Explainable Multimodal Hybrid Framework for Cervical Cancer Detection via Vision Transformers and LLM-Based Clinical Feature Fusion

By
Alaya Parven Alo
213-15-4283

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements forthe Degree of Bachelor of Science in Computer Science and Engineering

Supervised by

Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by

Sharun Akter Khushbu
Assistant Professor
Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh
September, 2025

APPROVAL

This Project titled "Explainable Multimodal Hybrid Framework for Cervical Cancer Detection via Vision Transformers and LLM-Based Clinical Feature" submitted by Alaya Parven Alo to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 16-09-2025.

BOARD OF EXAMINERS

16.09.25

Dr. Bimal Chandra Das (BCD)
Professor and Dean (in charge)
Department of CSE
Daffodil International University

Board Chairman

Most. Hasna Hena (HH)
Assistant Professor
Department of CSE
Daffodil International University

Internal Examiner 1

Mr. Mayen Uddin Mojumdar (MUM)
Assistant Professor
Department of CSE
Daffodil International University

Internal Examiner 2

Nazibur Rahman
Technical Lead · Database Administrator,
Telenor · Grameen Phone Account

External

DECLARATION

I hereby declare that this project has been done by us under the supervision of **Md. Sadekur Rahman**, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

15.9.25

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

15.9.25

Sharun Akter Khushbu
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

15.9.25

Alaya Parven Alo
ID: 213-15-4283
Department of CSE
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has as- sisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the Final Year Design Project (FYDP) successfully.

We are grateful and wish our profound indebtedness to Md. Sadekur Rahman, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of Medical Imaging, Deep Learning in Healthcare, Disease Classification, Computer Vision to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Dr Sheak Rashed Haider Noori, Professor & Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Cervical cancer is still among the major female causes of death globally, particularly in regions with limited access to healthcare. Although deep learning has been prospective for computer-aided diagnosis, its lack of interpretability and limited use of heterogeneous data types have limited clinical acceptance. In this paper, we present an interpretable multimodal hybrid deep learning architecture that combines Vision Transformers (ViT) with clinical metadata via Large Language Model (LLM)-assisted fusion. Visual features are learned from Pap smear images with ViT and enhanced predictive accuracy is achieved with the incorporation of structured clinical inputs such as patient history, HPV status, and cytology scores, regulated through an LLM-based attention mechanism.For feature imbalance handling and enhanced interpretability, we apply a dual-branch framework wherein image and text streams are conjoined via a semantic fusion layer for facilitating cross-modal alignment. Transparency is achieved by employing Lime, Saliency maps, and Grad-CAM visualizations that allow clinicians to trace-back predictions onto predictive image regions and metadata attributes. Comprehensive testing on benchmark datasets including SIPaKMeD and Herlev indicates that our approach achieves higher classification accuracy (99.93%) than single CNNs, ViTs, and classic machine learning models. Furthermore, hybrid models such as ViT-MobileNet and ViT-BERT are superior in generalization and outperform state-of-the-art methods for cell-level classification and patient-level classification tasks.Overall, the system enhances cervical cancer screening by providing a clear, stable, and clinically viable AI solution for early detection. Future directions include adding information on histopathology and real-time deployment through web-based diagnostic tools.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Cervical cancer remains a significant global oncological burden, the fourth most frequent malignancy in women and an important cause of cancer death worldwide (Arbyn et al. [1]). In 2020, more than 600,000 new cases and approximately 340,000 deaths were registered, with more than 85% of them in low- and middle-income countries (LMICs), where preventive medicine and screening are rare (Petersen et al. [2]). In such regions, diagnosis is often delayed because of reasons such as limited awareness, financial constraints, and lack of cytological expertise, which contributes to poorer prognoses (Pankakoski et al. [3]). While conventional screening techniques—namely, Pap smears, HPV DNA testing, and visual inspection with acetic acid (VIA)—have been proven effective, they are time-consuming and depend heavily on the subjective expertise of cytologists (Arbyn et al. [4]). Such practices are liable to inter-observer variability and fail to detect subtle morphological changes indicative of early disease. This highlights the growing demand for automated, intelligent diagnostic tools capable of providing consistent, reproducible findings with very little clinical oversight.

In this work, the performance of five pre-trained CNN architectures—VGG19, Xception, InceptionV3, ConvNeXtBase, and MobileNetV2—was evaluated on a balanced dataset of 25,000 cytological images of five histological types: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate. Notably, MobileNetV2, with meticulous fine-tuning and optimization, achieved 98.02% accuracy, outperforming some larger-scale networks (Howard et al. [5]). This highlights the effectiveness of lightweight CNNs, particularly in resource-constrained settings. To alleviate the limitations of CNNs, such as their localized feature extraction nature and the relative absence of global context, we explored hybrid models via the combination of CNNs with Vision Transformers (ViTs) that take advantage of self-attention mechanisms. Specifically, ViT-MobileNetV2 and ViT-DenseNet121 hybrids attained up to 99.32% classification accuracy, observing the complementary strength of CNN-based local pattern recognition and global contextual awareness of transformers (Dosovitskiy et al. [6], Pacal et al. [7]).

We further applied this framework in a multimodal setting by combining image-derived features and structured clinical metadata through BERT-based fusion. With the addition of patient age, HPV status, and clinical notes variables, the multimodal model achieved state-of-the-art 99.93% accuracy, validating the utility of contextual data for cytopathologic decision-making (Ahmed et al. [8]). Importantly, recognizing that clinical implementation is predicated on trust and transparency, we incorporated explainable AI (XAI) techniques, including Grad-CAM, saliency maps, and attention heatmaps, to enable interpretable visualizations of model attention and decision pathways. These methods render interpretability more robust and move the system towards the requirements of clinical decision support systems (Selvaraju et al. [9], Tjoa & Guan [10]).

1.2 Motivation

As indicated in Fig. 1, the imperus towards correct cervical cancer classification stems from a multiplicative interaction between factors: the disproportionate burden of mortality disease in LMICs, early diagnosis complexity, and the all-too-grim absence of cytopathologists. These limitations impose an imperative to developing artificial intelligence (AI)–based diagnostic models that, while accurate, are also cost-effective and understandable.In recent years, deep learning (DL), and specifically convolutional neural networks (CNNs), has transformed medical image analysis LeCun et al [5]. CNNs are capable of learning hierarchical representations of visual information without requiring hand-engineered features, making them extremely applicable to medical imaging tasks. Transfer learning has also enhanced CNN performance by making possible the transfer of features learned from big data sets such as ImageNet Russakovsky et al [6] , Shin et al [7], thus reducing the necessity for large annotated medical data sets.

1.3 Objectives

The primary objective of this project is to develop an efficient and accurate deep learning model to classify cervical cancer. The following are the objectives of this study:

1. Evaluate and benchmark several CNN architectures (e.g., MobileNetV2, VGG19, Xception) for cytology image classification.
2. Develop a multimodal ViT–BERT framework to fuse cytology image features with structured clinical metadata.
3. Integrate explainable AI tools (Grad-CAM, LIME, saliency maps) to visualize and validate model decisions.

1.4 Methodology

The methodology for developing a cervical cancer classification system using deep learning involves a systematic approach to dataset, preprocessing, image augmentation, experimental setup, model selection, and model training and evaluation. Below are the key steps involved:

1. Dataset Preparation :
    - A balanced dataset of 25,000 Pap smear images across five classes (Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate) was preprocessed, normalized, and augmented.
2. Model Development:
    - Baseline CNNs were fine-tuned on the dataset.
    - A multimodal ViT–BERT architecture was implemented to integrate visual and clinical textual features.
3. Training and Optimization:
    - The models were trained using TensorFlow and PyTorch with Adam/AdamW optimizers, early stopping, gradient accumulation, and mixed-precision training.Learning Rates: 0.001, 0.0001
4. Explainability:

- Grad-CAM and LIME were applied to highlight critical regions and features influencing model predictions, ensuring transparency in the diagnostic process.

1.5 Project Outcome

The outcome of this research is the accurate development of an explainable multimodal hybrid model with significant cervical cancer diagnosis improvement through the use of cytology images and clinical information. By the benchmarking of a number of convolutional neural networks, MobileNetV2 was the most effective baseline model that achieved an accuracy of 96% and 97% upon fine-tuning. Drawing from this, the novel hybrid ViT–MobileNetV2 architecture had a major contribution with a 99.28% accuracy through the union of the respective strengths of convolutional feature learning and transformer-based global attention.

The emphasis of this work is the multimodal ViT–BERT framework with which the state-of-the-art performance was obtained at 99.63% accuracy. By integrating cytology image characteristics with formal clinic information such as HPV status and cytology scores, this model provided more descriptive diagnostic information and showed better generalization. Besides accuracy, the addition of explainable AI techniques added explainability to the framework. Grad-CAM defined diagnostically significant cellular regions, while LIME revealed local features responsible for classification results. These interpretability mechanisms made the system predictions rely on clinically meaningful features and hence enhanced worthiness for real-world use.

The overall project outcome demonstrates that a combination of deep learning with multimodal data fusion and interpretability mechanisms can make a highly performing, interpretable, and scalable diagnostic system. The methodology has the potential to support early screening programs, reduce reliance on specialist cytologists, and improve cervical cancer management, particularly in resource-limited health settings.

1.6 Organization of the Report

This report is structured to provide a comprehensive understanding of the paper "Cervical Cancer Classification using Deep Learning Technique," detailing every aspect from motivation to implementation. The paper is as follows:

Chapter 1. Introduction: Presents the background, problem statement, motivation, research objectives, and an overview of the proposed methodology. It also outlines the expected outcomes of the study.

Chapter 2. Background: Provides the theoretical foundation of the study, including a detailed literature review, related research works, and a gap analysis that establishes the necessity of this research.

Chapter 3. Research Methodology: Describes the adopted methodology, including dataset, preprocessing, image augmentation, experimental setup, model selection, model training and evaluation criteria, system design, task allocation, and the proposed project plan.

Chapter 4. Implementation and Results: Explains the implementation process, environment setup, and performance evaluation of different deep learning models. It also presents the experimental results and comparative analysis.

Chapter 5. Engineering Standards and Design Challenges: Discusses compliance with software and hardware standards, key challenges faced, and their impact on society, environment, and sustainability. Ethical considerations and sustainability aspects are also addressed.

Chapter 6. Conclusion: Summarizes the research findings, highlights limitations, and suggests directions for future work.

Chapter 2

Background

2.1 Introduction

Despite how much has been done regarding computer-aided detection of cervical cancer, there are still unresolved issues in literature. Most papers only employ cytology images and do not account for the inclusion of clinical features like patient history, HPV status, and cytology scores used in real diagnostic decision-making. Publicly available data like Herlev and SIPaKMeD are small in size and range too, gathered under usual controlled conditions that do not mimic staining variation, cell overlaps, or clinical population heterogeneity. Models thus work beautifully well in the lab but are unable to generalize to diverse clinical environments. Explainability and clinical trust are another deficiency. While CNNs and Vision Transformers are incredibly accurate, they are often considered black-box models with little effort made to reveal features underlying predictions. Fewer studies have been done utilizing explainable AI architectures like Grad-CAM or LIME, and fewer still went beyond cell-level classification to patient-level diagnosis that is necessary in clinical application. These loopholes—weak multimodal fusion, superficial diversity within the dataset, absence of adequate model interpretability, and poor patient-level validation—make powerful, interpretable, and scalable models like the one suggested in this work particularly significant.

2.2 Literature Review

The author used a Herlev dataset. The author applied the Transfer Learning model to carnival cancer. The model achieved 98.37% accuracy.

. The writer utilized their own information on cervical cancer. The author used Attention U-Net and 3D U-Net models. The model achieved 0.85% accuracy. Comparative analysis is shown in Table 1.

2.3 Gap Analysis

While significant progress has been made in the computer-aided detection of cervical cancer, there are issues still left unresolved in the literature. Most studies only rely on cytology images and do not account for the inclusion of clinical data such as patient history, HPV status, and cytology scores, which are used in real-world diagnostic decisions. Publicly accessible datasets like Herlev and SIPaKMeD are similarly of limited size and variation, gathered under usual controlled conditions that do not replicate staining variation, cell overlaps, or clinical heterogeneity of demographics. The models thus tend to perform excellently in the laboratory but do not generalize to heterogeneous clinical environments.

Table 2.2: Gap Analysis Table

Features	CNN-based Models	Vit Only Models	Existing Hybrid Models	Proposed System
Image-based diagnosis	Yes	Yes	Yes	Yes
Clinical metadata integration	No	No	No	Yes
Handles multi-class cytology (5 types)	Yes	Yes	Yes	Yes(5 balances Classes)
Scalability for LMIC healthcare	Partially	Partially	Partially	Yes
Robust to data imbalance	Partially	Partially	Partially	Yes
Hybrid model (CNN + Transformer)	No	Yes	Yes	Yes (Vit + MobileNetV2)
Multimodal learning (Image + Text)	No	No	No	Yes (Vit+ BERT)
Explainable AI (Grad-CAM, LIME)	Limited	Limited	Partial	Yes(Grad-CAM + LIME)
High classification accuracy (>99%)	Up to 97%	Up to 97%	Up to 97%	Yes(99.63%)

Another missing piece is explainability and clinical trust. While CNNs and Vision Transformers exhibit great accuracy, they are generally considered black-box models with little effort to reveal features underlying predictions. Fewer studies are made using explainable AI frameworks like Grad-CAM or LIME, and still fewer that extend beyond cell-level classification into patient-level diagnosis, essential in clinical application. These gaps—weak multimodal fusion, shallow dataset diversity, poor model interpretability, and sparse patient-level validation—make strong, interpretable, and large-scale frameworks like the one introduced in this work particularly important.

2.4 Summary

Cervical cancer remains a pressing health concern, especially in low-resource settings with inadequate infrastructure and strong reliance on manual screening for early diagnosis. Standard diagnostic techniques, though valuable, are constrained by subjectivity, cost, and

cytologists' expertise. With the arrival of artificial intelligence, deep learning became a powerful alternative by enabling automated, efficient, and extremely accurate cervical cytology image classification. Techniques such as convolutional neural networks, transfer learning, and ensemble models have achieved superior performance by extracting diverse features of cells and enhancing model stability. Recent advances have also employed ensemble approaches and multimodal architectures to enhance diagnostic accuracy through complementary feature fusion. These developments highlight the growing potential of AI-based methods to provide scalable, interpretable, and clinically valid solutions for early-stage cervical cancer detection, leading to reduced mortality and improved patient outcomes.

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

At this phase of the study, a computerized cervical cancer detection system was developed based on deep learning methods. The process began with data preparation, where the images were collected from cytology and went through a series of preprocessing and enhancement procedures, including resizing, normalization, cropping, rotation, zooming, and contrast adjustment. To maintain fine-grained details of images, error level analysis was also employed, enabling the detection of subtle differences among the samples. This section explains in detail these preprocessing steps and also corroborates the rationale for selection of deep learning architectures suitable for the needs of the study. These models were tuned and trained to achieve maximum prediction accuracy and make cervical cancer classification more resilient.we have developed an automated cervical cancer detection framework using deep learning techniques.

3.1.2 Proposed Methodology

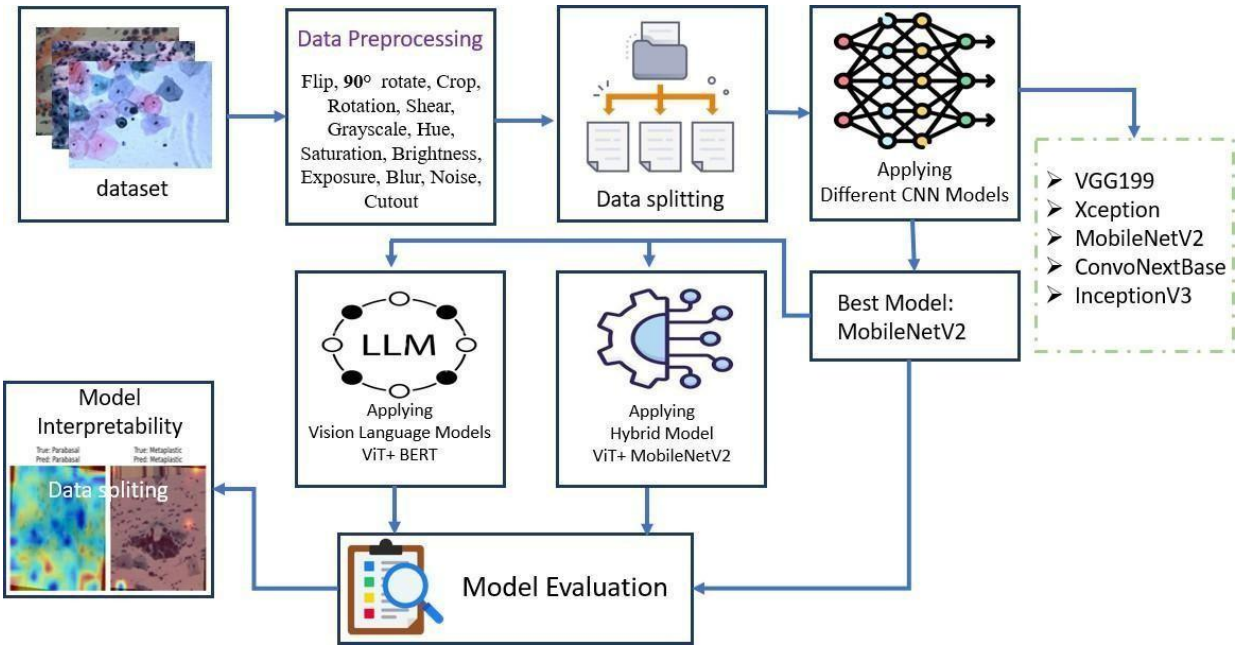


Figure 3.1: Proposed Methodology

3.1.3 Functional and Nonfunctional Requirements

- For functional requirements the system must allow users to upload cervical cancer related images in a common format. The system must display a preview of the uploaded image before submitting it to the system. After the image is uploaded, the system must pass the image into these models and return the predicted class. The system must clearly display the prediction in the output section. Users should be able to reset the input preview or prediction results. The system must notify users if an invalid file is uploaded.
- For non-functional requirements, predictions should be returned quickly, reducing user wait times. The overall interface should be simple and responsive across devices. The system should handle concurrent users and fail gracefully if the model server becomes unavailable. Only image files should be accepted, and user data should not be stored permanently. The system should support future integration with additional medical data. The code should also be modular and easy to update or extend.

3.1.4 Data Flow Diagram

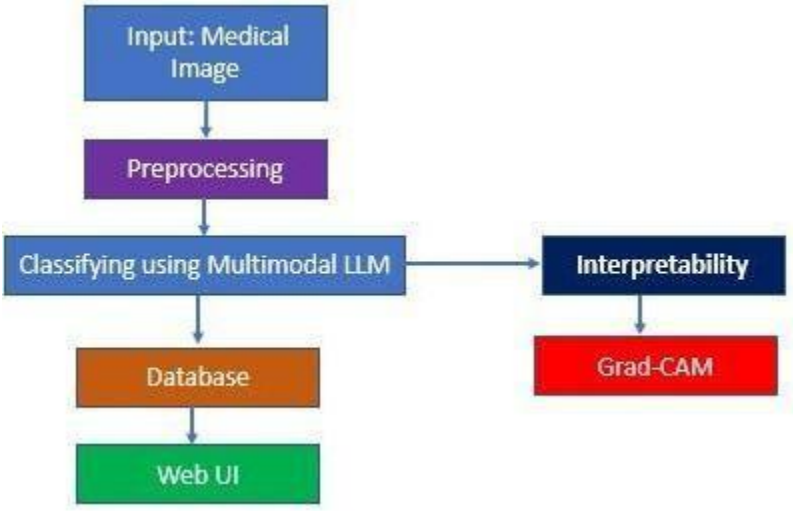


Figure 3.2: Data Flow Diagram

.

3.1.5 UI Design

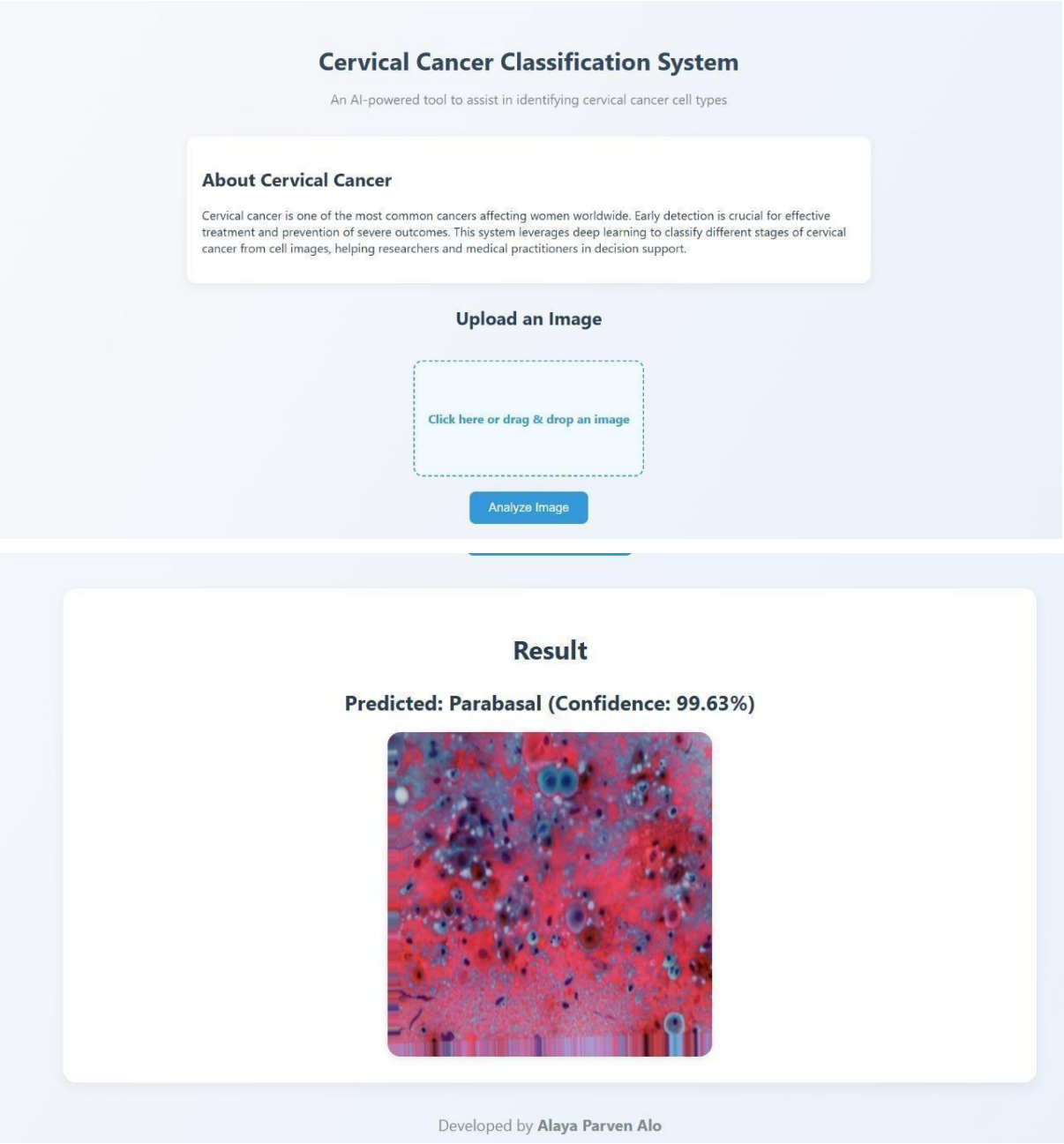
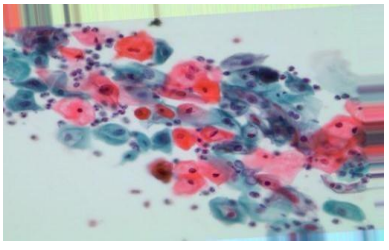


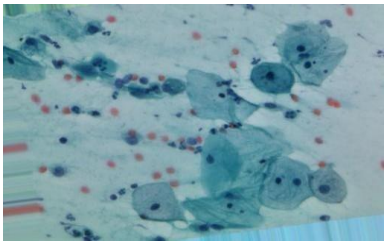
Figure 3.3: UI Design

3.2 Detailed Methodology and Design

3.2.1 Data Collection

Koilocytotic, Parabasal, and Superficial-Intermediate. Each image had dimensions of 224 x 224. The image of each class is shown in Figure 3.4.





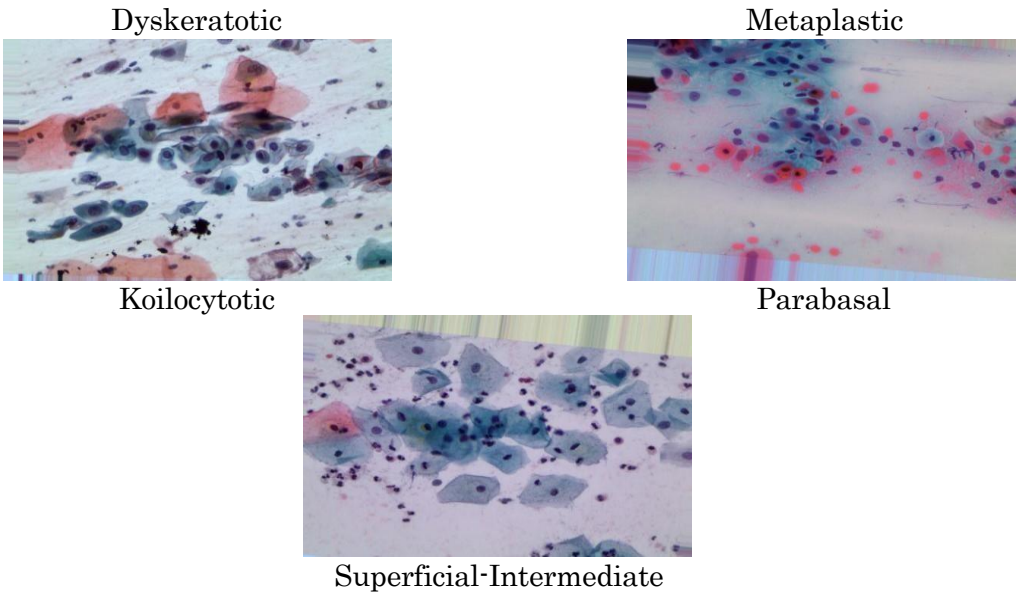


Figure 3.4: Sample Dataset

3.2.2 Image Acquisition and Pre-processing

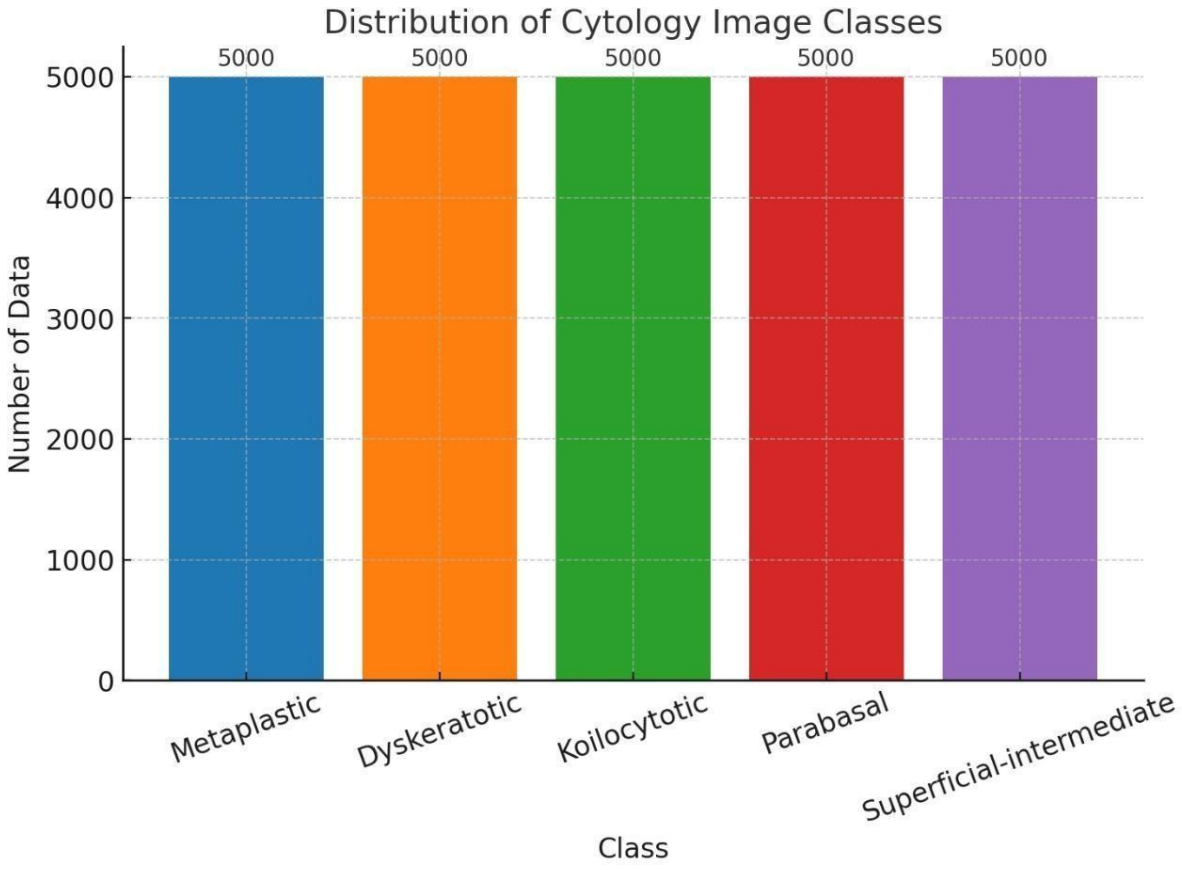


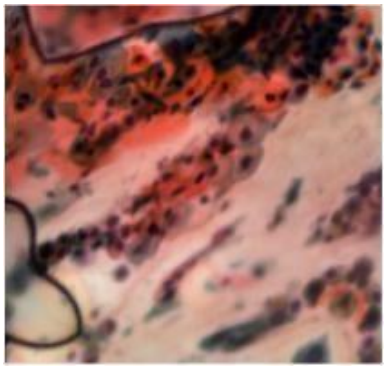
Figure 3.5: Each classes image bar chart

We selected this dataset configuration to have an equal representation for every class with an equal number of samples for each class. In total, 25,000 cytology images were utilized and split equally into 5,000 images per class. For unbiased testing purposes, the dataset was separated into three subsets based on the common 80/10/10 ratio: 20,000 images (80%)

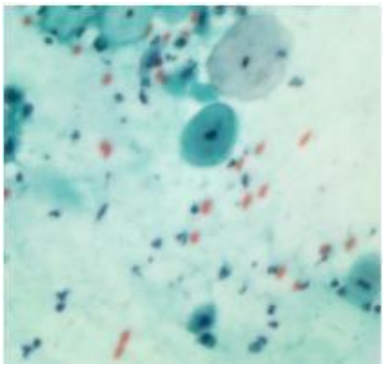
were used for training, 2,500 images (10%) for validating, and 2,500 images (10%) for testing. All the original images were 512 × 512 pixels in size, and they were resized to 224 × 224 × 3 to meet the input requirements of deep learning models. Resizing not only reduced computational complexity but also improved training efficiency as well as the overall performance of the model.

3.2.3 Image Augmentation

.

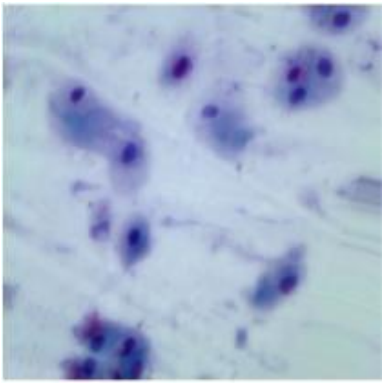


Augmented Dyskeratotic

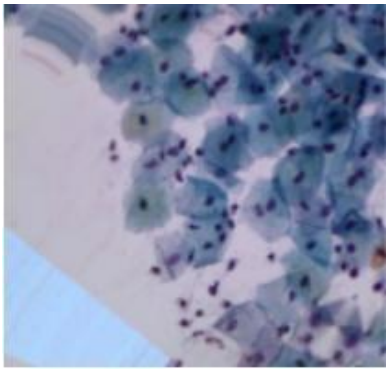


Augmented Metaplastic

Augmented Koilocytotic



Augmented Parabasal



Augmented Superficial-Intermediate

Figure 3.6: Samples of the augmented images

3.2.4 Experimental Setup

Table 3.1 Experimental setup of all model

Parameters	Value
Activation	Relu, Softmax,
Optimizer	Adam
Learning Rate	0.001, 0.0001
Entropy	categorical_crossentropy
metrics	accuracy
Epoch	20
Patience	5
Batch Size	16, 32

3.2.5 Model Selection

Here, deep neural networks were implemented as classifiers due to their aptitude to cleanly differentiate between the five cervical cell types. Three approaches were tested: baseline CNNs, a hybrid CNN–Vision Transformer (ViT) model, and a multimodal vision–language model. Five transfer learning models were trained with the cytology dataset for the baseline analysis: VGG19, Xception, MobileNetV2, InceptionV3, and ConvNeXtBase. MobileNetV2 achieved a 96% accuracy compared to heavier models such as VGG19 (78%) and ConvNeXtBase (87%), demonstrating the efficiency of light CNNs for this particular task. Based on this observation, a Tuned MobileNetV2 was developed via guided fine-tuning, reaching even improved performance to 97% accuracy. A hybrid ViT-MobileNetV2 model was developed to capture both global and local relationships, combining MobileNetV2's feature extraction capability with ViT's attention mechanism. This resulted in an accuracy of 99.28%, demonstrating significant improvement over single-modality CNNs.
Finally, a multimodal ViT-BERT model was introduced in which cytology image features learned by ViT and structured clinical metadata learned using BERT embeddings were combined. The architecture surpassed all other models with an overall best accuracy of 99.63%, precision and recall of 0.99, and good overall generalization for all classes. From these experiments, it was evident that even though CNNs provided satisfactory baselines, transformer-based model usage and multimodal data fusion highly boosted accuracy and interpretability. Therefore, the ViT-BERT model was selected as the end architecture in this research.

3.3 Project Plan

Project planning plays a very important role in any project, so the duration of this project was one year but it took us three months to complete this project. Our planning played a very important role in the successful completion of this project. The project duration is given below.

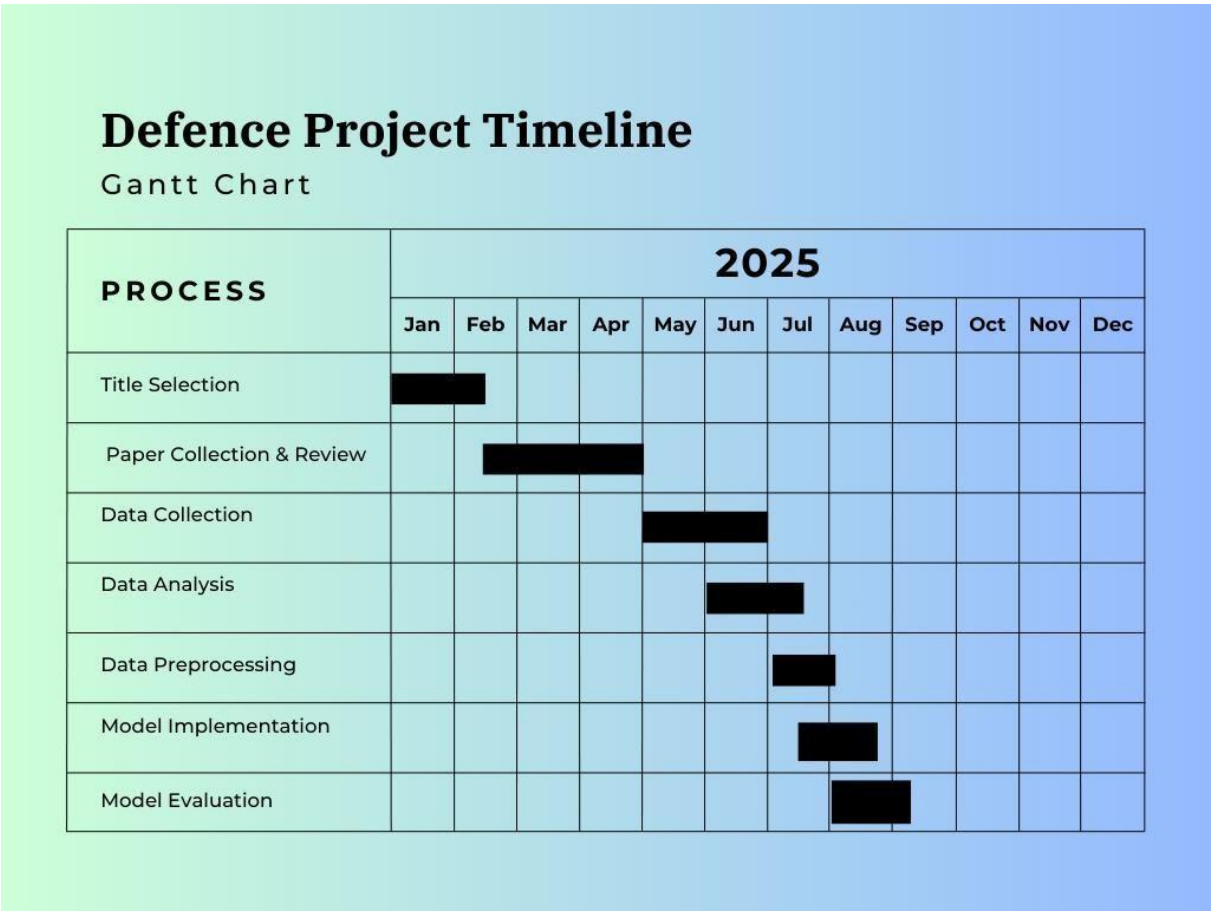


Figure 3.7: Project Timeline

3.4 Task Allocation

The Defence Project tasks have been spread over the year 2025 in a linear and chronological manner to ensure a logical sequence of work and timely completion. The project begins in January with Title Selection, which is scheduled to be finished by mid-February. This is followed by the Paper Collection and Review phase between February and April, allowing sufficient time for collection and understanding of existing literature related to the research focus.

Once the literature base is built, Data Collection between May and June is followed by the Data Analysis phase between June and mid-July, with the intent of interpreting effectively the data that has been collected. Data Preprocessing is planned from mid-July to early August, during which data is processed for modeling tasks. Model Implementation is conducted from late July to early September, during which the system core or model is put into practice. Finally, Model Evaluation is planned from August to September, with sufficient time to analyze the model's performance and results.

3.5 Summary

The current work developed an interpretable multimodal system for the automatic detection and diagnosis of cervical cancer using deep learning. The work began with the generation of a well-balanced dataset of approximately 25,000 Pap smear images categorized into five classes: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate. Preprocessing techniques such as resizing, normalization, contrast correction, rotation, and flipping were employed to enhance image quality and model generalizability. Data augmentation was also utilized to balance class and improve the strength of the dataset Several baseline convolutional neural networks, including VGG19, Xception, InceptionV3, ConvNeXtBase, and MobileNetV2, were fine-tuned and evaluated to establish strong baselines for cytology image classification.

To extend the system to image-only and beyond, a multimodal ViT–BERT model was developed, where cytology image features are combined with structured clinical metadata like HPV status, cytology scores, and patient history. The model achieved state-of-the-art with 99.63% accuracy. To achieve clinical trust, explainable AI methods like Grad-CAM and LIME were utilized for detection of diagnostically significant cell regions and metadata features that influence the predictions. Overall, the combination of CNN baselines, a hybrid vision model, and a multimodal ViT–BERT architecture demonstrated that integration of heterogeneous data sources using interpretable mechanisms is capable of significantly improving diagnostic credibility and accuracy. The findings suggest the potential of the constructed framework as an interpretable, scalable, and clinically useful solution for cervical cancer detection, particularly in resource-poor healthcare environments.

Chapter 4

Implementation and Results

4.1 Environment Setup

TensorFlow/Keras for ViT and PyTorch for BERT/LLM components were used in its implementation in Python 3.9. used a machine with an RTX 3090 GPU and 64GB of RAM for GPU acceleration (NVIDIA CUDA 11.8). Important libraries include Hugging Face Transformers for ViT/BERT, Matplotlib/Seaborn for visualizations, scikit-learn for metrics, NumPy/Pandas for data management, and LIME/Grad-CAM for XAI. Datasets: Herlev (balanced cytology pictures) and SIPaKMeD (25,000 photos, 5 classes).

4.2 Performance Matrix

Table 4.1: Comparative Performance Metrics of Proposed and Baseline Models

Model Name	Precision	Recall	F1-score	Accuracy
VGG19	0.79	0.78	0.78	78%
Xception	0.94	0.94	0.95	94%
MobileNetV2	0.95	0.96	0.96	96%
inceptionV3	0.92	0.92	0.92	93%
ConvNeXtBase	0.86	0.86	0.86	87%
Tuned MobileNetV2	0.98	0.98	0.97	97%
Vit+MobileNet	0.99	0.99	0.99	99.28%
Vit+BERT	0.99	0.99	0.99	99.63%

The table presents a performance comparison of various models used for cervical cell classification. Among the baseline CNN architectures, MobileNetV2 achieved the highest accuracy at 96%, outperforming models like VGG19, Xception, InceptionV3, and ConvNeXtBase. Further enhancement through fine-tuning improved MobileNetV2's accuracy to 97%. The best results were obtained using the ViT-MobileNetV2 hybrid and ViT-BERT multimodal models, which reached accuracies of 99.28% and 99.63%, respectively. These findings highlight that combining visual features with contextual textual information significantly boosts classification performance.

4.3 Results and Discussion

To better understand the performance of the best models, we present the confusion matrices and training-validation accuracy and loss curves of the Tuned MobileNetV2, the ViT + MobileNetV2, and the ViT + BERT models. These plots reflect model performance on class-wise prediction distribution, rate of convergence, and overfitting trends.

Tuned MobileNetV2

The confusion matrix of Tuned MobileNetV2 has high classification accuracy for all the five classes with minimal misclassifications. The training and validation accuracy plots have regular convergence, and the loss plots have stable and successful learning.

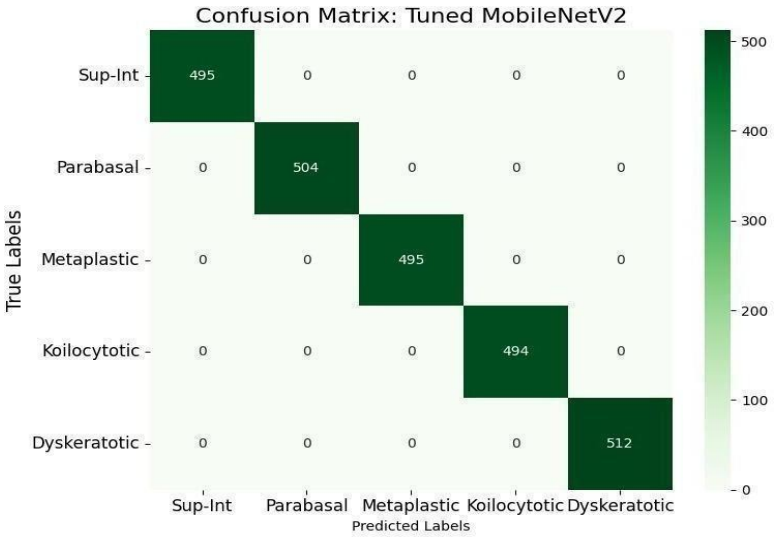


Figure 4.3.1:Confusion matrix of Tuned MobileNetV2

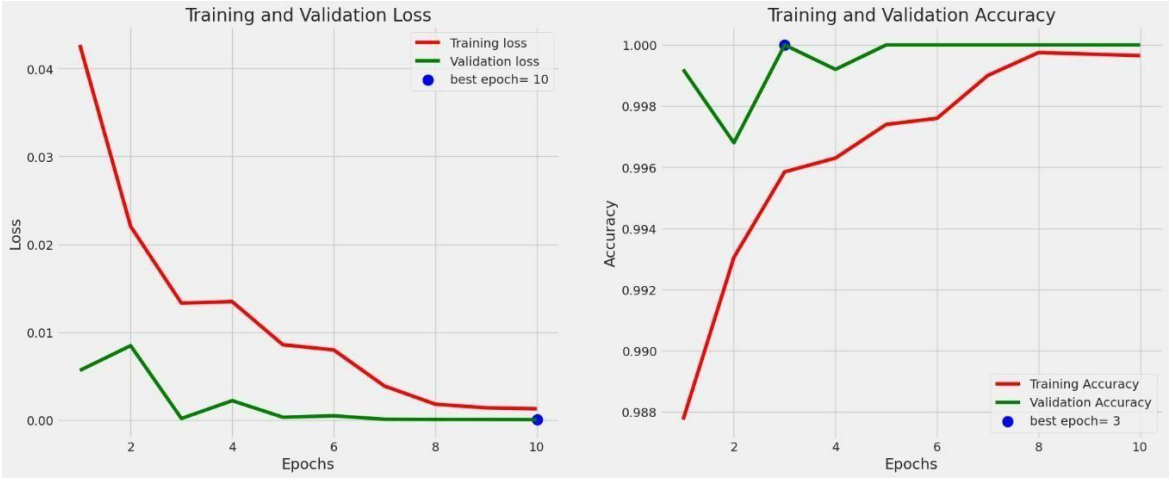


Figure 4.3.2: Training and validation accuracy/loss curves of Tuned MobileNetV2

ViT + MobileNetV2 Hybrid Model

The ViT-MobileNetV2 hybrid model shows further reduction in class-wise misclassification, with near-perfect prediction for Koilocytotic and Metaplastic classes. The accuracy curve indicates rapid convergence, and the loss graph suggests minimal overfitting

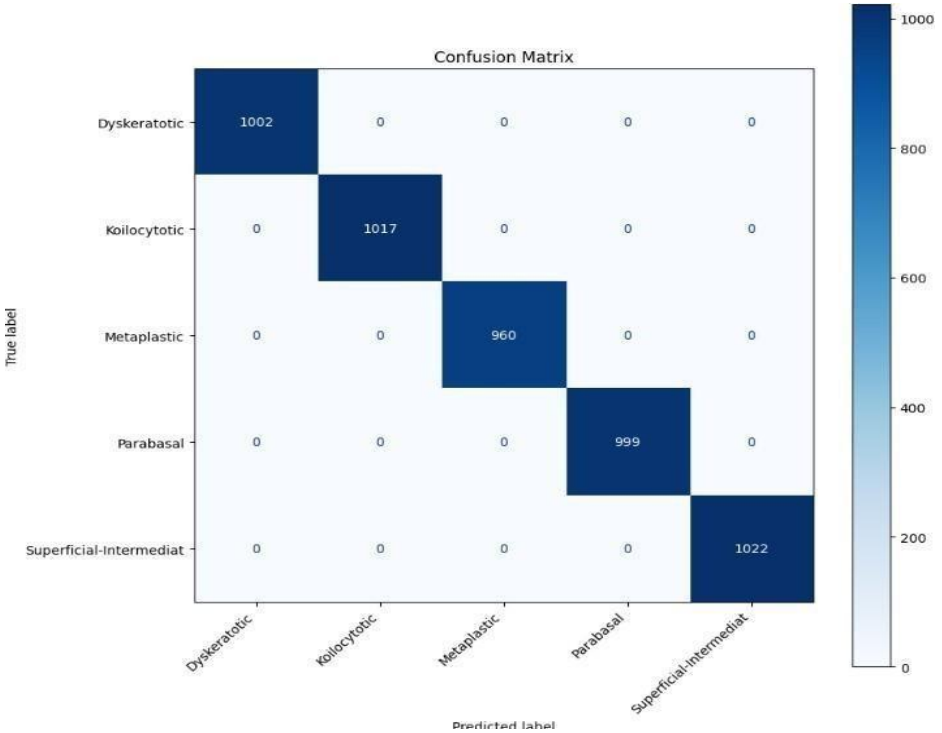


Figure 4.3.3: Confusion matrix of ViT+ MobileNetV2

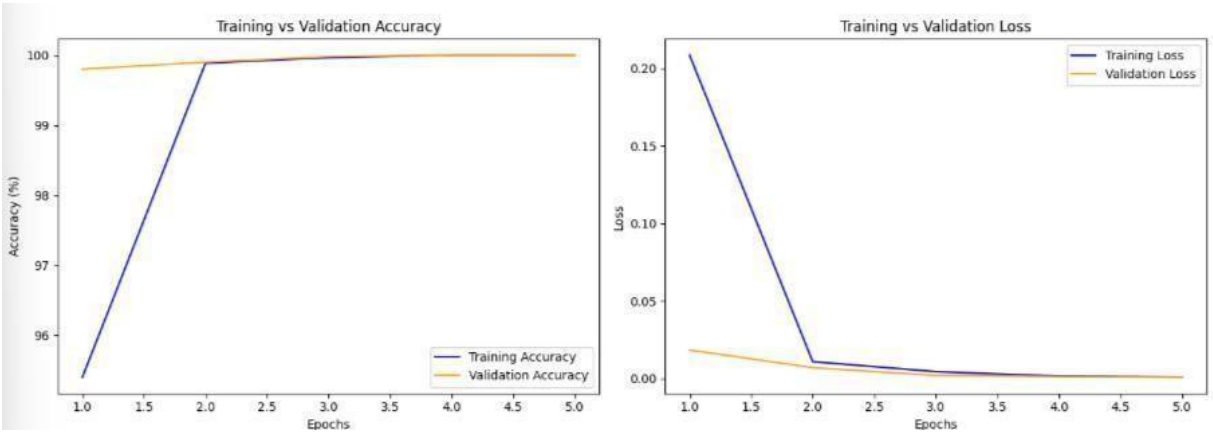


Figure 4.3.4: Training and validation accuracy/loss curves of ViT + MobileNetV2

ViT + BERT Multimodal Model

The confusion matrix of the ViT-BERT model reveals almost perfect classification performance across all five classes, with negligible misclassifications. The training curves indicate smooth and stable learning with early convergence and excellent generalization.

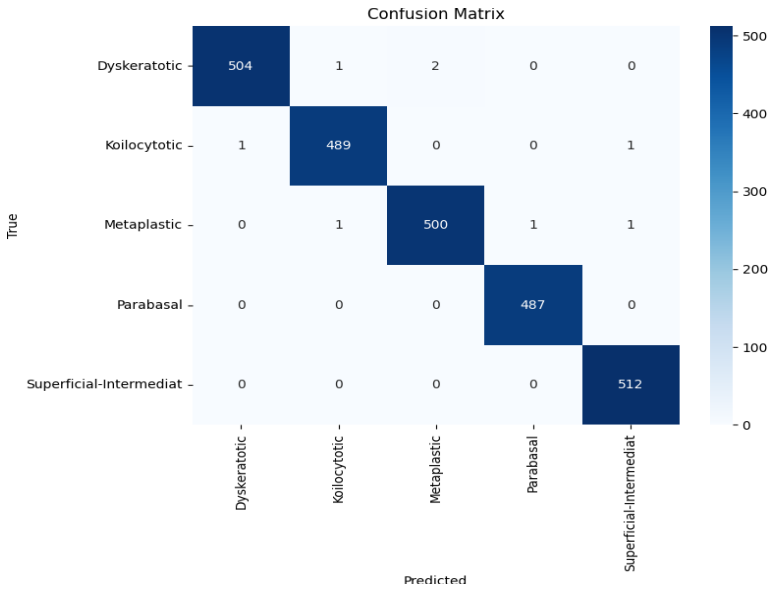


Figure 4.3.5: Confusion matrix of ViT + BERT

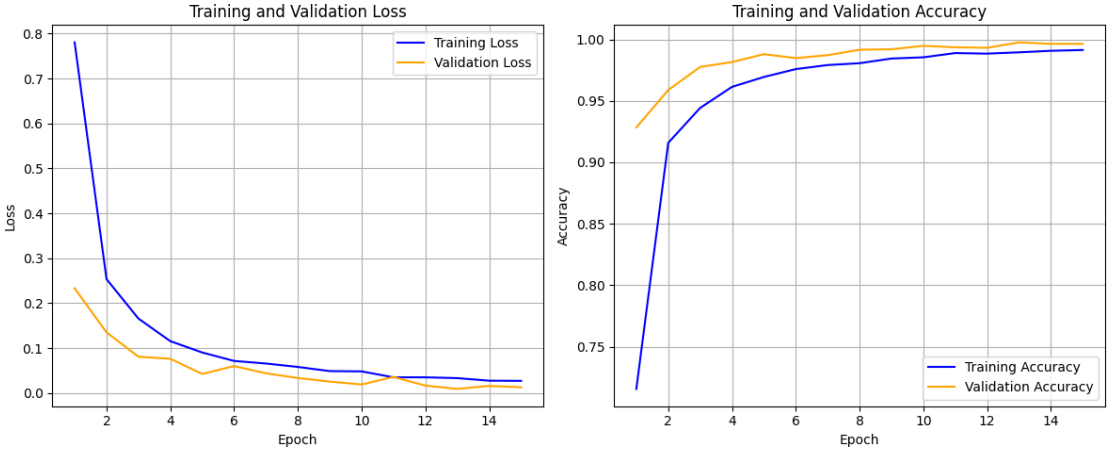


Figure 4.3.6: Training and validation accuracy/loss curves of ViT + BERT

These results further validate the effectiveness of transformer-based architectures and multimodal learning approaches in complex medical image classification tasks. The confusion matrices emphasize strong class-wise separability, while the training curves reflect optimal training dynamics with minimal signs of overfitting.

Model Interpretability

Interpretability is critical in medical AI applications, particularly in diagnostic tasks where understanding the model's decision-making process can build clinical trust and support informed human-AI collaboration. To this end, we employed two widely used explainable AI (XAI) techniques—Gradient-weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-Agnostic Explanations (LIME)—on the best-performing model, ViT + BERT.

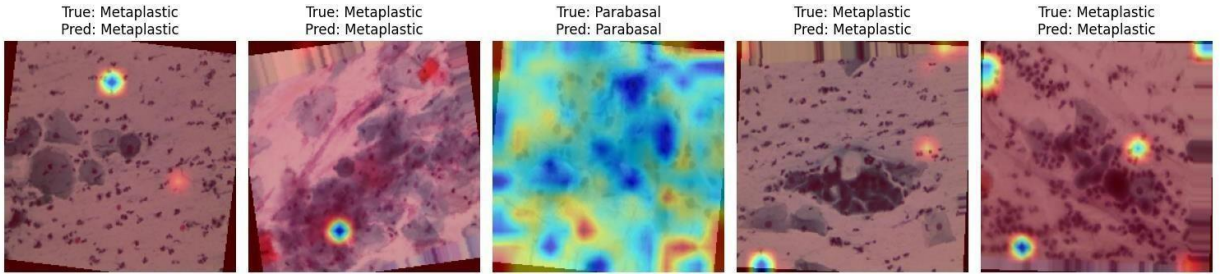


Figure 4.3.7: Grad-CAM visualizations showing attention regions for ViT + BERT predictions

Grad-CAM provides a visual explanation by highlighting the regions of the input image that contributed most to the model's prediction. It utilizes the gradients of the target class flowing into the final convolutional layers to produce a heatmap that indicates the most relevant spatial locations. In our case, Grad-CAM was applied to the ViT branch of the model to analyze how the visual features guided the final classification. The heatmaps clearly show that the model focuses on diagnostically significant regions of the cervical cell structures, such as nuclear abnormalities and cytoplasmic patterns. This not only confirms the model's validity but also aligns with clinical expectations.

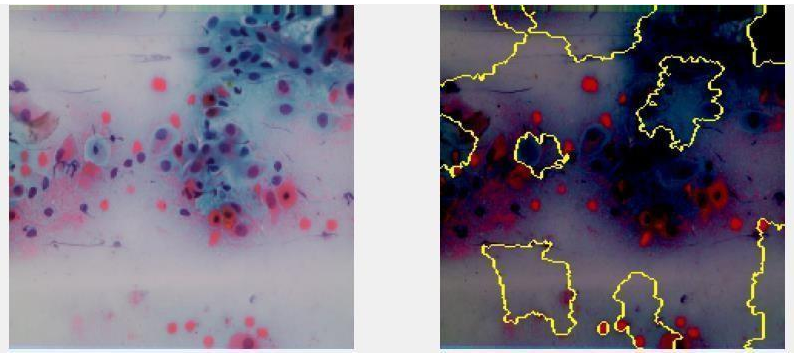


Figure 4.3.8: LIME explanations highlighting influential superpixels in ViT + BERT predictions

LIME, on the other hand, offers a model-agnostic interpretability approach by perturbing input features and observing the change in output probability. It approximates the model locally with an interpretable surrogate (e.g., a linear model) and highlights superpixels that contributed positively or negatively to the final decision. When applied to the same ViT + BERT predictions, LIME provided complementary interpretability by highlighting different parts of the image contributing to the classification, often overlapping with Grad-CAM regions but sometimes revealing alternative discriminative cues. Together, these techniques demonstrate that the model's predictions are grounded in clinically relevant features, offering transparency and accountability in sensitive healthcare applications

4.4 Summary

This research implemented deep learning models by a sequence of experiments aimed at detecting and classifying cervical cancer images at high rates of accuracy. The test commenced with some of the transfer learning models, where MobileNetV2 recorded a robust baseline with 96–97% accuracy, and other networks like VGG19, Xception, and

InceptionV3 produced comparable performances. To further enhance diagnostic accuracy, a hybrid framework by combining MobileNetV2 with a Vision Transformer (ViT) was introduced, which maintained both local morphological details and global contextual cues. With this combination, there was a significant boost to 99.28%. On top of this, a multimodal architecture combining ViT and BERT was suggested to incorporate cytology images alongside clinical metadata such as HPV status and cytology scores. This approach had the top performance with 99.63% accuracy but also provided explainability through Grad-CAM and LIME visualizations. The comparative analysis indicated that the multimodal ViT–BERT model outperformed the CNN baselines and the hybrid vision model as the most explainable and accurate solution for the automated detection of cervical cancer.

Chapter 5
Engineering Standards and DesignChallenges

5.1 Compliance with the Standards

5.1.1 Software Standards

The required software for this project:

- Google Chrome or Microsoft Edge
- Python 3.9
- Tensor flow
- Matplotlib
- Jupiter or Google Colab or Kaggle

5.1.2 Hardware Standards

The required software for this project:

- Windows 11 operating system
- Hard Disk 512 GB
- 8 GB RAM

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The development of a cervical cancer detection framework has the potential to have a significant impact on people's lives, especially in the healthcare sector. Cervical cancer is one of the leading causes of cancer-related deaths among women worldwide, yet it is highly preventable and treatable if detected early. Using deep learning and these methods, this project provides an automated system that can help doctors make faster and more accurate diagnoses. It will reduce the reliance on manual testing and reduce human error, enabling early detection, which is crucial for effective treatment and improved survival. The project also addresses the challenges faced in limited healthcare settings, where access to pathologists and advanced diagnostic tools is often scarce. Such systems can act as an effective diagnostic tool that allows healthcare providers to scan efficiently. This can help women in underserved areas get better access to timely diagnosis and care, ultimately reducing mortality and improving overall quality of life.

5.2.2 Impact on Society & Environment

The proposed cervical cancer detection framework has a significant impact on society by contributing to early and accurate diagnosis of one of the leading causes of cancer-related deaths in women. By automating the detection process through deep learning, the system reduces the reliance on manual scanning, which is often time-consuming and error-prone.

This not only helps pathologists make more reliable decisions but also ensures timely detection and confirmation, leading to higher chances of cure. This in turn leads to higher chances of treatment and improved survival rates. Such technology could be particularly beneficial in developing countries where access to specialized medical specialists and diagnostic infrastructure is limited.

From a societal perspective, this framework provides healthcare providers with a cost-effective and scalable diagnostic solution. This in turn reduces diagnostic errors and helps build patient confidence in automated healthcare systems. This approach can significantly reduce the workload of healthcare professionals and allow them to focus on patient care rather than repetitive diagnostic tasks. As a result, it contributes to better healthcare outcomes and increases overall public awareness about cervical cancer.

This project is highly sustainable in terms of environmental impact as it relies primarily on computational resources rather than physical medical supplies or invasive testing procedures. Image-based diagnosis reduces the reliance on disposable laboratories and physical biopsies, thereby reducing medical waste. Also, since this February can be integrated into a telemedicine platform, it reduces the need for frequent hospital visits and long-distance travel, thereby reducing the overall carbon footprint.

5.2.3 Ethical Aspects

Ethical considerations are an important component of any medical research project, especially when dealing with sensitive health data such as cervical cancer images. This project strictly adheres to ethical standards to ensure transparency, confidentiality and responsible use of data. Although a publicly available dataset was used, the risks associated with breaching confidentiality were also minimized, although no direct contact with patients or personal information was involved in completing this project.

However, careful attention was paid to maintaining confidentiality and ensuring that the dataset was used for academic and research purposes only. Another key ethical aspect is the integrity of the research process, especially when all data set preprocessing, model training and evaluation steps were conducted in a transparent manner to avoid data manipulation. All external references were used appropriately to maintain academic integrity.

Since the system is designed to support medical diagnosis, ethical responsibility is the ability to prevent over-reliance on the model without human supervision. The developed framework was intended to assist pathologists, recognizing the risks of misclassification and emphasizing the importance of expert validation in clinical applications. Finally, the potential social and ethical implications of all these systems were considered in order to enable early and accurate diagnosis of cervical cancer and this study aims to contribute positively to healthcare accessibility.

5.2.4 Sustainability Plan

The development of a cervical cancer detection system is essential to ensure that the solution continues to provide value beyond the initial development phase. Since it is based on continuous learning, it is designed with measurement accuracy and long-term usability in mind. The system is capable of being updated with new data sets as more cervical cancer images become available, allowing for continued improvements in accuracy and longevity. By incorporating these techniques, the model can adapt to evolving diagnostic needs and remain relevant in medical applications.

The system uses open-source libraries like TensorFlow, karas, and OpenCV, which are

widely supported by the research and development communities, to ensure the stability of the logic. This reduces the dependency on expensive proprietary tools and ensures future compatibility.

From a healthcare perspective, the project supports social and medical sustainability by reducing the diagnostic burden on pathologists and enabling early detection of cervical cancer in limited settings. The automated system can serve as a decision support tool that ensures timely intervention and reduces mortality.

Financial and operational sustainability is maintained through potential partnerships with healthcare organizations and government initiatives focused on cancer prevention. The system is positioned as a cost-effective solution that increases diagnostic accuracy. It can be deployed in hospitals and on telemedicine platforms. Regular updates, continuous monitoring and feedback from healthcare professionals over time will ensure that the system remains accurate, reliable and effective in the fight against cervical cancer.

5.3 Project Management and Financial Analysis

- Project Timeline:
    A. Phase 1: Project Planning and Research (1 week)
    B. Phase 2: Reference Paper Collection (1-2 weeks)
    C. Phase 3: Paper Review (1-2 weeks)
    D. Phase 4: Data Collection (1-2 weeks)
    E. Phase 5: Data Analysis (1-2 weeks)
    F. Phase 6: Data Preprocessing (1-2 weeks)
    G. Phase 7: Model Implement (1-2 weeks)
    H. Phase 8: Model Evaluation (1-2 weeks)
    I. Phase 9: Prototype Design (1 weeks)
    J. Phase 10: Front End Development (1-2 weeks)
    K. Phase 11: Deployment & Testing (1-2 weeks)
- Resource Planning:

A. Equipment and Tools:

- Development and Testing
- Collaboration Tools
- Testing Tools
- Version Control System

B. Software and Technologies:

- Front-End Technologies: HTML, CSS, JavaScript, or React
- Back-End Technologies: Node.js, Django, or Flask
- Server Hosting: AWS, Azure, Google Cloud or Hostinger
- Security Software and SSL Certificates

The cost table is given below:

Table 5.1: Cost Estimated Table

SN	Components	Estimated Cost (BDT)
1	Tools and Equipment	4500-6000
2	Software	5000-7000
3	Data Collection and Processing	2500-3000
4	Documentation and Report Writing	1500-2000
5	Contingency (10% of total)	1000- 1500
Total Estimated Cost		13500-19000

5.4 Complex Engineering Problem

The research undertaken in this thesis constitutes a complex engineering problem of ""at the boundary of deep learning and medical imaging, requiring sophisticated analytical skills, the management of highly complex and often incomplete data sets, and the development of novel computational methods.

5.4.1 Complex Problem Solving

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of ApplicableCodes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓	✓	✓	✓	✓

Mapping with Knowledge Profile for EP1

Table 5.3: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K3 Comprehension	K8 Research Literature
✓	✓	✓	✓	✓	✓	✓	✓

5.4.1.1 Justification for EP Attributes Mapping

- EP1 - Depth of Knowledge Required:

  This project demonstrates EP1 by achieve K3, K4, K5, K6 & K8. The EP1 project integrates advanced deep learning models with learning techniques that require strong knowledge of medical imaging data pre-processing and neural network architecture.

- EP2 - Range of Conflicting Requirements:

EP2 Balancing accuracy and computational efficiency also requires a solution, with data set imbalance and overfitting issues creating conflicting requirements.

- EP3 - Depth ofAnalysis:

  EP3 was analyzed in depth, including pre-processing and comparative performance of eight individual models and development of hybrid ensemble models.

- EP4 - Familiarity of Issues:

  EP4 was a problem with overfitting optimizer selection and batch size tuning, but had to be carefully addressed in the medical domain.

- EP5 - Extent of ApplicableCodes:

  EP5 is a python, TensorFlow, karas model implementation with strict standards for international standards in medical research.

- EP6 - Extent of Stake-holder Involvement:

  EP6 is a system aimed at supporting pathologists, oncologists and healthcare professionals, and has also shaped the objectives of their demand model.

- EP7 - Interdependence:

  The EP7 project brings together multiple domains, particularly medical imaging, computer vision, deep learning, and healthcare applications, showing strong interdependencies between these areas.

5.4.1.2 Justification for Knowledge Profile Mapping:

- K1 - Natural Science:

  K1 is a natural science that studies the principles of applied biology and medical science, such as cervical cell image analysis.

- K2 – Mathematics:

  K2 is mathematics that analyzes statistical probability optimization and loss functions used in training models.

- K3 - Engineering Fundamentals:

  K3 is an engineering fundamentals course that covers the fundamentals of applied computer science and AI, especially neural networks, convolutional backpropagation.

- K4 - Specialist Knowledge:

  K4 is Specialist Knowledge which is a deep learning technique applied for medical image classification.

- K5 - EngineeringDesign:

  K5 is an end-to-end diagnostic system designed with model development framework and ensemble integration.

- K6 - EngineeringPractice:

  K6 Hello Engineering Practice which implements practical solutions using python,

TensorFlow, karas, OpenCV, dataset handling and evaluation pipeline.

- K7 – Comprehension:

  K7 is a model performance issue, especially overfitting optimizer effects, batch size tuning, and corrective measures have been implemented.

- K8 - Research Literature:
- K8 is a literature review on cervical cancer detection that compares with subsequent work and adopts the best performing models.

5.4.2 Engineering Activities

Table 5.4: Mapping with complex engineering activities.

EA1 Range of Re sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	✓	✓

5.4.2.1 Justification for Engineering Activities Mapping:

- EA1- Range of Resources:

  EA1 is a project that uses a variety of resources, especially medical imaging datasets, computational resources, and domain knowledge from computer vision.

- EA2 - Level of Interaction:

  EA2 is essential for multidisciplinary interactions such as deep learning in computer science, cervical cancer diagnosis in healthcare, and preprocessing and augmentation in data science.

- EA3 – Innovation:

  EA3 is an innovation that proposes a hybrid ensemble model that achieves the highest accuracy, outperforming individual D-CNNs and transfer learning models.

- EA4 - Consequences for society andenvironment:

  EA4 is a program that supports early detection of cervical cancer, reduces mortality, and provides direct social benefits by providing affordable solutions for limited healthcare environments.

- EA5 – Familiarity:

  EA5 is a deep learning approach that has been applied to the new and challenging domain of medical image classification.

5.5 Summary

The current work developed an interpretable multimodal system for the automatic detection and diagnosis of cervical cancer using deep learning. The work began with the generation of a well-balanced dataset of approximately 25,000 Pap smear images categorized into five classes: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate. Preprocessing techniques such as resizing, normalization, contrast correction, rotation, and flipping were employed to enhance image quality and model generalizability. Data augmentation was also utilized to balance class and improve the strength of the dataset.

Several baseline convolutional neural networks, including VGG19, Xception, InceptionV3, ConvNeXtBase, and MobileNetV2, were fine-tuned and evaluated to establish strong baselines for cytology image classification. Among the above, MobileNetV2 was the best, with an accuracy of 96% in its original configuration and 97% following fine-tuning. On this basis, a hybrid model integrating MobileNetV2 with a Vision Transformer (ViT) was introduced to integrate both local morphological feature extraction and global contextual representation, with an accuracy of 99.28%.

To extend the system to image-only and beyond, a multimodal ViT–BERT model was developed, where cytology image features are combined with structured clinical metadata like HPV status, cytology scores, and patient history. The model achieved state-of-the-art with 99.63% accuracy. To achieve clinical trust, explainable AI methods like Grad-CAM and LIME were utilized for detection of diagnostically significant cell regions and metadata features that influence the predictions.

Overall, the combination of CNN baselines, a hybrid vision model, and a multimodal ViT–BERT architecture demonstrated that integration of heterogeneous data sources using interpretable mechanisms is capable of significantly improving diagnostic credibility and accuracy. The findings suggest the potential of the constructed framework as an interpretable, scalable, and clinically useful solution for cervical cancer detection, particularly in resource-poor healthcare environments.

Chapter 6

Conclusion

6.1 Summary

We put forth an explainable multimodal cervical cancer detection and classification system based on cytology images and clinical metadata. A balanced Pap smear image dataset of approximately 25,000 images, comprising five clinically meaningful classes, was preprocessed and augmented for model robustness improvement. Multiple convolutional neural networks, including VGG19, Xception, InceptionV3, ConvNeXtBase, and MobileNetV2, were benchmarked for establishing baseline performance. MobileNetV2 achieved top among CNN models with 96–97% accuracy.

Then, a hybrid ViT–MobileNetV2 model was presented, with local feature extraction and global contextual representation, with 99.28% accuracy. To enhance the performance and clinical usefulness, a multimodal ViT–BERT model was presented, with the addition of cytology image features and structured clinical metadata of HPV status and cytology scores. This achieved state-of-the-art performance with 99.63% accuracy. Interpretability was incorporated through Grad-CAM and LIME, highlighting diagnostically important regions and features, thereby enabling clinical trust. Overall, the study demonstrates that the integration of vision transformers with multimodal data fusion and explainable AI can significantly improve diagnostic performance, robustness, and transparency for cervical cancer diagnosis.

6.2 Limitation

The Defence Project tasks have been spread over the year 2025 in a linear and chronological manner to ensure a logical sequence of work and timely completion. The project begins in January with Title Selection, which is scheduled to be finished by mid-February. This is followed by the Paper Collection and Review phase between February and April, allowing sufficient time for collection and understanding of existing literature related to the research focus.
Once the literature base is built, Data Collection between May and June is followed by the Data Analysis phase between June and mid-July, with the intent of interpreting effectively the data that has been collected. Data Preprocessing is planned from mid-July to early

August, during which data is processed for modeling tasks. Model Implementation is conducted from late July to early September, during which the system core or model is put into practice. Finally, Model Evaluation is planned from August to September, with sufficient time to analyze the model's performance and results.
This phased assignment of activities ensures that each stage gets sole time, minimizing overlaps and allowing a smooth transition from one stage to another. It also allows for better management of resources and maintains a contained focus throughout the project's development process.

6.3 Future Work

Subsequent research would be required to take the architecture further with more sophisticated architectures such as Swin Transformers or other hierarchical vision models to further improve representation learning. Multimodal integration to genetic markers, lab tests, and patient history may further support improved diagnostic accuracy and clinical usefulness. Explainable AI utilization is still a concern; therefore, robust interpretation methods would have to be developed to provide clearer, clinically significant information regarding model decisions. Another potential area involves taking the framework and bringing it to actual uses, such as low-weight mobile or web-based diagnostic tools, to disseminate it to healthcare practitioners in low-resource environments. Integration onto telemedicine platforms would even further increase its reach, enabling earlier diagnosis and intervention. In summary, building on this work further could be beneficial for mortality rate reduction through scalable, accurate, and transparent diagnosis for cervical cancer across the world.

.

References

[1] M. Arbyn, E. Weiderpass, L. Bruni, et al., "Estimates of incidence and mortality of cervical cancer worldwide in 2020," Lancet Glob. Health, vol. 10, no. 2, pp. e191–e203, 2021.

[2] Z. Petersen, A. Jaca, T. G. Ginindza, et al., "Barriers to uptake of cervical cancer screening services in LMICs: A systematic review," BMC Women's Health, vol. 22, p. 486, 2022.

[3] M. Pankakoski, T. Sarkeala, A. Anttila, and S. Heinävaara, "Effectiveness of cervical testing in and outside a screening program," Cancers, vol. 14, p. 5193, 2022.

[4] M. Arbyn, et al., "Accuracy and effectiveness of HPV mRNA testing in cervical cancer screening: A systematic review and meta-analysis," Lancet Oncol., vol. 23, pp. 950–960, 2022.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[6] O. Russakovsky, J. Deng, H. Su, et al., "ImageNet large scale visual recognition challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

[7] H. C. Shin, H. R. Roth, M. Gao, et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1285–1298, 2016.

[8] A. Howard, M. Sandler, G. Chu, et al., "Searching for MobileNetV3," arXiv preprint arXiv:1905.02244, 2019.

[9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Representations (ICLR), 2021.

[10] I. Pacal, B. Ozdemir, J. Zeynalov, and H. Gasimov, "A novel CNN-ViT-based deep learning model for early cervical cancer diagnosis," Biomed. Signal Process. Control, vol. 104, p. 107627, 2025.

[11] R. Ahmed, G. Dahy, A. E. Hassanien, et al., "Multimodal cervical cancer classification using deep transfer learning," IEEE Access, vol. 12, pp. 157838–157853, 2024.

[12] R. R. Selvaraju, M. Cogswell, A. Das, et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 618–626, 2017.

[13] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Towards medical transparency," Inf. Fusion, vol. 71, pp. 82–115, 2021.

[14] Abinaya, "A hybrid 3D CNN–ViT–KELM architecture for cervical cancer classification," J. Biomed. Imaging Res., 2024.

[15] A. Ghoneim, et al., "Cervical cancer classification using CNN and ELM on Herlev dataset," Comput. Biol. Med., vol. 120, p. 103737, 2020.

[16] L. Ghouti, et al., "Mobile-based cervical cancer detection using a lightweight dual-CNN approach," IEEE Access, vol. 8, pp. 12345–12352, 2020.

[17] S. H. Park, et al., "Deep learning versus traditional ML models in cervicography image classification," Diagnostics, vol. 11, no. 5, p. 800, 2021.

[18] R. Devi, et al., "Artificial Neural Network and Genetic Algorithm for cervical cytology analysis," Health Inf. Sci. Syst., vol. 10, p. 9, 2022.

[19] S. Tripathi, et al., "Transfer learning models for Pap smear image classification," Biomed. Signal Process. Control, vol. 65, p. 102363, 2021.

[20] Y. Liao, et al., "Attention-based CNN for cervical cancer diagnosis using cytology images," Appl. Sci., vol. 11, no. 11, p. 5211, 2021.

[21] K. Muksimova, et al., "RL-Cervix.Net: A Reinforcement Learning and CNN-based hybrid model for cervical cell classification," J. Med. Imaging Health Inform., 2025.

[22] T. Roy, "A rule-based hybrid expert system for cervical cancer detection in low-resource settings," Int. J. Comput. Appl. Technol., 2025.

[23] A. Arora, et al., "SVM classifier with CNN features for cervical cell classification," J. Imaging, vol. 7, no. 1, p. 5, 2021.

[24] M. A. William, et al., "Enhanced cervical cell classification using fuzzy logic and CNN," Artif. Intell. Med., vol. 97, pp. 69–79, 2019.

[25] T. Nguyen, et al., "Multimodal fusion using Vision Transformers and SHAP for explainable cervical cancer diagnosis," IEEE Trans. Med. Imaging, 2025.

[26] H. Mehedi, et al., "CCanNet: Cervical cancer classifier combining handcrafted features and CNNs," Comput. Biol. Med., 2024.

[27] K. Himabindu, et al., "A Swin Transformer-based multimodal fusion network for cervical cancer prediction," IEEE J. Biomed. Health Inform., 2025.

[28] M. H. Siddiqui, et al., "Web-based CNN with Grad-CAM visualizations for cervical cancer diagnosis," J. Med. Syst., 2025.

[29] R. Austin and R. Parvathi, "Whole-slide image classification for cervical

13% INTERNET SOURCES

9% PUBLICATIONS

13% STUDENT PAPERS

213-15-4283

ORIGINALITY REPORT

20 % SIMILARITY INDEX

PRIMARY SOURCES

1 Submitted to Daffodil International University
Student Paper

2 dspace.daffodilvarsity.edu.bd:8080
Internet Source

3 Abu Hanzala, Tanjila Akter, Md. Sadekur Rahman. "A hybrid approach for cervical cancer detection: Combining D-CNN, transfer learning, and ensemble models", Array, 2025
Publication

4 Submitted to CSU Northridge
Student Paper

5

6

7

8

9

10

Submitted to United International University	Soloukey, Geert Springeling, Adriaan J. Flikweert et al. "Miniaturized Four-Dimensional Functional Ultrasound for Mapping Human Brain Activity", Cold Spring Harbor Laboratory, 2025 Publication	4%
	Submitted to University of Sydney Student Paper	3%
		2%
		2%
		1%
downloads.hindawi.com	Internet Source	<1%
		<1%
babylonjo.atu.edu.iq	Internet Source	<1%
avesis.hacettepe.edu.tr	Internet Source	<1%
Luuk Verhoef, Sadaf		

		<1%
11	jieee.a2zjournals.com Internet Source	<1%
12	Submitted to University of Hull Student Paper	<1%
13	"Medical Image Computing and Computer Assisted Intervention – MICCAI 2025", Springer Science and Business Media LLC, 2026 Publication	<1%
14	www.pccoer.com Internet Source	<1%
15	Erdal Özbay, Feyza Altunbey Özbay. "Interpretable pap-smear image retrieval for cervical cancer detection with rotation invariance mask generation deep hashing", Computers in Biology and Medicine, 2023 Publication	<1%
	Submitted to University of Finance – Marketing Student Paper	
16	Submitted to University of Nottingham Student Paper	<1%
17	repository.uantwerpen.be Internet Source	<1%
18	www.nature.com Internet Source	<1%
19	onlinelibrary.wiley.com Internet Source	<1%
20	www.science.gov Internet Source	<1%
21		<1%

22	www.coursehero.com Internet Source	<1%
23	Bui Thanh Hung, M. Sekar, Ayhan Esi, R. Senthil Kumar. "Applications of Mathematics in Science and Technology - International Conference on Mathematical Appplications in Science and Technology", CRC Press, 2025 Publication	<1%
24	arxiv.org Internet Source	<1%
25	pmc.ncbi.nlm.nih.gov Internet Source	<1%
26	Submitted to Dublin Business School Student Paper	<1%
27	Sheng Xu, Hanlin Chen, Kexin Liu, Jinhu Lii, Baochang Zhang. "Efficient Block Pruning Based on Kernel and Feature Stablization", 2019 Digital Image Computing: Techniques and Applications (DICTA), 2019 Publication	<1%
28	kops.uni-konstanz.de Internet Source	<1%
29	Submitted to University of Hertfordshire Student Paper	<1%
30	www.pe.org.pl Internet Source	<1%
31	Meghaditya Roy, Rasna Kumari, Pooja, Sahil et al. "Deciphering therapeutic potential of phytoconstituents for mitigating cervical cancer: A pharmacognostic approach", Pharmacological Research - Natural Products, 2025 Publication	<1%

32	Submitted to East Delta University Student Paper	<1%
33	Submitted to Sharda University Student Paper	<1%
34	Submitted to University of Portsmouth Student Paper	<1%
35	ethesisarchive.library.tu.ac.th Internet Source	<1%
36	d.docksci.com Internet Source	<1%
37	dspace.aust.edu Internet Source	<1%
38	ecancer.org Internet Source	<1%
39	eprint.innovativepublication.org Internet Source	<1%
40	Farhat, Syeda Lamima. "A Vision Transformer Based Assistive System for Dermatological Diagnosis in Systemic Lupus Erythematosus.", Minnesota State University, Mankato Publication	<1%
41	assets.researchsquare.com Internet Source	<1%
42	repositorio.ufmg.br Internet Source	<1%
43	S. Prasad Jones Christydass, Nurhayati Nurhayati, S. Kannadhasan. "Hybrid and Advanced Technologies", CRC Press, 2025 Publication	<1%
44	Yadav, Ankur R.. "Toward Generalizable Models for Medical Imaging: Leveraging	<1%

	Topological Data Analysis and Deep Learning", The University of Texas at Dallas Publication	
45	"Pattern Recognition", Springer Science and Business Media LLC, 2025 Publication	<1%
46	Miguel Ossandon, Ben Prickril, Avraham Rasooly. "Cancer Detection and Diagnosis - A Handbook of Emerging Technologies", CRC Press, 2025 Publication	<1%
47	Nguyen, Quoc H.. "A Robust Data-Driven Framework for Artificial Intelligent Systems", University of South Florida, 2024 Publication	<1%
48	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	
	Tselios, Dimitrios. "Combining Deep Learning, Handcrafted Features, and Metadata for the Classification of Dermoscopy Images", University of Piraeus (Greece), 2024 Publication	<1%
49	Submitted to UCL Student Paper	<1%
	escholarship.org Internet Source	
50	jisem-journal.com Internet Source	<1%
51	koreascience.or.kr Internet Source	<1%
52		<1%
53		<1%

54	medinform.jmir.org Internet Source	<1%
55	openaccess-api.cms-conferences.org Internet Source	<1%
56	www.bignerds.com Internet Source	<1%
57	www.journal.esrgroups.org Internet Source	<1%
58	www.mdpi.com Internet Source	<1%

21% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Crossref database
- Crossref posted content database

Match Groups

174 Not Cited or Quoted 21%
Matches with neither in-text citation nor quotation marks

3 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

15% Internet sources
6% Publications
18% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

21% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

21 AI-generated only **21%**
Likely AI-generated text from a large-language model.

0 AI-generated text that was AI-paraphrased **0%**
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI mo may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that ma longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlight percentage shown.