Breast Cancer Diagnosis Using Hybrid Deep Learning with CNN, Vision Transformers, and Self-Supervised Learning for Histopathology Imaging

1st Manasvi Khandelwal
2022UCA1888
Netaji Subhas University of Technology
Delhi, India

2nd Savvya Dahiya
2022UCA1906
Netaji Subhas University of Technology
Delhi, India

3rd Riya Sherawat
2022UCA1923
Netaji Subhas University of Technology
Delhi, India

Mr. Gaurav Singal
Professor, Computer Science and Engineering
Netaji Subhas University of Technology
Delhi, India

***Abstract*—This paper proposes a hybrid deep learning model for breast cancer diagnosis using Convolutional Neural Networks (CNN), Vision Transformers (ViT), and Self-Supervised Learning (SSL) for histopathology image classification. The model leverages CNNs for feature extraction, ViT for improved spatial understanding, and SSL to reduce the dependency on labelled data, all while achieving high diagnostic accuracy**

I. Introduction

Breast cancer is the most common cancer worldwide, accounting for a significant percentage of cancer-related deaths. Early and accurate diagnosis is critical for improving patient outcomes. Traditional histopathology-based cancer diagnosis relies heavily on the manual examination of tissue slides by pathologists, a time-consuming and error-prone process. The growing volume of histopathological data further strains diagnostic efficiency, making it essential to develop automated solutions.

Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable success in image classification tasks. However, their performance is sometimes limited by their inability to capture sequential or contextual relationships in image data. To address this, our project proposes a hybrid deep learning approach combining CNNs, Vision Transformers (ViT), and Self-Supervised Learning (SSL) techniques to enhance accuracy, efficiency, and interpretability.

The project aims to build a robust model capable of accurately classifying breast cancer subtypes from histopathology images, providing valuable insights for pathologists and aiding in early diagnosis and treatment planning.

No external funding was received for this research.

A. Introduction to Domain

Breast cancer is one of the most common types of cancer affecting women worldwide. Accurate early diagnosis is crucial for improving survival rates. Histopathological images, typically examined by pathologists, provide critical information for cancer diagnosis. However, the manual analysis of such images is time-consuming, and the volume of data makes it difficult for pathologists to detect all patterns accurately.

B. Problem Statement

Manual analysis of breast cancer histopathology images is prone to errors and lacks scalability. As a result, automated diagnostic systems leveraging deep learning models are essential for improving the accuracy and efficiency of breast cancer diagnosis.

C. Objectives

- Develop a hybrid deep learning model to improve the accuracy and interpretability of breast cancer subtype classification.
- Integrate Convolutional Neural Networks (CNNs), Vision Transformers (ViT), and Self-Supervised Learning (SSL) techniques to address challenges such as insufficient labeled data, image complexity, and diagnostic speed.
- Improve model performance by using CNNs for fine-grained feature extraction and ViT for capturing global spatial relationships within histopathology images.
- Reduce the dependency on large labeled datasets by incorporating Self-Supervised Learning for pretraining the model on unlabeled data.
- Enhance classification accuracy, precision, recall, and F1 score to ensure reliable breast cancer subtype classification.

- Prioritize model interpretability, ensuring that pathologists can trust and understand model decisions for real-world clinical applications.

D. Contributions

This paper presents the following contributions:

- A hybrid CNN-InceptionV3 model for automated classification of breast cancer histopathology images.
- The integration of Vision Transformers for enhanced spatial and contextual feature extraction.
- Use of Self-Supervised Learning to improve model performance using limited labeled data.
- Benchmarking the hybrid model against state-of-the-art CNN models.

E. Paper Organization

The rest of the paper is organized as follows:

- Section 2 reviews related work and background studies.
- Section 3 presents the methodology and the approach used in the model.
- Section 4 provides the experimental setup, results, and analysis.
- Section 5 concludes the paper and suggests directions for future work.

II. Literature/Related Work

Deep learning models, particularly CNNs, have been widely used for breast cancer classification, excelling at spatial feature extraction. However, these models often miss contextual and sequential relationships in histopathology images. This paper proposes a hybrid model combining InceptionV3 for advanced feature extraction, Vision Transformers (ViT) for capturing long-range dependencies, and Self-Supervised Learning (SSL) to reduce reliance on large labeled datasets.

InceptionV3 is known for its ability to capture hierarchical features in complex medical images, while ViT uses self-attention mechanisms to focus on both local and global spatial patterns, improving classification accuracy. SSL further enhances the model's performance by learning meaningful representations from unlabeled data, making it more efficient and robust. By integrating these techniques, the proposed model offers improved accuracy and interpretability, addressing the limitations of traditional CNN-based approaches and providing a more effective solution for breast cancer diagnosis.

A. Background Study

In the past decade, deep learning models have revolutionized medical image analysis. CNNs have been widely used for their ability to extract hierarchical features from images. Vision Transformers, with their self-attention mechanisms, have shown promise in capturing complex patterns in large datasets, while Self-Supervised Learning reduces the need for extensive labeled data, making models more scalable.

III. Methodology

A. Data Preparation

- Dataset: BreakHis dataset.
- Preprocessing: Image resizing, normalization, and augmentation.
- Dataset Split: Training, validation, and testing.

B. Model Development

- InceptionV3 used for efficient and deep spatial feature extraction.
- CNN layers added to refine local patterns.
- Vision Transformers (ViT) applied to capture global spatial dependencies using self-attention mechanisms.
- SSL applied for pretraining on unlabeled data using contrastive learning.
- Transfer learning and fine-tuning on labeled datasets for final classification.

C. Training Strategy

- Supervised fine-tuning of pretrained networks.
- Optimizers: Adam and SGD with learning rate scheduling.
- Regularization techniques: dropout, batch normalization.
- Evaluation metrics: accuracy, precision, recall, and F1-score.

D. Approach

- Images are preprocessed and passed through InceptionV3.
- Features are enhanced using ViT.
- SSL is used during pretraining to improve learning.

E. Algorithm

A hybrid approach that combines CNN with ViT and SSL. Fine-tuning is done using transfer learning.

F. Hardware/Software

- Hardware
  - The model was trained on high-performance workstations with powerful processors and sufficient RAM. No GPUs were used due to resource limitations.
- Software
  - Frameworks: TensorFlow and Keras.
  - Programming Language: Python.
  - Libraries: NumPy, Pandas, Matplotlib, OpenCV, PIL.
  - Development Environment: Jupyter Notebook, PyCharm.
- Dataset Used
  - BreakHis

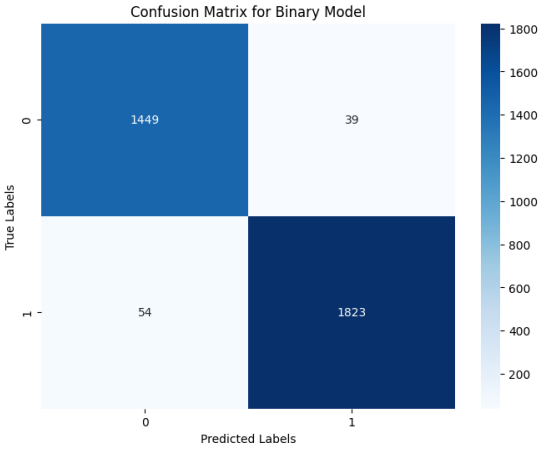
IV. Result Analysis

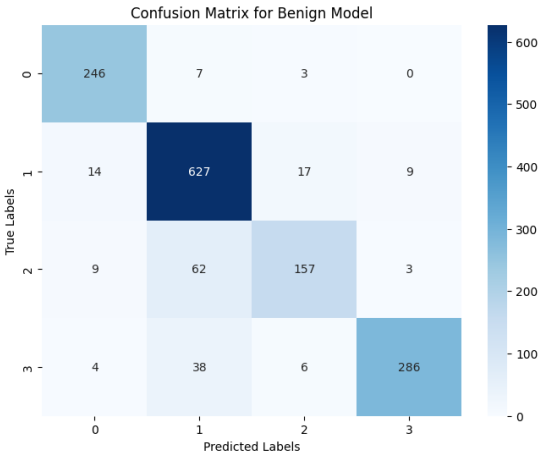
A. Experimental Setup

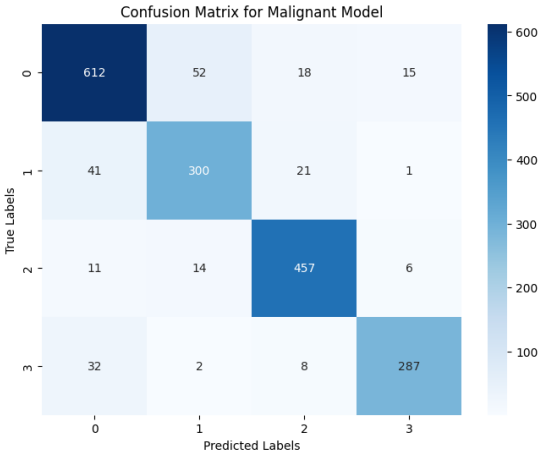
- The model was trained using a hybrid deep learning architecture combining InceptionV3, Vision Transformers (ViT), and Self-Supervised Learning (SSL).
- The dataset used was the BreakHis histopathology image dataset, consisting of benign and malignant subtypes.
- Standard preprocessing was applied including resizing, normalization, and extensive data augmentation.
- The training was conducted using GPU acceleration in Google Colab with optimizers such as Adam and learning rate scheduling for convergence.
- Evaluation was done using metrics such as Precision, Recall, and F1-Score for both binary and multi-class classification.

B. Result Graph

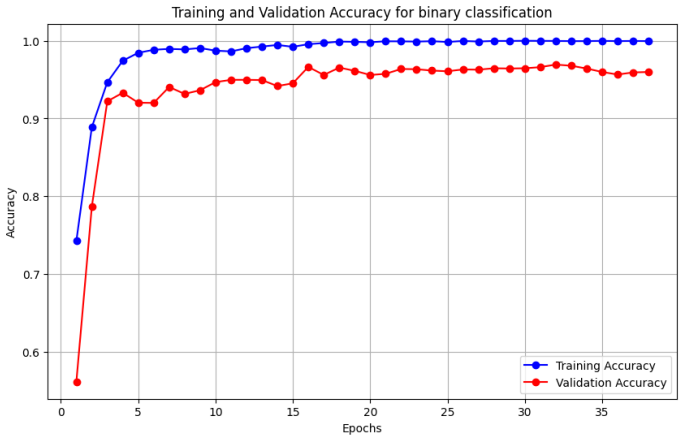
- Below are the visualizations representing the model's classification performance:
  - **Confusion Matrix:** Illustrates true positives, false positives, false negatives, and true negatives for each class.

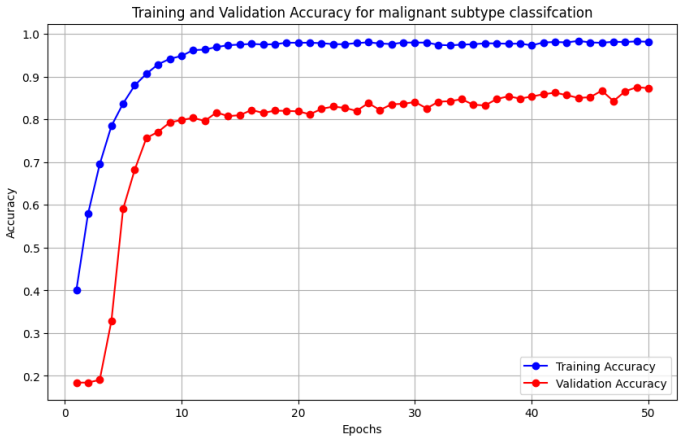




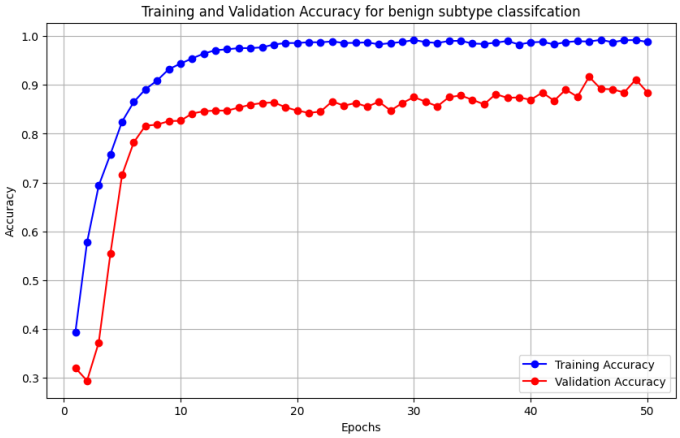


  - **Accuracy and Epochs Curve:** Training and validation curves show model convergence and highlight any overfitting or underfitting.





- These graphs provide insights into how the model performed over different training epochs and across different subtypes.



C. Observations and Inference

- **Overall Performance:**
  - **Precision:** 0.9791
  - **Recall:** 0.9712
  - **F1-Score:** 0.9751
  - The model demonstrates high accuracy with minimal false positives and negatives, indicating strong generalization ability.
- **Benign Subtype Performance (Adenosis, Fibroadenoma, Tubular Adenoma, Phyllodes Tumor):**
  - **Precision:** 0.8932
  - **Recall:** 0.8592
  - **F1-Score:** 0.8722
  - While the model is accurate in identifying benign cases, the slightly lower recall indicates a need to reduce false negatives in this class.
- **Malignant Subtype Performance (Ductal, Lobular, Mucinous, Papillary Carcinomas):**
  - **Precision:** 0.8825
  - **Recall:** 0.8783
  - **F1-Score:** 0.8801
  - Balanced and strong performance in detecting malignant tumors highlights the model's utility for early cancer diagnosis.

V. Conclusion

- The proposed hybrid model integrating InceptionV3, Vision Transformers (ViT), and Self-Supervised Learning (SSL) significantly improved classification accuracy for breast cancer diagnosis.
- The approach addresses the challenges of limited labeled data and resource constraints through SSL and transfer learning.
- Despite its high accuracy, the model still has limitations in terms of interpretability and computational complexity.
- These limitations may impact clinical trust and deployment in low-resource healthcare environments.

VI. Future Work

- Enhance model efficiency and reduce computational requirements to make it more feasible for deployment in resource-constrained settings.
- Expand the dataset to address issues of class imbalance and scarcity, which affect generalization performance.
- Focus on improving model interpretability through explainable AI techniques to gain clinicians' trust.
- Validate the model in real-world clinical settings to test its robustness and practical applicability.
- Implement continuous learning by regularly updating the model with new patient data to adapt to changing diagnostic patterns.

VII. References

1) Dosovitskiy, A., et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale."
2) He, K., et al. "Deep Residual Learning for Image Recognition."
3) Szegedy, C., et al. "Rethinking the Inception Architecture for Computer Vision."
4) Chen, T., et al. "A Simple Framework for Contrastive Learning of Visual Representations."