopathological images
- Implementation of several Vision Transformers for classification
- Incorporation of various attention mechanisms to enhance model performance
- Comparative analysis of multiple models to evaluate performance

This project highlights the potential of deep learning models to automate and improve the accuracy of oral cancer diagnoses, offering a valuable tool to assist pathologists in clinical settings.

II. Task Assigned

During the internship, the primary responsibilities included:

- **Data Collection and Annotation:** Collaboration with pathologists was conducted to gather and annotate a comprehensive dataset consisting of 3,763 histopathological images. This task involved ensuring the accurate representation of various classes, including oral squamous cell carcinoma (OSCC), dysplasia, and images without dysplasia. The annotations provided essential information for the effective training of deep learning models.
- **Data Processing:** A systematic data processing approach was implemented following data collection. This included examining image quality, removing low-quality or misclassified images, and separating the annotated images into distinct directories according to their respective classes. Data augmentation techniques, such as rotation, flipping, and scaling, were also applied to enhance the dataset's diversity, improving the robustness of the models during training.
- **Model Implementation:** Various deep learning models for image classification were implemented and fine-tuned. This included the development of multiple convolutional neural network (CNN) architectures and the leveraging of vision transformers (ViT) and other advanced models. Each model was optimized through hyperparameter tuning and training strategies to achieve optimal performance in classifying histopathological images.

- **Incorporation of Attention Mechanisms:** Several attention mechanisms, including Squeeze-and-Excitation (SE) blocks, Convolutional Block Attention Module (CBAM), Bottleneck Attention Module (BAM), and global context attention, were integrated to enhance model performance. These mechanisms enabled the models to focus on the most relevant features within the images, significantly improving classification accuracy.
- **Experimental Analysis:** A comprehensive performance analysis of all implemented models was conducted using various metrics, including balanced accuracy (BCC), accuracy (ACC), precision, recall, and F1-score. This comparative analysis was crucial for identifying the strengths and weaknesses of each model, guiding further refinements and helping to select the most effective architecture for the classification task.

III. Work Carried Out

A. Team Contributions

The team worked collaboratively to ensure the success of the project, with contributions spanning multiple facets:

- **Data Preparation:** Collaborated with pathologists to meticulously annotate a dataset comprising 3,763 histopathological images. These images were categorized into three distinct classes: oral squamous cell carcinoma (OSCC), dysplasia, and without dysplasia, ensuring a balanced representation for effective model training.
- **Model Development:** Developed and evaluated a variety of deep learning models, including convolutional neural networks (CNNs) and transformer-based architectures, to assess their effectiveness in classifying the annotated images.

B. Individual Contributions

Key individual contributions included:

- **Data Processing:** Conducted comprehensive data preprocessing and augmentation. This involved resizing images to 224×224 pixels, applying random rotations, and flipping images to enhance the diversity of the training dataset, ensuring robust training conditions.
- **Model Implementation and Evaluation:** Focused on implementing and assessing various deep learning models, with a particular emphasis on the ConvNeXt Tiny model, which achieved a balanced accuracy of 93% using Squeeze-and-Excitation (SE) attention. Developed performance evaluation metrics, including balanced accuracy (BCC), accuracy (ACC), precision, recall, and F1-score to gauge the effectiveness of the models.
- **Training and Validation:** Executed the training process for the ConvNeXt Tiny model, utilizing class-weighted loss functions to address class imbalance. Implemented learning rate scheduling techniques to optimize training efficiency and improve convergence.

C. Model Implementation

The implementation of the ConvNeXt Tiny model with Squeeze-and-Excitation (SE) block involved the following key steps:

- **Data Preparation:** Developed a custom dataset class for loading and augmenting images, ensuring robust training through stratified splits into training and validation sets.
- **Model Architecture:** Utilized the pre-trained ConvNeXt Tiny model from the timm library, modifying the final layer for classification and integrating Squeeze-and-Excitation (SE) blocks to enhance feature representation through attention mechanisms.
- **Training Configuration:** Unfroze all layers for comprehensive training, employing the Adam optimizer and a learning rate scheduler to optimize training.
- **Evaluation:** Assessed model performance using balanced accuracy, precision, recall, and F1-score, while generating a classification report to evaluate strengths and weaknesses.

The ConvNeXt Tiny model achieved a highest balanced accuracy of 93% on the validation set, demonstrating its effectiveness in classifying histopathological images.

D. Comparative Analysis

A detailed comparative analysis of various models, including CNNs, ViT, and transformer architectures, was conducted to evaluate their performance in classifying histopathological images:

1) Convolutional Neural Networks (CNNs): CNNs are known for their strong performance in image classification tasks due to their hierarchical feature extraction capabilities. Key models analyzed include:

- **ResNet50:** Utilized residual connections to prevent gradient vanishing, achieving balanced accuracy of 92.3%.
- **EfficientNet-B0:** Optimized for model size and accuracy, resulting in a balanced accuracy of 92.0%.
- **DenseNet121:** Employed dense connections to improve gradient flow, yielding a balanced accuracy of 91.1%.
- **MobileNetV3:** Lightweight architecture suitable for mobile applications, with a balanced accuracy of 91.3%.

2) Vision Transformers (ViT): ViT models leverage self-attention mechanisms to capture long-range dependencies in images. Notable models include:

- **ViT Small (Patch16-384):** Achieved a balanced accuracy of 92.8%, showcasing the potential of transformer-based architectures in image classification.
- **DeiT Tiny (Patch16-224):** Combined distillation and transformer architectures, resulting in a balanced accuracy of 92.7%.

3) Attention Mechanisms in ConvNeXt: Models incorporating advanced attention mechanisms, such as Squeeze-and-Excitation and Bottleneck Attention Module, demonstrated significant improvements:

- **ConvNeXt Tiny (with SE):** Achieved a remarkable balanced accuracy of 93%, highlighting the effectiveness of SE attention in enhancing model performance.
- **ConvNeXt Tiny (with BAM):** Achieved a balanced accuracy of 92.8%, demonstrating the efficiency of the Bottleneck Attention Module in improving feature representation and classification performance.

E. Model Performance Summary

Performance metrics for the various models tested during the internship are summarized in Table I.

TABLE I
MODEL PERFORMANCE SUMMARY

Model Names	BCC	ACC	Precision	Recall	F1
maxvit_tiny_tf_224	92.4	90.0	89.9	90.0	90.0
efficientnet_b0	92.0	89.7	87.9	89.7	88.7
resnet50	92.3	89.7	89.3	89.7	89.5
vit_small_patch16_384	92.8	90.3	90.7	90.3	90.5
densenet121	91.1	88.5	85.7	88.5	86.9
mobilenetv3_large_100	91.3	88.8	87.1	88.8	87.8
regnety_016	91.2	88.5	88.2	88.5	88.3
xception65	91.7	89.2	87.3	89.2	88.1
inception_v3	88.3	84.6	85.6	84.6	85.0
convnext_tiny	93.0	92.2	92.8	91.2	92.1
dm_nfnet_f0	91.8	89.3	87.7	89.3	88.5
deit_tiny_patch16_224	92.7	90.1	92.0	91.0	91.5
cait_m36_384	91.7	88.9	90.8	88.9	89.8
swin_s3_tiny_224	92.9	90.7	91.4	90.7	91.7
focalnet_tiny_lrf	92.9	90.5	91.2	90.5	90.8

This structured and detailed approach ensured thorough evaluation and optimization of the models, culminating in the successful classification of histopathological images with high accuracy.

IV. EXPERIENCE

This internship provided invaluable insights into the practical applications of deep learning techniques in medical imaging. Key learnings included:

- **Data Quality and Annotation:** I learned the critical importance of high-quality data and precise annotations for training robust models. Collaborating with pathologists highlighted the nuances involved in accurately labeling histopathological images.
- **Deep Learning Architectures:** Gained a deeper understanding of the comparative strengths and weaknesses of various deep learning architectures, including CNNs and transformers, and their impact on classification performance.
- **Deployment Challenges:** Experienced firsthand the practical challenges of deploying deep learning models in a clinical setting, emphasizing the need for thorough validation and model interpretability.
- **Research Methodology:** Acquired knowledge in research methodology, from literature review to data collection and analysis, which is crucial for academic pursuits in this field.

V. FUTURE PROSPECTS

The project holds significant potential for future research and application, including:

- **Enhanced Model Performance:** Future work could explore advanced models and optimization techniques to improve classification accuracy further. Investigating ensemble methods may also lead to better generalization across different datasets.
- **Dataset Expansion:** Increasing the size and diversity of the dataset can result in more generalizable models, enhancing their applicability in real-world scenarios.
- **Writing Research Papers:** Preparing a research paper to document the findings and methodologies of this internship project will be crucial for sharing knowledge with the academic community. This paper could focus on the innovative approaches used in data processing, model implementation, and evaluation.
- **Clinical Integration:** The ultimate goal is to integrate these models into clinical practice, assisting pathologists in accurate and timely diagnosis.

VI. CONCLUSION

This internship has been an invaluable experience, providing a robust platform to apply and expand my knowledge in deep learning and medical image analysis. The collaborative nature of the project significantly enhanced my learning process and facilitated effective problem-solving. Under the expert guidance of Dr. Deepak Ranjan Nayak, we successfully achieved our research objectives. The insights gained and the work accomplished during this internship have laid a strong foundation for future developments in automated diagnostics and further research in the field of medical image analysis.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to my supervisors and colleagues at MNIT Jaipur for their unwavering support and guidance throughout the internship. Special thanks to Dr. Deepak Ranjan Nayak for his mentorship, which significantly contributed to my learning experience, and to the pathologists who provided their expertise and insights, which were instrumental in the success of the project.

REFERENCES

[1] B. M. S. Maia, M. C. F. R. de Assis, L. M. de Lima, M. B. Rocha, H. G. Calente, M. L. A. Correa, D. R. Camisasca, and R. A. Krohling, ”Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer,” 2024. [Online].

[2] A. Vaswani, N. Shard, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, ”Attention is all you need,” in *Advances in Neural Information Processing Systems*, pp. 5998-6008, 2017.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ”An image is worth 16x16 words: Transformers for image recognition at scale,” 2020. [Online]. Available: https://arxiv.org/abs/2010.11929.