PATHOVISION: MULTIMODAL DEEP LEARNING FOR ADVANCING PATHOLOGY IMAGING WITH EXPLAINABLE ARTIFICIAL INTELLIGENCE

***Project Reference No.**: 48S_BE_5007*

College : Ramaiah Institute of Technology, Bengaluru
Branch : Artificial Intelligence and Machine Learning
Guide(s) : Dr. Meeradevi
Dr. Sowmya B J
Student(s) : Mr. Prathik Boppudi
Ms. Janya V
Ms. Neeha Ritvika Mandava
Mr. C S Parthasarathy

Keywords:

Multimodal Deep Learning, Pathology Imaging, Explainable AI, PLIP, Grad-CAM

Introduction:

Medical diagnostics, particularly in the field of pathology, play a critical role in early detection, treatment planning and patient care. Traditional diagnostic workflows often rely heavily on manual interpretation by pathologists, which can be time-consuming, subjective, and prone to variability. In recent years, artificial intelligence (AI), especially deep learning, has shown significant promise in automating pathology image analysis. However, most AI models function as "black boxes," lacking transparency in their decision-making, which limits clinical adoption. Moreover, many models focus only on image data, neglecting valuable contextual information present in clinical text or reports.

PathoVision aims to address these gaps by developing a multimodal AI system that combines visual features from pathology images with semantic information from clinical text. The model uses PLIP (Pathology Language–Image Pretraining) for generating multimodal embeddings and integrates a language model (PathChat) to enhance text understanding. To ensure trust and interpretability, the system employs Grad-CAM for generating visual heatmaps that highlight the image regions influencing the model’s predictions. PathoVision has the potential to assist medical professionals, improve diagnostic turnaround times, reduce workload in pathology labs and serve as a powerful educational tool for medical students and trainees, providing interactive learning with visual explanations. The ultimate goal is to create a reliable, accessible, and scalable AI solution that democratizes pathology diagnostics.

Objectives:

1. To build a multimodal deep learning model that integrates pathology images and clinical text for accurate cancer diagnosis.
2. To incorporate explainable AI using Grad-CAM for visual insights into model predictions.
3. To deploy the model using cloud services for scalable, real-time access.
4. To create an intuitive web-based interface for pathologists and learners.
5. To evaluate the model on performance metrics like accuracy, AUC, and F1-score.

Methodology:

The project follows the Agile methodology, enabling iterative development and regular feedback integration.

Figure 1: Agile Methodology

- Data Collection and Preprocessing: Using the OpenPath dataset with over 208,000 image-text pairs, images are resized, normalized, and augmented. Text is cleaned and tokenized.
- Feature Extraction: PLIP is used to generate multimodal embeddings from pathology images and associated clinical text.
- Model Development: A transformer-based architecture is used, integrating pre-trained image and text encoders. Class imbalance is handled with weighted loss functions.
- LLM Integration: A PathChat-like large language model is integrated for interpreting and generating clinical insights from text data.
- Explainability with Grad-CAM: Heatmaps are generated to show which regions of the image influenced the model's prediction.
- Diagnosis Generation: The system provides predictions with interpretability.
- Deployment: The model is deployed on Amazon SageMaker with a web UI for pathologists.
- User Testing & Feedback: Medical users interact with the system and provide feedback for refinement.

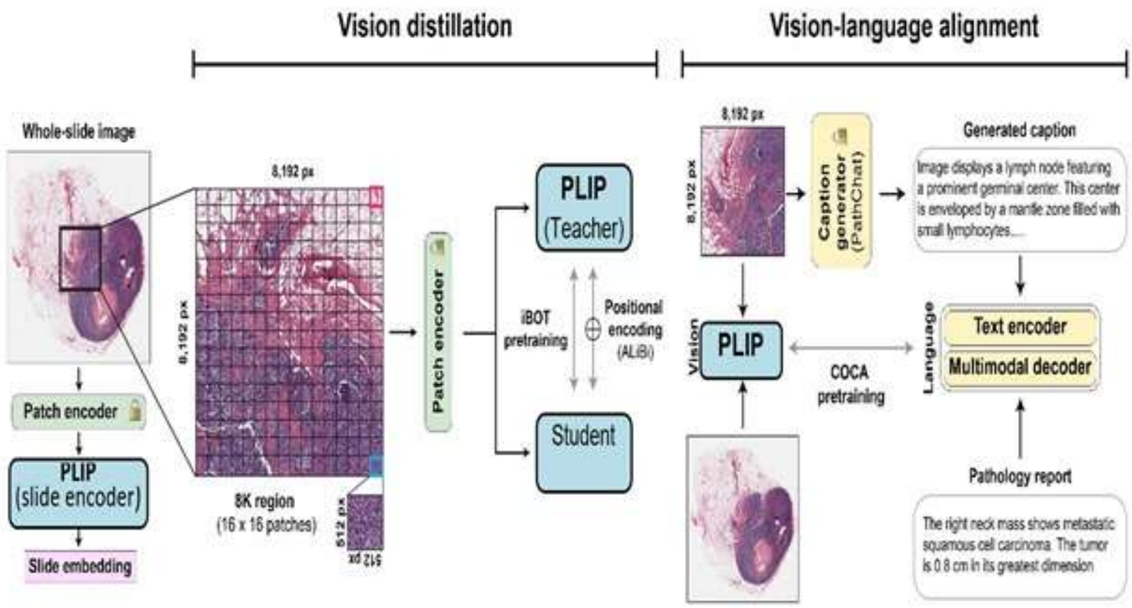

Figure 2: Methodology

Result and Conclusion:

The PathoVision project aims to develop an AI-powered diagnostic tool that integrates pathology images and clinical text for accurate and interpretable disease diagnosis. The model is designed to be trained and evaluated on the OpenPath dataset, which contains over 200,000 image-text pairs, ensuring a rich and diverse training environment.

The use of PLIP for multimodal feature extraction and a transformer-based deep learning model is expected to significantly enhance prediction accuracy. Grad-CAM will provide visual explanations for image-based decisions, helping clinicians understand the reasoning behind AI predictions which improves trust and enabling adoption of the system in clinical settings.

The system will be deployed on Amazon SageMaker, allowing real-time diagnostics with a web-based interface. Performance will be measured using standard metrics such as accuracy, F1-score, and AUC-ROC.

The integration of clinical text interpretation using an LLM will further strengthen context-aware diagnostics.

In conclusion, this project shows strong potential for real-world impact in clinical diagnostics and medical education. It lays a foundation for scalable, explainable, and accessible AI systems that can be adapted across healthcare institutions, especially in underserved areas.

Project Outcome & Industry Relevance:

The project is expected to result in a fully functional, AI diagnostic tool that assists pathologists by providing accurate and interpretable predictions. It will showcase the effectiveness of multimodal deep learning by combining pathology images with clinical text, significantly improving the precision and speed of diagnosis. The integration of Grad-CAM makes the model transparent, fostering greater clinical trust.

From an industry perspective, the system has strong relevance in hospitals, diagnostic laboratories, telemedicine platforms, and healthcare AI startups. Its deployment on Amazon SageMaker ensures scalability, cloud accessibility, and low-latency performance, making it suitable for real-time diagnostics. The system can be integrated into existing healthcare infrastructure to reduce diagnostic backlog and enhance clinical decision-making.

Moreover, the project can support healthcare education by serving as a learning tool for students and interns, offering visual and textual explanations of diagnostic cases. With minor adaptations, it can be extended for use in remote or under-resourced hospitals, making it highly impactful for equitable healthcare delivery.

Working Model vs. Simulation/Study:

This project involves the creation of a working model deployed on the cloud (Amazon SageMaker), capable of real-time input-output functionality with explainable AI features.

Project Outcomes and Learnings:

- Developed a real-time multimodal diagnostic tool integrating image and text inputs.
- Gained hands-on experience with PLIP, Grad-CAM, Amazon SageMaker, and LLMs.
- Learned the importance of model interpretability in medical AI applications.
- Understood the workflow from data preprocessing to cloud deployment.
- Collaborated in an Agile environment and incorporated expert feedback iteratively.

Future Scope:

- Expansion to Other Diseases: The current model focuses on cancer subtype classification, but future iterations can be extended to detect a broader range of diseases, including infectious and autoimmune conditions.
- Integration with Electronic Health Records (EHRs): Structured data can be incorporated from patient records to provide more comprehensive, context-aware diagnostics.

- Mobile and Edge Deployment: The system can be optimized for deployment on mobile devices or edge platforms, enabling on-the-go diagnostics in rural or underserved areas without requiring cloud access.
- Improved Explainability Features: Advanced XAI techniques like SHAP or LIME can be integrated alongside Grad-CAM to offer multi-dimensional explanations for both image and text inputs.
- Clinical Trials and Real-World Validation: The system can be tested in collaboration with hospitals and pathology labs to evaluate its impact in real-world diagnostic workflows and further refine its performance.