Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

DIVISION OF COMPUATIONAL BIOLOGY

MSc. Med (by dissertation) in Bioinformatics

Enhancing Detection of Cervical Cancer Through Deep Learning: A Comparative study of Histological Image-Based Algorithms

Palesa Tjale

Student number: TJLPAL002

PS Number: 1889961

TJLPAL002@myuct.ac.za

University of Cape Town

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In fulfilment of the requirements for this degree

Supervisor: Dr. Sinkala Musalula – musalula.sinkala@uct.ac.za, Department of Integrative Biomedical Sciences, University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Declaration

I, *Palesa Tjale*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Table of Contents

Acknowledgements i
LIST OF ABBREVIATIONS ii
LIST OF TABLES ii
LIST OF FIGURES iv
Thesis Organisation v
ABSTRACT vi
CHAPTER 1: INTRODUCTION 1
1.1. AIMS AND OBJECTIVES 2
CHAPTER 2: LITERATURE REVIEW 4
2.1. Deep Learning in Medical Imaging 5
Image Pre-processing Techniques 6
Model Training and Evaluation 6
2.2. Visual Prompting Techniques 7
2.3. Comparison of Deep Learning and Visual Prompting Techniques 8
2.4. Advancements in Imaging and detection Techniques 10
2.5. Importance of High-Quality Imaging for Accurate Diagnosis 11
2.6. Related work 12
CHAPTER 3: METHODOLOGY 16
3.1. Implementation Workflow 16
3.2. Data Collection 16
3.3. Image Pre-processing 17
3.4. Environment Setup 17
3.5. Model Development 18
3.6. Visual Prompting Using LandingLens 19
3.7. Explainability and Evaluation Metrics 20
CHAPTER 4: RESULTS 21
4.1. Dataset Overview 21
4.2. Quantitative Results 22
Performance of CNN Models 23
CNN Models Performance on ROC Analysis 28
Visual Prompting Model Performance on ROC Analysis 30
4.3. Model Explainability using Grad-CAM 31
4.4. Comparative Summary 40
CHAPTER 5: DISCUSSION 41
5.3. Limitations 45

5.4. Future Directions 45
CHAPTER 6: CONCLUSION 46
REFERENCES 47

Acknowledgements

First and foremost, I want to thank God. Your grace, strength, and guidance have carried me through this journey, giving me the perseverance to see it through. I could not have done this without You.

To my supervisor, Dr. Musalula Sinkala—thank you for your unwavering support, patience, and guidance. Your insights and encouragement have shaped not only this research but also my growth as a researcher. I truly appreciate your dedication and the time you invested in helping me navigate this project.

Kuhle Mcinga and Sibongiseni Msipa, thank you both for being my pillars of strength. Your constant encouragement, understanding, and belief in me have made all the difference. I am beyond grateful for your emotional support throughout this journey.

I also want to express my sincere appreciation to the organizations that made this work possible. UCT Master's Financial Aid, thank you for supporting me during my first year and easing the financial burden of my studies. Ada & Bertie Levenstein Bursary Award and Marion Beatrice Waddel Bursary Award, your generosity in my second year allowed me to focus on my research without worry. I am truly grateful for your support.

Finally, to everyone who has been part of this journey—whether through encouragement, advice, or simply being there when I needed it—thank you. Every bit of support has meant the world to me, and I will always be grateful.

LIST OF ABBREVIATIONS

- AI – Artificial Intelligence
- AUC – Area Under the Curve
- CCD – Charge-Coupled Device
- CNNs – Convolutional Neural Networks
- DL – Deep Learning
- DNA – Deoxyribonucleic Acid
- Grad-CAM – Gradient-weighted Class Activation Mapping
- HPV – Human Papillomavirus
- LMICs – Low- and Middle-Income Countries
- ML – Machine Learning
- MVP – Medical Visual Prompting
- NLP – Natural Language Processing
- PMs – Pre-trained Models
- ROC – Receiver Operating Characteristic
- SVM – Support Vector Machine
- TL – Transfer Learning
- WSIs – Whole-Slide Images

LIST OF TABLES

Table 1: Summary of recent studies employing deep learning techniques for cervical cancer detection. The table highlights the methods used, datasets analysed, and key performance metrics, including accuracy, sensitivity, specificity, and AUC _____21
Table 2: Comparison of accuracy, macro-average precision, recall, and F1-score for SqueezeNet, EfficientNet, ResNet50, and the Visual Prompting Model._____26

LIST OF FIGURES

Figure 1: Cell images of five categories: (a) Superficial-Intermediate, (b) Parabasal, (c) Koilocytotic, (d) Dyskeratotic, (e) Metaplastic _____21
Figure 2: Confusion matrix illustrating the classification performance of ResNet50 for cervical cell classification._____24
Figure 3:Confusion matrix illustrating the classification performance of SqueezeNet for cervical cell classification._____25
Figure 4:: confusion matrix illustrating the classification performance of EfficientNet for cervical cell classification._____26
Figure 5:Confusion matrix illustrating the classification performance of the Visual Prompting Model for cervical cell classification._____27
Figure 6: ROC curve illustrating the AUC value for the ResNet50 model in cervical cell classification. 28
Figure 7:ROC curve illustrating the AUC value for the SqueezeNet model in cervical cell classification. _____29
Figure 8: ROC curve illustrating the AUC value for the EfficientNet model in cervical cell classification. _____30
Figure 9: ROC curve illustrating the AUC value for the Visual Prompting Model in cervical cell classification._____31
Figure 10:: Grad-CAM heatmap illustrating the regions of interest focused on by the ResNet50 model in cervical cell classification The SqueezeNet model (Fig. 11) demonstrated better localization compared to ResNet50, with Grad-CAM heatmaps showing more concentrated focus on cell regions. This improvement was particularly evident for Koilocytotic and Metaplastic cells, where the heatmaps were more precise. However, some subtle features were still not fully captured, indicating room for improvement. _____34
Figure 11:Grad-CAM heatmap illustrating the regions of interest focused on by the SqueezeNet model in cervical cell classification._____36
Figure 12:Grad-CAM heatmap illustrating the regions of interest focused on by the EfficientNet model in cervical cell classification._____38
Figure 13:Grad-CAM heatmap illustrating the regions of interest focused on by the Visual Prompting Model in cervical cell classification_____40

Thesis Organisation

This thesis is organized into six chapters:

- **Chapter One** provides a general background on cervical cancer, including its epidemiology, current diagnostic methods, and the role of artificial intelligence in medical imaging. It also introduces the aims and objectives of this study.
- **Chapter Two** presents a detailed literature review on deep learning applications in histopathological image analysis. It explores conventional CNN-based approaches and introduces visual prompting techniques as an emerging method for enhancing model interpretability and accuracy.
- **Chapter Three** focuses on the methodology used in this study. It describes the dataset, image preprocessing techniques, and the experimental setup. Additionally, it details the development and evaluation of ResNet50, SqueezeNet, EfficientNet, and the Visual Prompting Model, along with the explainability techniques employed, such as Grad-CAM.
- **Chapter Four** presents the results obtained from model training and evaluation. It includes performance comparisons among the models based on accuracy, sensitivity, specificity, and F1-score. The chapter also provides insights from Grad-CAM heatmaps to assess the interpretability of model predictions.
- **Chapter Five** discusses the findings in relation to existing research, highlighting the advantages of visual prompting in medical AI applications. It also addresses the study's limitations and suggests areas for future research, including potential improvements in AI-assisted cervical cancer diagnostics.
- **Chapter Six** concludes the thesis by summarizing key findings and their implications for clinical practice. It also outlines potential research directions for advancing AI-based cervical cancer screening.

ABSTRACT

Cervical cancer remains a leading cause of cancer-related mortality among women worldwide, particularly in low- and middle-income countries (LMICs) where access to screening is limited. Early detection is crucial for improving patient outcomes; however, traditional diagnostic methods, such as Pap smears and histological examinations, often suffer from variability, subjectivity, and limited sensitivity. Recent advancements in artificial intelligence (AI), particularly deep learning (DL) and visual prompting techniques, present promising opportunities to enhance the accuracy, efficiency, and interpretability of cervical cancer diagnosis from histology images.

In this thesis, I investigate the application of DL models—ResNet50, SqueezeNet, EfficientNet, and a Visual Prompting Model—for classifying cervical cells using histopathological images. I conduct a comparative analysis to evaluate these models based on accuracy, sensitivity, specificity, and interpretability. To enhance model explainability, I employ Grad-CAM to visualize model decisions, offering insights into the diagnostic relevance of highlighted features. My results indicate that the Visual Prompting Model outperforms conventional DL models, achieving the highest accuracy (98%) and F1-score (0.99) while also demonstrating superior localization of diagnostically significant regions. EfficientNet follows closely with an accuracy of 97% and an F1-score of 0.97, while SqueezeNet achieves 95% accuracy and an F1-score of 0.95. In contrast, ResNet50 shows lower performance, with an accuracy of 91% and an F1-score of 0.91, indicating limitations in feature extraction and localization.

A key finding of my study is that integrating visual prompting significantly enhances model explainability, addressing a critical challenge in AI-driven medical imaging. By directing attention to clinically relevant areas within histological images, visual prompting reduces misclassification rates, potentially aiding pathologists in making more informed diagnostic decisions. Additionally, the computational efficiency and ease of training of Visual Prompting Models suggest their feasibility for deployment in resource-constrained settings where expert pathology review is limited.

Overall, my findings underscore the transformative potential of AI, particularly visual prompting, in improving cancer detection. These AI-assisted diagnostic tools promise not only to enhance accuracy but also to improve interpretability, making them highly relevant for

clinical integration. I suggest that future research should focus on validating these AI models across diverse clinical settings, optimizing computational efficiency, and exploring hybrid AI approaches that incorporate molecular and genomic data for a more comprehensive approach to cervical cancer diagnostics.

Keywords: Cervical Cancer, Deep Learning, Visual Prompting, Histopathological Imaging, Explainable AI

CHAPTER 1: INTRODUCTION

Cervical cancer is a leading cause of cancer-related mortality among women, primarily affecting the cervix. The development of cervical cancer is closely linked to persistent human papillomavirus (HPV) infection, viral replication, and the growth of infected cells within the cervical transformation zone [1]. Early detection and effective treatment of cervical cancer significantly improve the chances of recovery [2]. However, current diagnostic methods, primarily relying on histological and morphological examinations, lack well-defined sensitivity and specificity. Regular screening procedures such as Pap smears, colposcopy, and HPV testing are employed to detect cervical cancer, emphasizing the importance of early detection through routine check-ups, as the disease often remains asymptomatic in its initial stages [3].

Histopathological images, including Pap smear images, offer valuable information for diagnosing and classifying cervical cancer. Machine learning (ML) techniques provide an opportunity to leverage these images, enabling more accurate and reliable detection, diagnosis, and monitoring of cervical cancer[1]. The digitization of histopathological images facilitates the extraction of quantitative data, such as cell size, tissue abnormalities, and abnormal cell counts, which are essential for diagnostic purposes [4].

In the realm of medical imaging, ML, and particularly DL algorithms, have demonstrated promise in predicting early disease symptoms. ML is a multidisciplinary field that integrates principles from cognitive psychology, neuroscience, computational science, and statistics to mine data and make performance predictions[5]. Researchers continuously strive to develop advanced models capable of analysing medical data to extract hidden knowledge, thereby supporting diagnostic decision-making processes[6]. While ML presents powerful tools for addressing challenges in the medical field, it is essential to acknowledge potential errors and biases inherent in the models, like any other medical research or practice.

DL, a subset of ML, utilizes neural networks with multiple layers to learn complex data representations. DL has been successfully applied in various studies to predict cervical cancer, yielding encouraging outcomes. Medical applications frequently utilize DL methods to process diverse data input formats, including 3D data, which align with existing architecture capabilities[7]. A distinctive advantage of DL is its ability to autonomously learn relevant features without preselected variables or human assumptions, discerning significant variables through the training process.

Visual prompting is a technique that reformulates the downstream visual classification task using prompt learning methods. It draws inspiration from the success of prompt learning in the field of natural language processing (NLP) and aims to achieve task consistency between pre-training and downstream applications in computer vision[8]. The theoretical relevance of visual prompting lies in its potential to leverage the knowledge encoded in large-scale pre-trained models (PMs) for improving diagnostic accuracy and decision-making processes in computer vision tasks. By reformulating the visual classification task as a generative masked visual token modelling task, visual prompting enables the exploitation of pre-trained model knowledge with the same objectives as in pre-training tasks. This allows for a more effective utilization of the information captured by PMs, leading to potentially enhanced performance in downstream applications[8].

By providing real-time visual cues and highlighting potentially relevant areas within the histological images, visual prompting algorithms can assist pathologists in their assessments and reduce the likelihood of overlooking important diagnostic clues[9]. Moreover, visual prompting can aid in decision-making processes by providing explanations or justifications for the model's predictions[10]. The use of verbalizers helps establish connections between the predicted visual tokens and the downstream labels, allowing for interpretability and transparency in the decision-making process. This can be particularly valuable in critical applications where human experts need to understand and validate the model's predictions before making informed decisions[11].

Here, Iaim to harness the combined potential of DL and visual prompting techniques to address the limitations of traditional histological analysis. By evaluating and comparing the performance of DL models such as ResNet50, SqueezeNet, EfficientNet, and a visual prompting algorithm, Iseek to determine the most effective and clinically applicable approach for the cervical cancer detection. Overall, the results of this study will provide a framework and proof-of-concept for designing computer-aided diagnostic systems, that enhance both diagnostic efficiency and accuracy and the interpretability of histology image diagnosis, thus ultimately improving patient outcomes.

1.1. AIMS AND OBJECTIVES

Main Aim

To develop and evaluate deep learning and visual prompting algorithms for cervical cancer diagnosis, identifying the most effective model for clinical application based on model performance characteristics and interpretability.

Specific Objectives

1. To process histological images for deep learning applications by implementing image pre-processing techniques that enhance quality and address challenges such as noise, staining variability, and resolution inconsistencies.

2. To train deep learning models for cervical cancer classification by developing and evaluating ResNet50, SqueezeNet, EfficientNet, and a Visual Prompting Model for categorizing cervical cells into Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate types.

3. To compare deep learning model performance using multiple metrics by assessing accuracy, sensitivity, and specificity through precision, recall, F1-score, and confusion matrices.

4. To evaluate model interpretability for clinical use by investigating the effectiveness of visual prompting algorithms and DL in generating interpretable outputs using Grad-CAM to aid pathologists in clinical decision-making.

CHAPTER 2: LITERATURE REVIEW

Cervical cancer continues to be a major global health concern, ranking as the fourth most common cancer among women worldwide, with an estimated 570,000 new cases and 311,000 deaths reported annually[12,13].Although advancements in screening and vaccination have significantly reduced the incidence of cervical cancer in developed countries, the disease remains a leading cause of cancer-related deaths in LMICs, which bear nearly 85% of the global cervical cancer burden[13,14]. The widespread lack of access to screening programs and HPV vaccination in LMICs contributes to the persistent high incidence and mortality rates in these regions[13].

Cervical cancer is primarily caused by persistent infection with high-risk HPV strains, particularly HPV-16 and HPV-18, which are responsible for approximately 70% of cases[14] .Vaccination against HPV and the early detection of pre-cancerous lesions through regular screening, including Pap smears and HPV testing, can significantly reduce the burden of cervical cancer[12]. Nevertheless, millions of women, especially in under-resourced regions, remain vulnerable due to limited access to healthcare services and education[14].

The early detection of cancer, including those of the cervix, is vital for reducing mortality and improving patient outcomes. When detected at early stages, the five-year survival rate of cervical cancer patients exceeds 90%, but this figure drops significantly in advanced cases, falling to below 20% in some instances[13],[14].Screening methods such as the Pap test and HPV testing have proven to be highly effective in identifying pre-cancerous lesions. HPV DNA testing has shown superior sensitivity compared to traditional cytology-based tests, allowing for earlier and more accurate identification of high-risk individuals[12]. However, the lack of infrastructure and resources in LMICs often means that women present at later stages, making treatment more difficult and less effective[14]. Therefore, expanding access to affordable and accurate screening programs is critical for improving survival rates[12].

Future research should focus on developing and evaluating the novel screening methods namely, hybrid models that combine traditional techniques with advanced imaging and machine learning, as well as self-testing methods such as high-risk HPV self-tests[15]. These approaches hold promise for increasing screening participation and improving early detection rates. Moreover, longitudinal studies assessing the long-term effectiveness of these innovative

methods in real-world settings will be essential to establish their utility in routine clinical practice.

2.1. Deep Learning in Medical Imaging

The field of medical imaging has witnessed significant advancements with the integration of AI and DL. AI refers to the development of computational models that mimic human intelligence, enabling machines to perform tasks such as image recognition, decision-making, and predictive analysis[16]. AI encompasses several subfields, with ML being one of the most influential. ML models use statistical algorithms to learn patterns from data and make predictions without explicit programming. In medical imaging, ML has been applied to tasks such as tumour detection, disease classification, and segmentation of medical images. Traditional ML methods often rely on feature engineering, where domain experts manually extract relevant features such as cell texture, shape, and intensity variations for analysis[17].

DL, a specialized branch of ML, has transformed medical imaging by eliminating the need for manual feature extraction. Neural Networks (NNs) are a class of ML models inspired by the human brain, consisting of layers of interconnected neurons that process information[18]. These networks are particularly useful for medical imaging, as they can detect patterns, classify abnormalities, and enhance image resolution. DL models utilize multi-layered NNs to extract hierarchical feature representations directly from raw image data, making them highly effective for applications such as histopathological image analysis[16]. Among the various DL architectures, Convolutional Neural Networks (CNNs) have become the gold standard for image analysis in medical imaging. CNNs are specifically designed for pattern recognition in images, making them ideal for histopathological image classification and segmentation[19].

Popular CNN architectures such as AlexNet, VGGNet, ResNet, and more advanced models like DenseNet have been applied successfully in cervical cancer diagnosis[20,21]. These models consist of convolutional layers for feature extraction, followed by pooling layers for dimensionality reduction, and fully connected layers for final classification[22]. Another promising approach is ensemble learning, where multiple deep learning models are combined to improve performance. Techniques such as CervixNet, which uses a combination of CNNs and hierarchical learning strategies, have shown excellent results in classifying cervical cancer images[23].

In addition to CNNs, transfer learning has been a key technique in medical image analysis, allowing models trained on large datasets to be fine-tuned for cervical cancer detection with

smaller datasets[20,22]. Transfer learning enhances model performance by leveraging pre-trained weights, which speeds up training and increases accuracy, particularly when the available data is limited[24]. This approach significantly improves performance by leveraging pre-trained feature representations, thereby reducing training time and increasing classification accuracy, particularly in data-limited medical applications.

Image Pre-processing Techniques

Image pre-processing is a critical step in improving the performance of deep learning models in cervical cancer detection. Pre-processing techniques such as resizing, normalization, and augmentation are applied to ensure that the input data is consistent and suitable for model training[25]. For instance, resizing ensures that all images have the same dimensions, which is crucial for batch processing in CNNs. Normalization adjusts pixel intensity values, typically scaling them to a range between 0 and 1, which accelerates model convergence during training[24].

Data augmentation techniques, such as flipping, rotation, and zooming, are used to artificially increase the size of the training dataset. This helps to mitigate overfitting, a common issue when training deep learning models on small datasets, as is often the case in medical image analysis[26]. In one study, an ensemble pre-processing approach that combined techniques like histogram equalization and image sharpening significantly boosted the accuracy of cervical cancer segmentation models[25].

The effectiveness of image pre-processing is evident in studies where these techniques have led to marked improvements in model performance. For example, in a CNN-SVM-based cervical cancer detection system, pre-processing increased the accuracy of classification from 94% to over 99%[26].

Model Training and Evaluation

Training deep learning models for cervical cancer detection typically involves backpropagation and optimization techniques such as stochastic gradient descent and Adam optimizer to minimize a loss function[20,27]. During training, the model iteratively adjusts its weights based on the error it makes on the training data. A key aspect of training is early stopping, which prevents overfitting by halting the process when the model's performance on a validation set begins to deteriorate.

Common performance metrics used to evaluate deep learning models in cervical cancer detection include accuracy, sensitivity, specificity, precision, and the F1 score[26,27]. Sensitivity (recall) measures the ability of the model to correctly identify positive cases (cancerous), while specificity measures its ability to correctly identify negative cases (non-cancerous). The F1 score is a harmonic mean of precision and recall, providing a balanced metric for models dealing with imbalanced datasets, which is often the case in medical imaging[26,27].

Despite the high accuracy of deep learning models, they are not without limitations. CNNs, for instance, require large amounts of data to train effectively, and their performance can degrade when applied to data with significant variation from the training set. This is where transfer learning and ensemble models become particularly useful, as they help mitigate these issues by leveraging pre-trained weights and multiple models[20].

2.2. Visual Prompting Techniques

Visual prompting is a technique originally inspired by prompt engineering in NLP and adapted for computer vision. Visual prompting provides cues, or "prompts," within images to guide pre-trained models on specific tasks without modifying the model itself, often making use of minimal adjustments to the input image, such as small patches or inpainting techniques[8,28]. In medical imaging, visual prompting can assist in focusing the model's attention on clinically relevant areas, such as regions with potential abnormalities in histological images, thereby enhancing diagnostic accuracy without the need for task-specific fine-tuning[29].

Visual prompts work by altering the visual input space, such as adding structured patterns or inpainting sections of an image. This helps the model prioritize certain image sections over others, leading to more relevant feature extraction and improved interpretability[30,31]. For example, prompts can be used in histopathology to guide models in identifying tumour regions within whole-slide images, improving detection accuracy and aiding pathologists by highlighting critical areas for examination[32].

Research on visual prompting in cervical cancer diagnosis has shown promising results. One study demonstrated that visual prompts applied to histological images of cervical cells significantly improved the model's ability to classify pre-cancerous and cancerous lesions by directing attention to cellular structures indicative of malignancy[33]. Another application involves segmentation tasks where visual prompts provide guidance for delineating boundaries

of abnormalities, a crucial feature in identifying tumour spread and cellular changes associated with cervical cancer progression[29].

Comparative studies between visual prompting and traditional deep learning methods have shown that visual prompting not only enhances model interpretability without sacrificing accuracy but also offers significant advantages in ease of training. Unlike CNNs, which require extensive data, computational resources, and coding expertise, visual prompting models can be trained with fewer images, require minimal technical knowledge, and achieve competitive performance with far less computational effort. This makes them highly accessible, even to non-technical users, positioning visual prompting as a transformative approach for AI training in medical diagnostics and beyond[8,28].

The application of visual prompting in medical imaging offers several benefits for clinical decision-making. By focusing the model's attention on specific areas of interest, visual prompting reduces the cognitive load on clinicians, enhances diagnostic accuracy, and aids in identifying regions that warrant further investigation[34]. In clinical workflows, these techniques provide an additional layer of interpretability, making it easier for medical professionals to validate the model's predictions and trust the output[34].

The advantage of visual prompting over training CNN models is that it significantly simplifies the model training process, making it more accessible and efficient. Unlike traditional CNNs, which require large datasets, extensive computational resources, and technical expertise, visual prompting models can be trained with fewer images and without a deep understanding of coding. This reduces the barriers to entry, enabling even non-technical users to train and deploy AI models effectively. In addition, visual prompting enhances model interpretability, making it easier to explain and trust the results, which is crucial in medical applications. Despite these benefits, challenges remain in integrating visual prompting into clinical workflows, such as ensuring intuitive model explanations for clinicians and addressing potential increases in processing time, particularly with high-resolution histopathology images. These challenges highlight the need for continued research to balance the advantages of visual prompting with the demands of real-time diagnosis[31,32].

2.3. Comparison of Deep Learning and Visual Prompting Techniques

When evaluating deep learning models and visual prompting techniques for cervical cancer detection, key metrics such as accuracy, sensitivity, specificity, and F1 score illustrate the comparative strengths of each approach. Traditional deep learning models, including CNN-

based architectures such as ResNet, DenseNet, and VGG, have consistently shown high performance on cervical cancer datasets, often achieving sensitivity and specificity levels over 90% when trained on large datasets. For instance, DenseNet-121, fine-tuned on cervical histological images, has reached an accuracy of 98.2%, demonstrating the powerful feature extraction capabilities of deep learning models[20,23,31].

Visual prompting techniques have shown that models guided by visual prompts can achieve comparable performance to fully trained deep learning models, with studies reporting high sensitivity and specificity in tasks like tumour segmentation within whole-slide images (WSIs)[29,32]. One marked advantage of the Visual Prompting Model is its computational efficiency, as it can operate effectively with fewer parameters and limited labelled data**,** making it a viable option for resource-constrained settings[28,30].

A specific advantage of visual prompting is its adaptability to various clinical imaging tasks. For example, models with visual prompts in tumour segmentation and lesion detection have reached accuracy levels that align closely with CNN-based models, often within 1-2% variance, while reducing the training time and computational load[31,33]. This advantage makes visual prompting suitable for real-time applications where quick decision-making is essential.

The alignment with clinicians' diagnostic processes aids in building trust in AI-supported decisions. Studies have shown that visual aids in diagnostic models improve metacognitive judgment calibration, enabling clinicians to better assess diagnostic confidence and accuracy [29,34]. Traditional deep learning models, while highly accurate, often rely on abstract feature representations that require supplementary tools, like Gradient-weighted Class Activation Mapping (Grad-CAM), to explain predictions. Although these techniques improve interpretability, they add complexity, potentially limiting usability in clinical workflows[8,30].

Both DL and visual prompting techniques hold promise for clinical integration, but each presents unique implementation pathways. DL models are well-suited to high-throughput screening environments where accuracy and consistency are critical, as their robustness in performance can meet the demands of standardized clinical workflows [26]. However, these models depend on extensive labelled datasets and computational power, which may restrict their application in low-resource or small-scale settings[8,23].

Visual prompting techniques, conversely, offer flexibility and scalability, as they can be adapted to new tasks with minimal additional training, making them ideal for settings with limited data. Their ability to highlight diagnostically relevant areas also makes them

advantageous in clinical decision support, where interpretability and clinician interaction with model outputs are prioritized. Visual prompting's resource efficiency further allows for use in diverse healthcare environments, from small clinics to large hospitals, where real-time decision-making support is valuable[29,33].

In summary, while DL models remain the standard for high-accuracy diagnostic tasks, visual prompting offers a promising alternative with comparable accuracy, greater interpretability, and resource-efficient adaptability. Future developments in both areas could lead to hybrid models that leverage the strengths of deep learning's feature extraction with the interpretability and efficiency of visual prompting, potentially offering optimal solutions for real-world cervical cancer screening and diagnostics[31,34].

2.4. Advancements in Imaging and detection Techniques

Recent advancements in imaging techniques, combined with the application of AI /ML for image analysis, offer promising avenues to enhance cervical cancer detection. Many methods have been proposed that utilize CNNs for cervical cell classification directly from Pap smear images, bypassing traditional segmentation challenges[35].

The integration of DL into histopathological image analysis has transformed the landscape of cervical cancer diagnosis. For instance, recent studies have emphasized the potential of computer-assisted screening systems that employ digital image processing to analyse Pap smear images[36]. These systems utilize advanced algorithms for feature extraction and classification, resulting in accuracies that improve upon manual methods.

Advancements in medical technology have led to the development of automated, AI-driven methods for cervical cancer detection, offering solutions to some of the limitations of traditional screening. One of the most significant advancements is the application of deep learning models, particularly CNNs, in the analysis of cytological and histopathological images[37,38]. These models can automatically extract features from medical images, eliminating the need for manual feature selection and interpretation, which reduces variability and improves consistency in diagnostic results[27].

AI-driven approaches, such as computer-aided diagnosis systems, have been shown to significantly improve the accuracy of cervical cancer detection by analysing whole-slide

images (WSIs) and detecting subtle patterns that may be missed by human observers[36,39]. Studies have demonstrated that CNN-based systems can achieve high sensitivity and specificity, outperforming traditional methods like Pap smears in some cases[40]. These AI systems also allow for faster processing of large datasets, enabling quicker diagnoses and reducing the burden on healthcare providers[36,37].

In addition to diagnostic accuracy, AI technologies offer the potential for more widespread implementation of cervical cancer screening in resource-limited settings. AI-guided digital colposcopy, for example, can assist less experienced clinicians by providing automated image analysis and suggesting biopsy sites, which addresses the shortage of skilled colposcopists in LMICs[41].

Overall, the transition to AI-based diagnostic systems represents a significant step forward in cervical cancer detection. By enhancing the precision, speed, and scalability of screening processes, these technologies hold promise for improving early detection and patient outcomes, particularly in underserved populations[37].

2.5. Importance of High-Quality Imaging for Accurate Diagnosis

The quality of imaging plays a vital role in the diagnostic process. High-resolution images enhance the ability to detect subtle abnormalities in cervical cells, thereby improving screening outcomes. As highlighted by various studies, including those investigating deep learning techniques for image classification, the accuracy of cervical cancer detection is closely tied to the quality of the input images[42,43].

Moreover, the development of biomarkers for primary screening represents an innovative step towards enhancing the diagnostic process. This shift from traditional cytology towards biomarker-based screening could complement existing methods and lead to more reliable and efficient detection strategies[44].

Future research should focus on developing and evaluating the novel screening methods namely, hybrid models that combine traditional techniques with advanced imaging and machine learning, as well as self-testing methods such as high-risk HPV self-tests[15]. These approaches hold promise for increasing screening participation and improving early detection rates. Moreover, longitudinal studies assessing the long-term effectiveness of these innovative

methods in real-world settings will be essential to establish their utility in routine clinical practice.

In conclusion, the field of cervical cancer detection is evolving rapidly, with advancements in imaging techniques and machine learning presenting exciting opportunities for improving diagnostic accuracy and patient outcomes. Continued research and innovation are essential to bridge existing gaps and ensure that the benefits of these technologies are realized in diverse healthcare contexts.

2.6. Related work

Table 1:Summary of recent studies employing deep learning techniques for cervical cancer detection. The table highlights the methods used, datasets analysed, and key performance metrics, including accuracy, sensitivity, specificity, and AUC

Paper	**Method**	**Data**	**Results**
Supervised deep learning embeddings for the prediction of cervical cancer diagnosis[45]	Reviews various CNN architectures for image segmentation and classification of pap smear and colposcopy images	Medical records of 858 patients from Hospital Universitario de Caracas in Caracas	Top area under the curve (AUC) = 0.6875
Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images[42]	VGG19 (TL) model, and CYENET	5679 colposcopy photographs obtained from the cervical screening data collection by	The classification accuracy for VGG19 was 73.3%. CYENET exhibited high sensitivity,

Paper	Method	Data	Results
		Intel and Smartphone ODT	specificity, and kappa scores of 92.4%, 96.2%, and 88%, respectively.
Deep Transfer Learning for Histopathological Diagnosis of Cervical Cancer Using Convolutional Neural Networks with Visualization Schemes[46]	CNN	4993 cervical histological images (2503 benign and 2490 malignant)	Achieved 95.88% sensitivity, 98.93% specificity, 97.42% accuracy, 94.81% Youden's index and 99.71% area under the receiver operating characteristic curve
Classification of Histopathological Biopsy Images Using Ensemble of Deep Learning Networks[47]	VGG19, MobileNet, and DenseNet.	the Bioimaging 2015 breast histology classification challenge and the BreaKHis dataset	Achieves an accuracy of 96.41%, 92.40%, 97.42% respectively.

Paper	Method	Data	Results
Cervix Image Classification for Prognosis of Cervical Cancer using Deep Neural Network with Transfer Learning[48]	Develops a novel deep learning model based on DenseNet-121 for cervical cancer detection using pap smear images	Kaggle dataset consisting of a total of 5278 labelled images	Achieves an accuracy of 98.2% and improves the performance of DenseNet-121 by fine-tuning
Automated Pap Smear Cervical Cancer Screening Using Deep Learning[49]	Mask R-CNN	liquid-based histological slides obtained from Thammasat University (TU) Hospital	Obtained had an accuracy of 89.8%, sensitivity of 72.5%, and specificity of 94.3%

Table 1 summarizes some of the recent papers that have applied deep learning methods for cervical cancer detection using pap smear images. The table shows the main method, data and results of each paper. The studies encompass various deep learning architectures, including CNNs, Transfer Learning (TL) models like VGG19 and DenseNet-121, and specialized models such as Mask R-CNN. These architectures are applied to diverse datasets, ranging from histological images to colposcopy photographs, obtained from different medical institutions and research datasets. The papers demonstrate that deep learning methods can achieve high accuracy and reduce the false rate in cervical cancer detection.

The variety of approaches discussed here, including ensemble models, CNNs, hybrid models, and cutting-edge architectures like DenseNet, demonstrates that researchers are looking into

several different ways to enhance cervical cancer diagnosis using deep learning. This diversity is encouraging since it shows a desire to identify the best solutions. It is promising that these models routinely show excellent accuracy rates, ranging from 96.7% to 98.6%. This shows how deep learning has the potential to increase the accuracy of cervical cancer diagnosis, which is essential for early detection and improved patient outcomes.

High accuracy is essential, but the next crucial step is to apply these models in actual clinical situations. Their practical usefulness will be determined by elements like interpretability, explainability, regulatory clearances, and integration into current healthcare operations.

This review explored the roles of deep learning and visual prompting techniques in cervical cancer detection, emphasizing their potential in overcoming the limitations of traditional screening methods. Deep learning models, particularly CNNs and transfer learning architectures, have shown remarkable diagnostic accuracy, often exceeding that of traditional methods such as Pap smears and colposcopy. These models have improved the efficiency and consistency of cervical cancer detection, reducing human error in image interpretation. Meanwhile, visual prompting has emerged as a valuable approach in enhancing model performance with minimal fine-tuning, offering computational efficiency while maintaining high accuracy by directing model attention to diagnostically relevant areas. Both techniques demonstrate unique strengths: deep learning for robust feature extraction and visual prompting for its interpretability and adaptability to new tasks with limited labelled data.

CHAPTER 3: METHODOLOGY

This section outlines the systematic approach Iadopted to address the research objectives of our study focusing on enhancing detection of cervical cancer using deep learning and visual prompting techniques.

3.1. Implementation Workflow

The workflow included dataset preparation, pre-processing, CNN model training, Grad-CAM integration, LandingLens visual prompting, and performance evaluation. This methodology provides a robust framework for achieving the study objectives and advancing cervical cancer detection techniques.

3.2. Data Collection

I obtained the histological images for this study from the SIPaKMed Dataset, a publicly available cervical cell classification dataset hosted on Kaggle (https://www.kaggle.com/datasets/prahladmehandiratta/cervical-cancer-largest-dataset-sipakmed). The dataset consists of 5,014 images**,** categorized into five classes**:** Dyskeratotic (1,035), Koilocytotic (1,063), Metaplastic (1,064), Parabasal (895), and Superficial-Intermediate (957)**.**

Superficial-Intermediate cells, the most abundant cell type in Pap smear samples, are mature squamous epithelial cells**.** In contrast, Parabasal cells are immature squamous epithelial cells, significantly smaller than other epithelial cells in cervical smears. Metaplastic cells result from the transformation of columnar epithelial cells into squamous cells, a normal physiological process in the cervix. However, abnormal metaplasia can indicate early precancerous changes, warranting further examination. Koilocytotic cells**,** which are a hallmark of HPV infection**.** These cells are characterized by a perinuclear halo, hyperchromatic nuclei, and nuclear irregularities**.** The final category, Dyskeratotic cells**,** represents a more advanced pathological state**.** These cells exhibit abnormal keratinization, irregular nuclear contours, and a high nuclear-to-cytoplasmic ratio**,** indicating malignant transformation**.**

To ensure a balanced and structured dataset for model training and evaluation, I divided the SIPaKMed Dataset into training (3,003 images; 60%)**,** validation (1,003 images; 20%)**,** and testing (1,008 images; 20%) **s**ubsets[50]. Additionally, I utilized PyTorch's DataLoader to

efficiently batch and shuffle the data, optimizing performance during model training and evaluation[51].

3.3. Image Pre-processing

Pre-processing is a crucial step in deep learning-based image analysis as it standardizes input data, reduces variability, and enhances feature extraction, thereby improving model performance. To ensure uniformity and compatibility with DL models, I resized all images to 224 × 224 pixels, a commonly used input dimension CNNs, which balances computational efficiency with sufficient spatial resolution for feature extraction[52].

Following resizing, I normalized pixel values to a range of [0,1]**,** ensuring consistent intensity values across images and accelerating convergence during model training to enhance dataset diversity and reduce overfitting, I applied data augmentation techniques such as random rotation, horizontal flipping, zooming, and contrast adjustment. I chose to utilise these transformations because they introduce variability while preserving essential morphological features, effectively improving the model's generalization ability.

To further improve image clarity, I applied Gaussian filters for noise reduction. I chose the utilise the Gaussian filtering, because the approach is widely used in medical image processing to smooth images while preserving edges, effectively removing high-frequency noise that may obscure relevant structures[53]. By blurring irrelevant pixel variations, this step enhances the visibility of diagnostically significant regions, leading to improved feature extraction and classification accuracy.

3.4. Environment Setup

I conducted all experiments using Jupyter Notebook (Python 3.11.5), utilizing PyTorch as the primary deep learning framework. PyTorch provides efficient tensor computation, automatic differentiation, and dynamic computational graphs, making it highly suitable for deep learning applications[51]. I trained and evaluated all models in a GPU-enabled environment to ensure accelerated computations and efficient processing. For image preprocessing, including reading, resizing, and augmentation, I employed OpenCV, a widely used open-source computer vision library. OpenCV is optimized for real-time applications and provides a broad range of functions

for image processing, motion tracking, and object detection. OpenCV has been extensively used in various domains, including medical imaging, facial recognition, and automated detection systems[54]. In my image processing computation pipeline, I specifically used OpenCV to process histological images before feeding them into DL models, ensuring consistent formatting and enhancing model robustness. To improve model interpretability, I applied Grad-CAM – this method generates visual heatmaps that highlight the most influential regions in an image, providing insights into how CNN models make classification decisions[55]. Grad-CAM has been particularly valuable in medical imaging where the technique allows researchers and clinicians to verify whether the model is focusing on relevant diagnostic features rather than non-informative artifacts. Thus I used Grad-CAM by overlaying activation maps on input histology images, to enhances the interpretability of my AI-based diagnostics approach and helps identify potential biases or limitations in the model.

To evaluate the applicability of visual prompting for cervical cancer classification from histology images, I utilized LandingLens to implement visual prompting, enhancing model interpretability by directing attention to diagnostically relevant regions [citation]. I uploaded histological images and manually annotated key areas of interest to generate a dataset optimized for visual prompting. Using LandingLens's automated pipeline, I trained a classification model that learned to emphasize these annotated features, improving its ability to differentiate cervical cell types while ensuring greater transparency in the decision-making process through interpretable model outputs.

3.5. Model Development

I selected Three CNNs, ResNet50, SqueezeNet, and EfficientNet for comparative analysis in the classification of cervical cell types. These models were implemented using the PyTorch deep learning framework, leveraging transfer learning to enhance feature extraction. To improve generalization, I preprocessed and augmented the dataset before training I resized training images to 224 × 224 pixels, applied random flipping, and normalized them using ImageNet's mean and standard deviation. I resized the validation images to 256 pixels, center-cropped them to 224 × 224 pixels, and normalized the data accordingly. I split the dataset into training (60%) and validation (40%) subsets and loaded them into PyTorch's DataLoader with a batch size of 32, balancing computational efficiency and model performance.

For ResNet50, I fine-tuned a pre-trained model with ImageNet weights, replacing the fully connected layer with a custom classifier tailored to the five-class cervical cell classification task. I trained the model using cross-entropy loss and Adam optimizer with a learning rate of 0.001, running for 50 epochs with early stopping based on validation accuracy. Training involved a forward pass, loss computation, and backpropagation to update model weights, with the best-performing model saved for testing. I applied similar fine-tuning techniques to SqueezeNet, modifying its classifier while leveraging its lightweight architecture for faster training. I introduced dropout layers to prevent overfitting and performed grid search optimization to fine-tune learning rates and batch sizes. For EfficientNet, I incorporated depth-wise convolutions and squeeze-and-excitation layers for optimized feature extraction. I replaced the classifier head and introduced batch normalization to stabilize training. To further enhance performance, I conducted hyperparameter tuning via grid search, optimizing learning rates, batch sizes (16, 32, and 64), and training epochs (50–100). I used the Adam optimizer across all models due to its adaptive learning rates and convergence stability. I evaluated model performance using validation accuracy, weighted F1-score, and confusion matrices. Additionally, I employed Grad-CAM visualizations to interpret model decisions and highlight the most influential regions in classification.

3.6. Visual Prompting Using LandingLens

To integrate visual prompting, I utilized the LandingLens platform in classification mode, implementing manual annotations to guide the model toward key histological features. First, I uploaded cervical cell images to the platform and annotated diagnostically relevant regions. These annotations were incorporated into the training process, directing the model's focus toward significant features.

The rationale for using LandingLens lies in its ability to streamline the visual prompting process, offering an interactive and automated environment to mark, train, and validate models without requiring extensive manual feature engineering. This approach improves reliability and explainability, ensuring that the model focuses on clinically relevant structures rather than irrelevant background artifacts. After training, I compared the LandingLens model's outputs against CNN-based models to assess consistency and classification effectiveness. Unlike traditional CNNs, which rely purely on learned representations, visual prompting allows for a more human-in-the-loop approach, improving model reliability by incorporating expert knowledge directly into the training process. This enhances explainability, making it easier to

validate whether the model is focusing on clinically relevant regions rather than irrelevant background artifacts.

Once trained, the outputs from the LandingLens Visual Prompting model were compared against those from Python-implemented CNN models – ResNet50, SqueezeNet, and EfficientNet – to evaluate consistency and effectiveness. The results helped me assess whether integrating visual prompting into deep learning pipelines could improve interpretability and accuracy for cervical cancer classification, potentially making AI-assisted diagnosis more transparent and clinically useful.

3.7. Explainability and Evaluation Metrics

Interpretable AI is essential in medical imaging to ensure that deep learning models make decisions based on meaningful features rather than irrelevant artifacts. To improve transparency and trust in model predictions, Grad-CAM was employed to generate class activation maps, which highlight the image regions most influential in each classification decision[56].

By overlaying heatmaps on the original inputs, I used Grad-CAM to qualitatively assess model decision-making, verifying whether the networks focused on diagnostically relevant structures when classifying images. This approach provided valuable insights into both the strengths and limitations of the learned representations, allowing me to make targeted refinements to the model architecture and training process.

I conducted a comparative analysis to systematically evaluate baseline CNN models against those enhanced with visual prompting. I assessed model performance using accuracy, precision, recall, and F1-score, while Grad-CAM visualizations provided additional context on how each model localized crucial features. The integration of quantitative performance metrics and qualitative visualization outcomes demonstrated the effectiveness of visual prompting in guiding the networks toward diagnostically relevant regions, ultimately improving both classification accuracy and model interpretability.

CHAPTER 4: RESULTS

In this study, I evaluated the performance of four deep learning models ResNet50, SqueezeNet, EfficientNet, and the Visual Prompting Model in classifying cervical cells into five categories**:** Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate**.** The models were assessed using accuracy, precision, recall, F1-score, confusion matrices, and receiver operating characteristic (ROC) curves to determine their classification effectiveness. Additionally, Grad-CAM heatmaps were generated to evaluate model interpretability, providing insights into the decision-making process of each model.

4.1. Dataset Overview

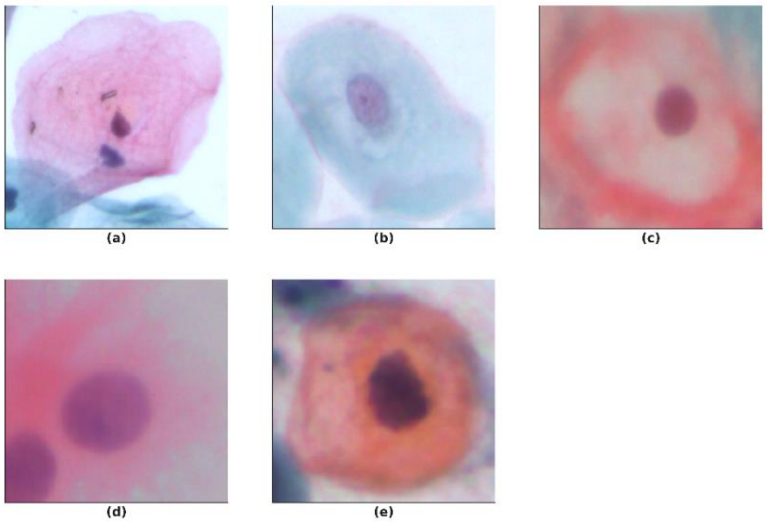

Figure 1: Cell images of five categories: (a) Superficial-Intermediate, (b) Parabasal, (c) Koilocytotic, (d) Dyskeratotic, (e) Metaplastic

The SIPaKMeD dataset is a publicly available cervical cell image database designed for automated classification of normal and pathological cells in Pap smear images. It consists of 4,049 manually cropped images of isolated cells, extracted from 966 clustered cell images of Pap smear slides. These images were acquired using a CCD camera adapted to an optical microscope and were expertly annotated to ensure diagnostic accuracy[57]. The dataset is divided into five categories:

1. Superficial-Intermediate cells (Figure 1a), which are mature squamous epithelial cells predominantly found in Pap smear samples with a well-defined polygonal cytoplasm and small condensed nucleus;
2. Parabasal cells (Figure 1b), which are immature squamous epithelial cells with a cyanophilic cytoplasm and large vesicular nucleus, often confused with metaplastic cells due to their morphological similarities;
3. Koilocytotic cells (Figure 1c), characterized by a perinuclear halo, hyperchromatic nuclei, and nuclear irregularities, with some displaying binucleation or multinucleation at different stages of viral transformation;
4. Dyskeratotic cells (Figure 1d), which represent a more advanced pathological state associated with high-grade cervical intraepithelial neoplasia or invasive squamous cell carcinoma, exhibiting abnormal keratinization and irregular nuclear contours indicative of malignant transformation; and
5. Metaplastic cells (Figure 1e), which arise from the transformation of columnar epithelial cells into squamous cells, a normal physiological process, but abnormal metaplasia may indicate early precancerous changes.

To facilitate model training and evaluation, the dataset was partitioned into training (60%) with 3,003 images, validation (20%) with 1,003 images, and testing (20%) with 1,008 images.

4.2. Quantitative Results

The Visual Prompting Model achieved the highest accuracy (98%) and F1-score (0.99), followed closely by EfficientNet (97%). SqueezeNet and ResNet50 demonstrated competitive but slightly lower performance, achieving accuracies of 95% and 91%, respectively (Table 2).

Table 2:Comparison of accuracy, macro-average precision, recall, and F1-score for SqueezeNet, EfficientNet, ResNet50, and the Visual Prompting Model.

Model	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Visual Prompting	**0.98**	**0.99**	**0.98**	**0.99**
EfficientNet	**0.97**	**0.97**	**0.97**	**0.97**
SqueezeNet	**0.95**	**0.95**	**0.95**	**0.95**
ResNet50	**0.91**	**0.91**	**0.91**	**0.91**

The confusion matrices for each model (Fig. 2-5) provide a detailed view of their classification performance across the five cervical cell categories: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate. Each matrix highlights the number of correctly and incorrectly classified instances, offering insight into specific areas of strength and weakness for each model.

Performance of CNN Models

The ResNet50 model (Fig. 2), which achieved an overall accuracy of 91%, exhibited limitations in differentiating between morphologically similar classes. Dyskeratotic cells were frequently misclassified, with 19 instances being predicted as Koilocytotic and 3 as Metaplastic. This suggests that the model struggled with overlapping features in these categories. Similarly, ResNet50 had difficulty in correctly identifying Metaplastic cells, with 19 samples incorrectly classified as Koilocytotic and 8 as Dyskeratotic. The confusion was less pronounced in the Parabasal and Superficial-Intermediate categories, where the model correctly classified most instances, though one Parabasal cell and two Superficial-Intermediate cells were still misclassified. Overall, the model faced significant challenges in distinguishing Dyskeratotic, Koilocytotic, and Metaplastic cells, leading to a higher rate of misclassification in these categories.

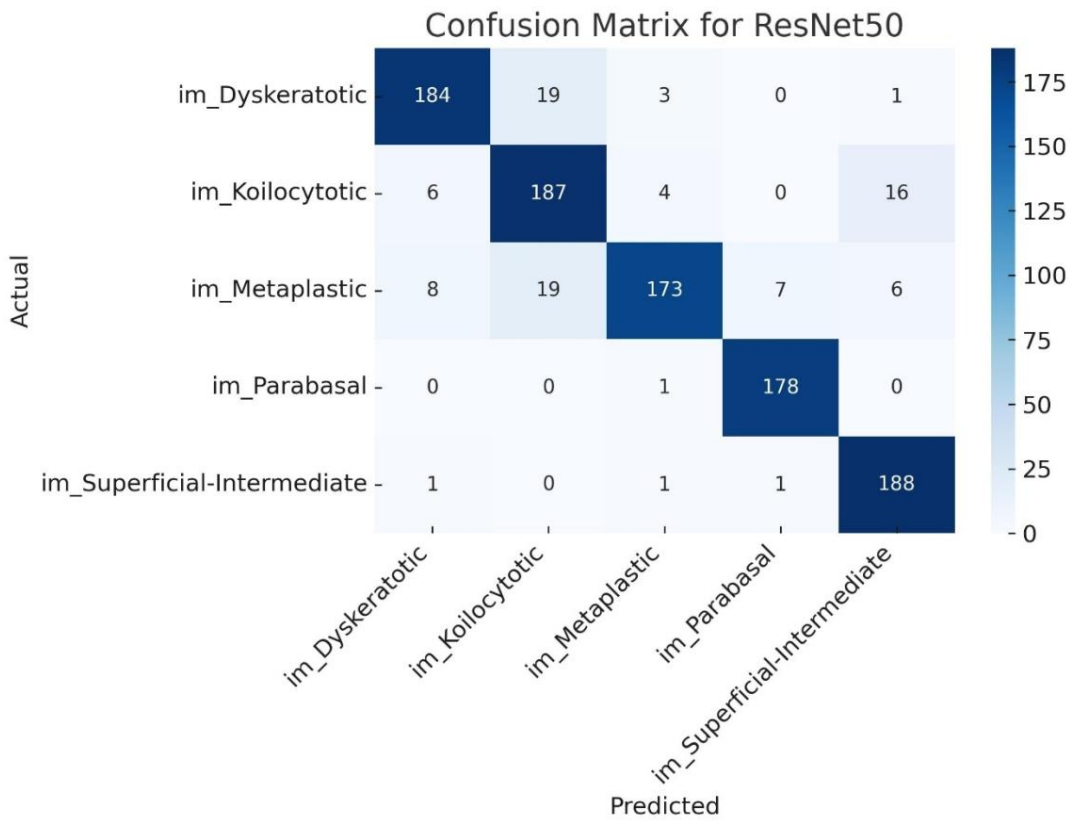

Figure 2: Confusion matrix illustrating the classification performance of ResNet50 for cervical cell classification.

The SqueezeNet model (Fig. 3), which achieved an accuracy of 95%, demonstrated improved classification performance compared to ResNet50. The frequency of misclassifications decreased, particularly for Dyskeratotic and Koilocytotic cells. Only 11 Dyskeratotic cells were misclassified as Koilocytotic, a noticeable improvement over ResNet50. However, Koilocytotic cells were still challenging, with 12 instances wrongly classified as Superficial-Intermediate. The model showed better differentiation in the Metaplastic category, with 11 samples misclassified as Koilocytotic and 5 as Dyskeratotic. In the Parabasal category, SqueezeNet exhibited strong performance, misclassifying only two instances as Superficial-Intermediate. Similarly, for the Superficial-Intermediate class, just one instance was wrongly classified as Dyskeratotic and another as Koilocytotic. These results indicate that SqueezeNet provided a more refined feature extraction compared to ResNet50 but still struggled in distinguishing Koilocytotic from Superficial-Intermediate cells.

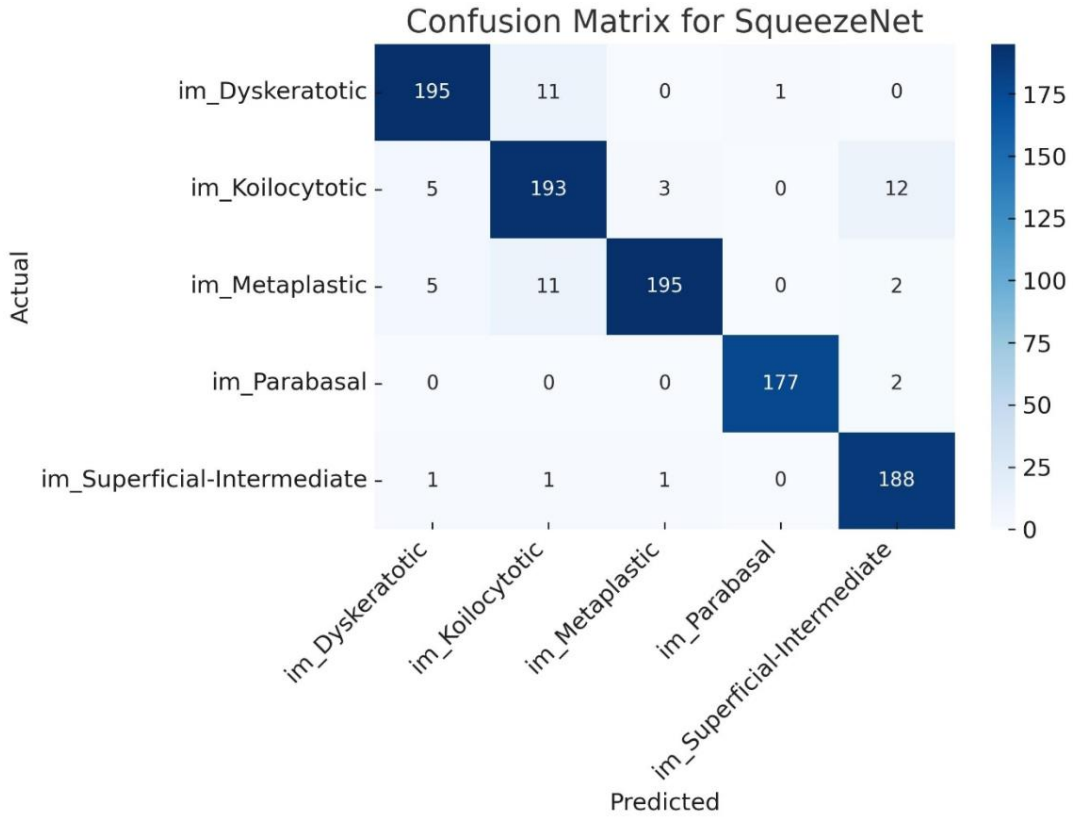

Figure 3:Confusion matrix illustrating the classification performance of SqueezeNet for cervical cell classification.

The EfficientNet model (Fig. 4), with an accuracy of 97%, outperformed both ResNet50 and SqueezeNet, demonstrating high classification precision across all categories. Misclassifications were minimal, with only nine Dyskeratotic cells misclassified as Koilocytotic and three as Superficial-Intermediate. Koilocytotic classification improved significantly, with only three instances being misclassified as Superficial-Intermediate. In the Metaplastic category, EfficientNet performed exceptionally well, with just four misclassified as Koilocytotic and two as Superficial-Intermediate. Parabasal and Superficial-Intermediate classifications were highly accurate, with only one Parabasal and two Superficial-Intermediate cells misclassified. These results indicate that EfficientNet had superior feature extraction and differentiation capabilities, reducing classification errors across all categories.

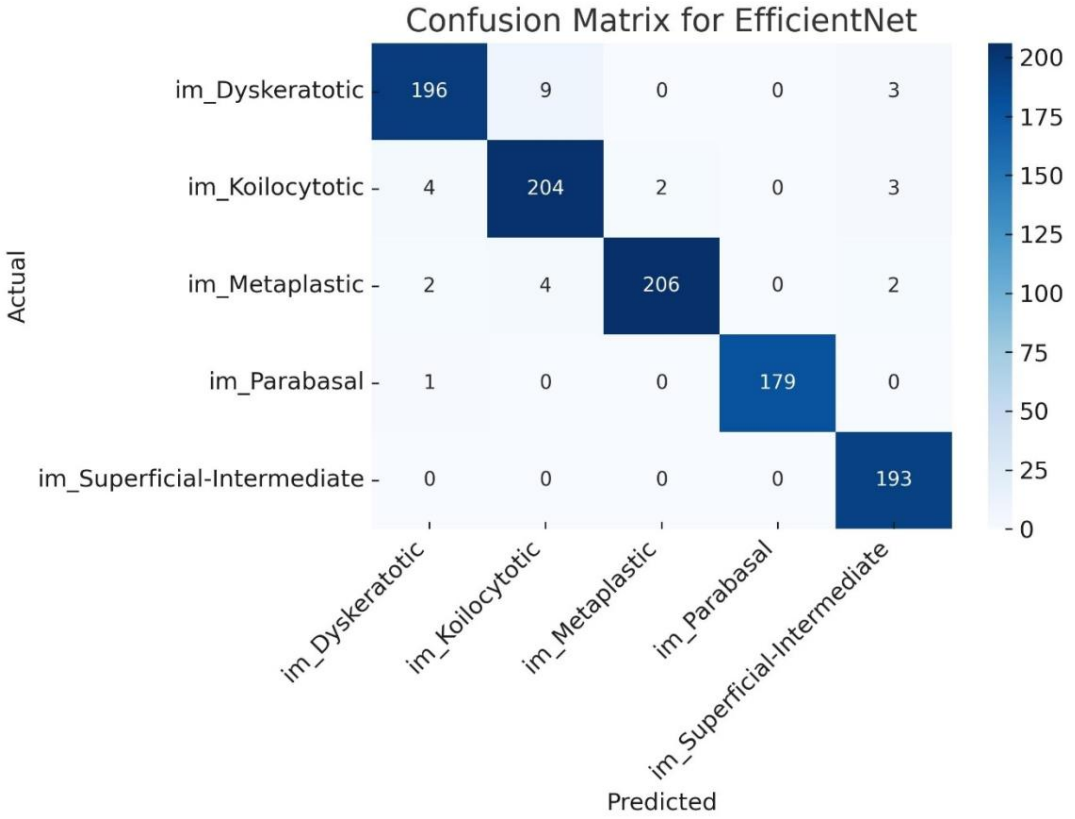

Figure 4:: confusion matrix illustrating the classification performance of EfficientNet for cervical cell classification.

The Visual Prompting Model (Fig. 5) surpassed all CNN models, achieving an accuracy of 98% and demonstrating near-perfect classification across all categories. Dyskeratotic cells, which were frequently misclassified by CNNs, were now classified with much higher precision, with only five instances misclassified as Koilocytotic. Similarly, Koilocytotic classification was highly refined, with only one sample misclassified as Superficial-Intermediate. The Metaplastic category, which had significant errors in ResNet50 and SqueezeNet, showed further improvement, with just three misclassified as Koilocytotic. Notably, the Parabasal and Superficial-Intermediate categories were classified with 100% accuracy, highlighting the model's ability to resolve classification ambiguities. The visual prompting approach provided additional guidance, enabling the model to focus on diagnostically significant features within the images, which reduced the overall misclassification rate.

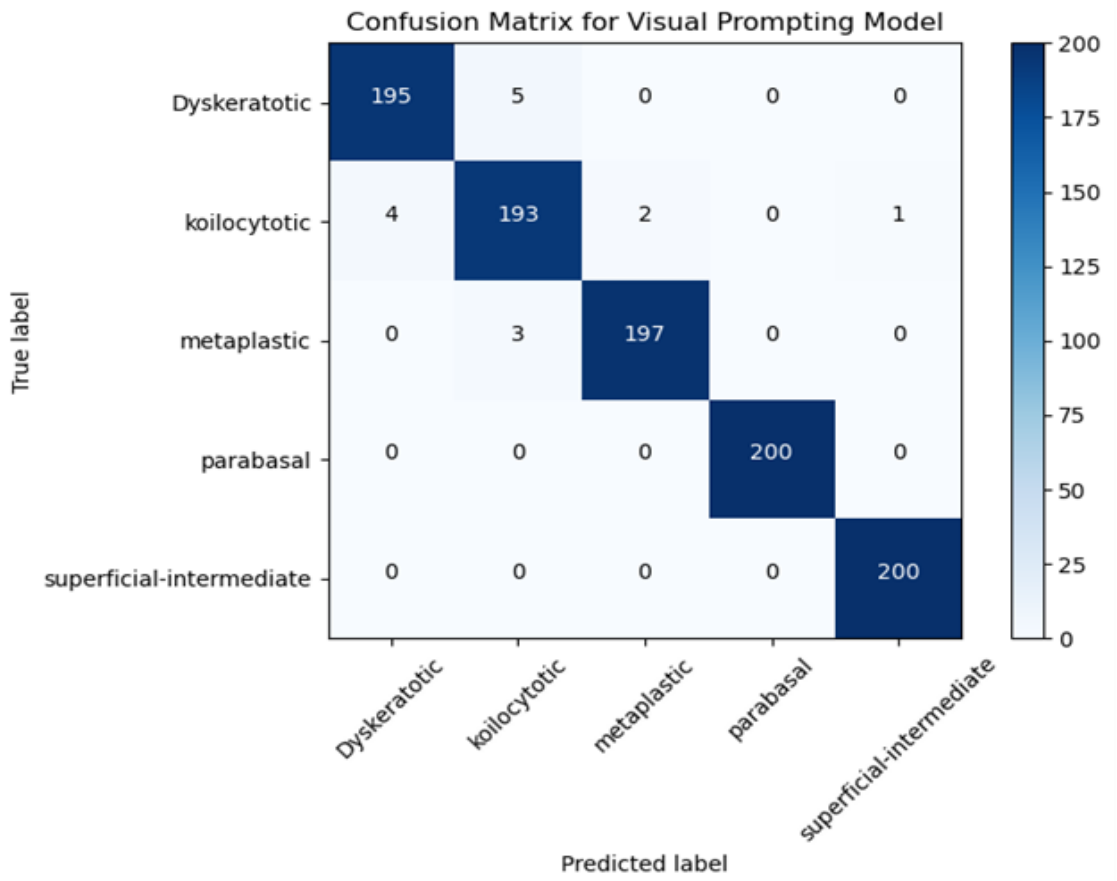

Figure 5:Confusion matrix illustrating the classification performance of the Visual Prompting Model for cervical cell classification.

The confusion matrices reveal a clear progression in model performance, with ResNet50 exhibiting the most classification challenges, SqueezeNet offering moderate improvements, and EfficientNet achieving high accuracy with minimal misclassification. The Visual Prompting Model outperformed all CNN models, demonstrating significant reductions in errors, particularly for Dyskeratotic and Koilocytotic cells, which posed the greatest challenges in previous models. These findings reinforce the effectiveness of visual prompting in guiding deep learning models to focus on clinically relevant features, making it a promising approach for enhancing cervical cancer diagnosis.

ROC curves were generated for each model (Fig. 6-9) to assess their discriminative performance across the five cervical cell categories: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate. The AUC values provide a measure of how well each

model distinguishes between the classes, with a value closer to 1.00 indicating near-perfect classification performance.

CNN Models Performance on ROC Analysis

The ResNet50 model (Fig. 6) exhibited strong classification performance, with AUC values ranging from 0.98 to 1.00. Despite an overall high AUC, individual class-level performance revealed some inconsistencies. Dyskeratotic cells had an AUC of 0.99, suggesting that the model was highly capable of distinguishing them from other classes. However, the Koilocytotic class had a slightly lower AUC of 0.98, reflecting challenges in distinguishing these cells from others, particularly Metaplastic and Superficial-Intermediate. The Metaplastic class demonstrated an AUC of 0.99, reinforcing the model's ability to differentiate this category effectively. Meanwhile, the Parabasal and Superficial-Intermediate classes achieved perfect AUC values of 1.00, indicating that ResNet50 was highly confident in classifying these two cell types.

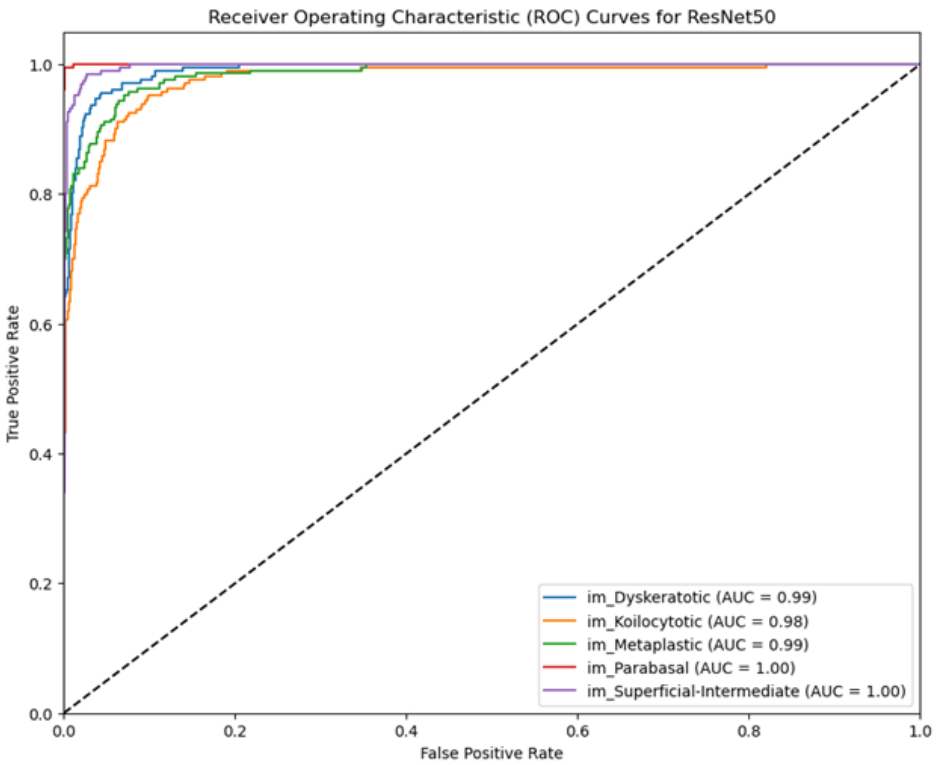

Figure 6: ROC curve illustrating the AUC value for the ResNet50 model in cervical cell classification.

The SqueezeNet model (Fig. 7), similar to ResNet50, achieved AUC values between 0.98 and 1.00. Dyskeratotic and Metaplastic cells had an AUC of 0.99, showing that the model

effectively distinguished these classes. However, the Koilocytotic class had an AUC of 0.98, indicating continued challenges in differentiating it from closely related categories. The Parabasal and Superficial-Intermediate classes, like in ResNet50, achieved an AUC of 1.00, suggesting strong confidence in their classification. Compared to ResNet50, SqueezeNet exhibited slightly more stability in distinguishing overlapping categories, but still faced minor difficulties with Koilocytotic misclassifications.

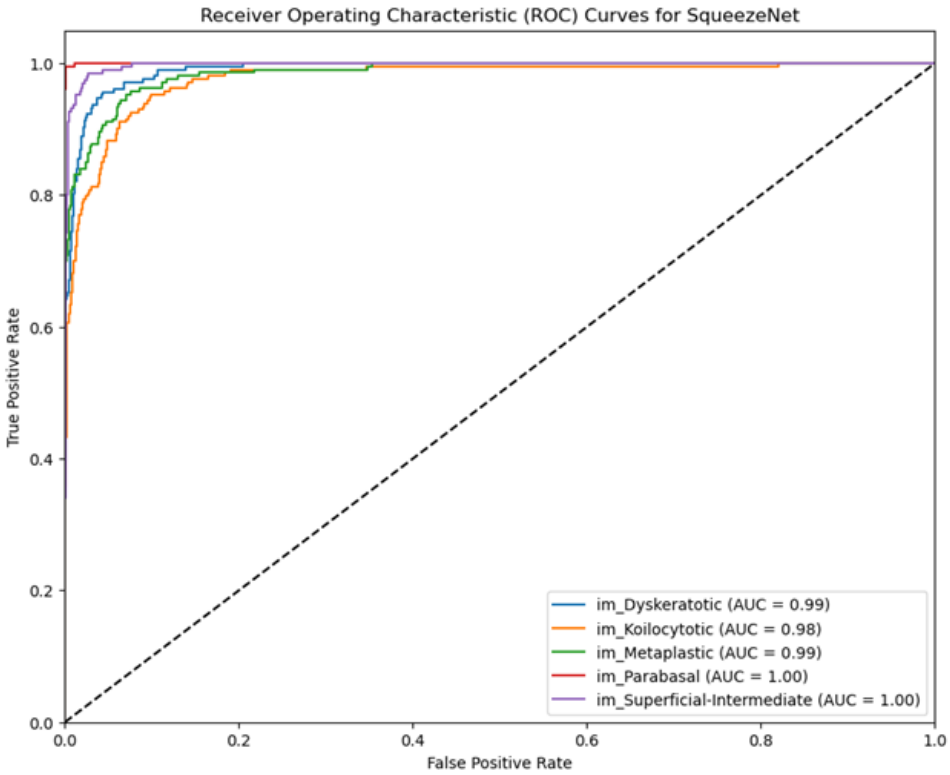

Figure 7:ROC curve illustrating the AUC value for the SqueezeNet model in cervical cell classification.

The EfficientNet model (Fig. 8) demonstrated superior classification performance, achieving perfect AUC values of 1.00 across all five classes. Unlike ResNet50 and SqueezeNet, which exhibited slight variations in AUC values, EfficientNet consistently identified Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate cells with maximum discriminative power. This result highlights EfficientNet's ability to extract meaningful features across all categories, reducing classification ambiguities observed in previous CNN models.

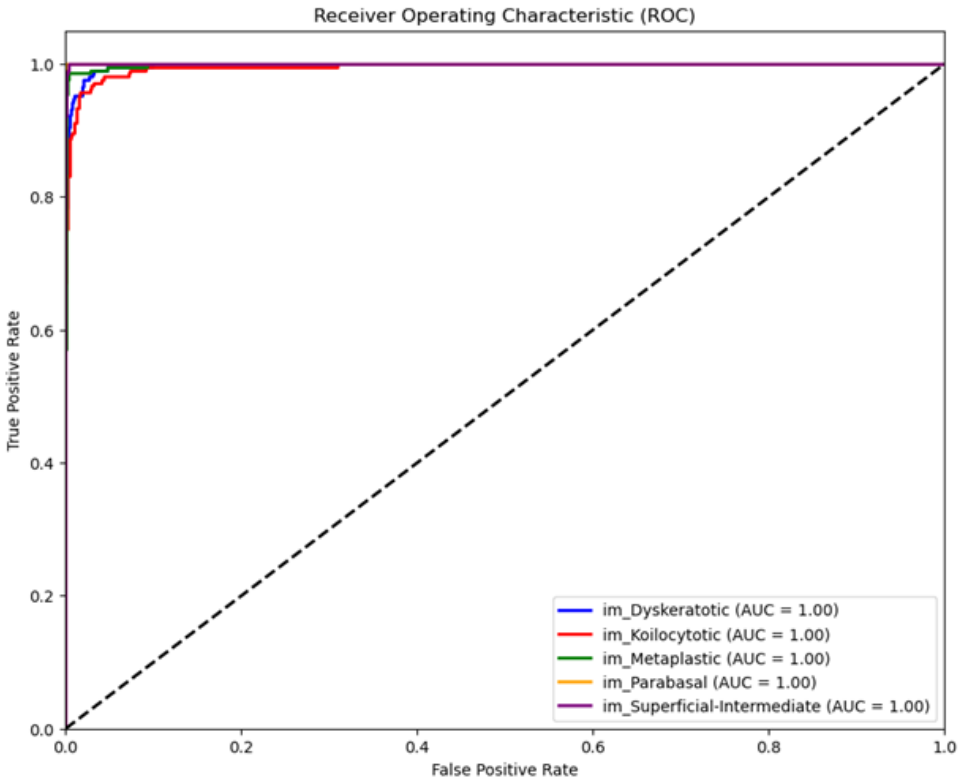

Figure 8: ROC curve illustrating the AUC value for the EfficientNet model in cervical cell classification.

Visual Prompting Model Performance on ROC Analysis

The Visual Prompting Model (Fig. 9) achieved AUC values comparable to EfficientNet, with near-perfect classification performance across all five categories. Dyskeratotic, Metaplastic, Parabasal, and Superficial-Intermediate cells were classified with an AUC of 1.00, ensuring high sensitivity and specificity in predictions. Koilocytotic cells, which had posed challenges for ResNet50 and SqueezeNet, were now distinguished with an AUC of 0.99, further reinforcing the model's ability to address misclassification errors. The Visual Prompting Model's near-perfect AUC values validate its ability to improve decision-making accuracy by focusing on diagnostically significant regions, effectively reducing false positives and false negatives.

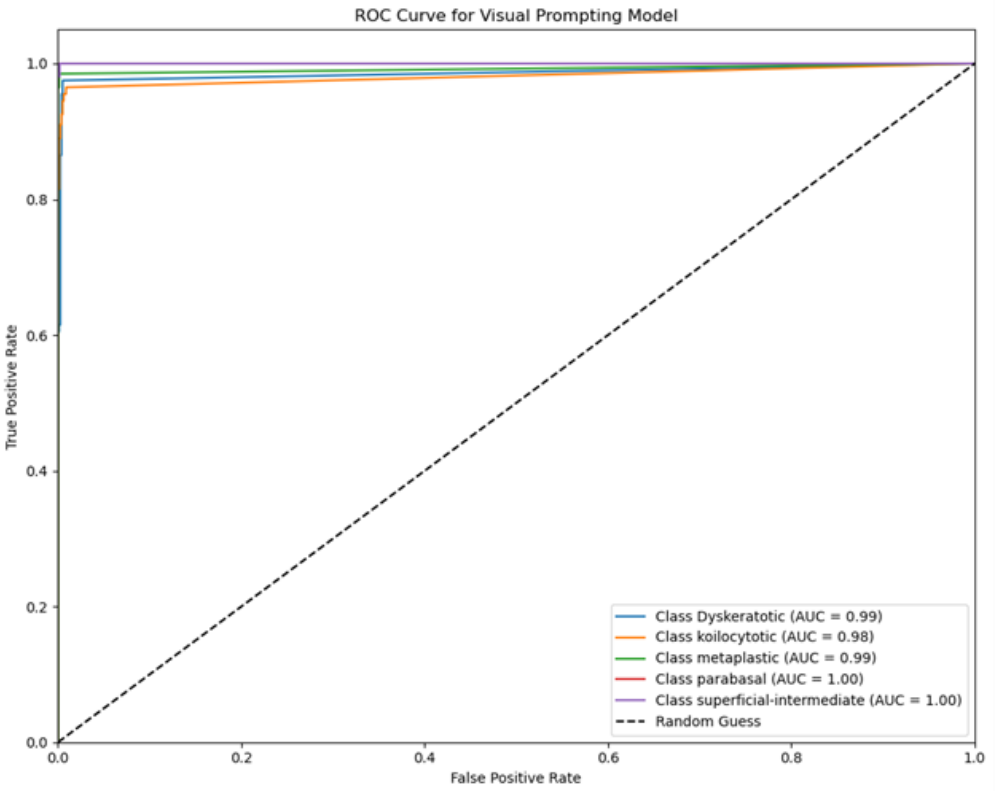

Figure 9: ROC curve illustrating the AUC value for the Visual Prompting Model in cervical cell classification.

The ROC analysis reveals a clear trend of increasing classification performance across models. ResNet50 and SqueezeNet performed well, but exhibited slightly lower AUC values for Koilocytotic cells, suggesting some difficulty in distinguishing them from other categories. EfficientNet outperformed these CNN models, achieving perfect AUC values of 1.00 across all five classes. However, the Visual Prompting Model matched EfficientNet's classification performance while providing enhanced interpretability through visual prompting, making it the most effective and reliable model for cervical cell classification. These findings reinforce the effectiveness of combining deep learning with visual prompting to enhance diagnostic accuracy and decision transparency in medical imaging.

4.3. Model Explainability using Grad-CAM

Grad-CAM heatmaps were generated (Fig. 10-13) to analyse the interpretability of each model's predictions. These heatmaps illustrate the regions of interest within the histological images that contributed most to the classification decisions.

The ResNet50 model produced Grad-CAM heatmaps (Fig. 10) that showed moderate localization of critical regions. However, some activations extended to unrelated areas,

reducing interpretability. For example, Dyskeratotic cells exhibited focus near the nucleus but scattered activations in the background, highlighting the model's struggles to consistently

identify relevant areas.

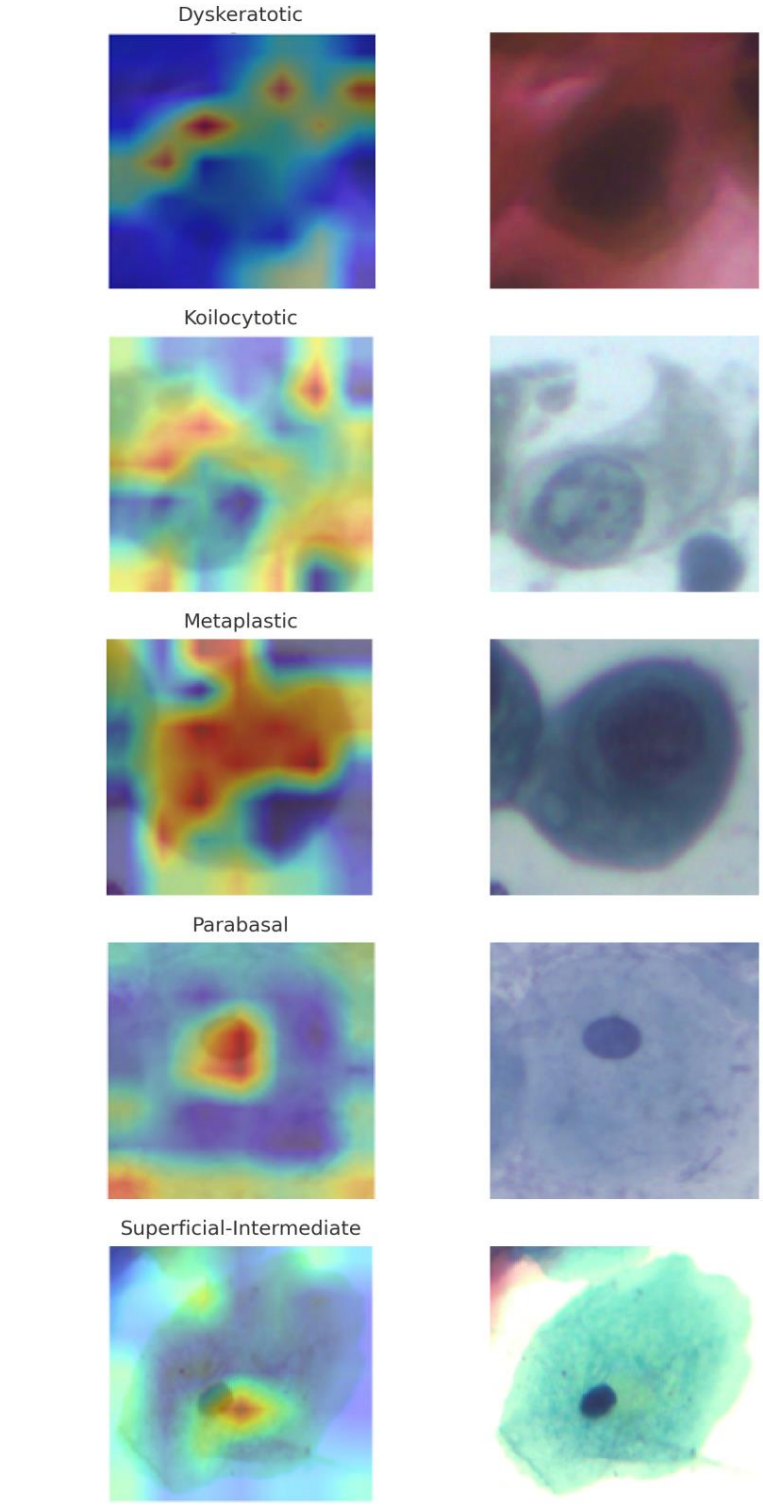

Figure 10:: Grad-CAM heatmap illustrating the regions of interest focused on by the ResNet50 model in cervical cell classification

The SqueezeNet model (Fig. 11) demonstrated better localization compared to ResNet50, with Grad-CAM heatmaps showing more concentrated focus on cell regions. This improvement was particularly evident for Koilocytotic and Metaplastic cells, where the heatmaps were more precise. However, some subtle features were still not fully captured, indicating room for improvement.

Original Images and Their Grad-CAM Visualizations (SqueezeNet)

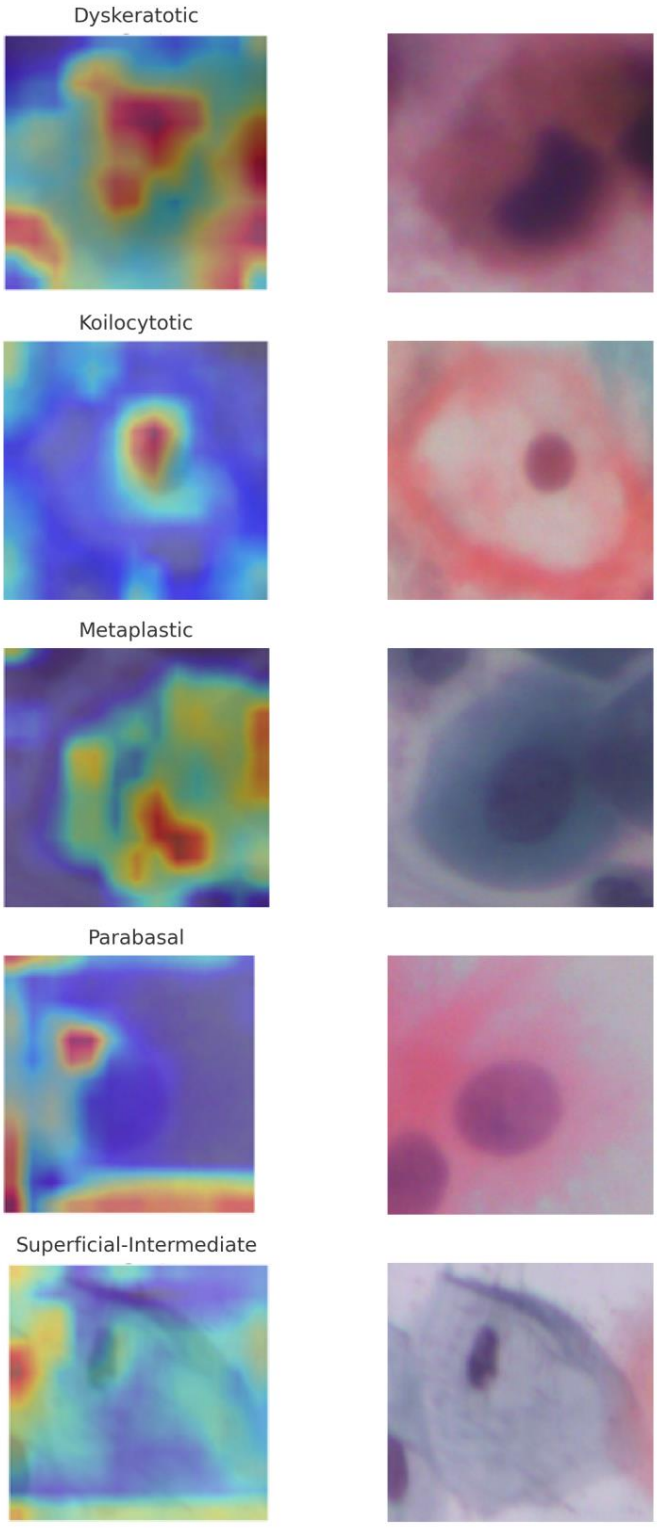

Figure 11:Grad-CAM heatmap illustrating the regions of interest focused on by the SqueezeNet model in cervical cell classification.

EfficientNet consistently localized key regions in the histological images. Its Grad-CAM heatmaps were well-aligned with diagnostically relevant features. Metaplastic cells displayed focused attention on cellular abnormalities with minimal irrelevant activations, making EfficientNet highly interpretable.

Original Images and Their Grad-CAM Visualizations (EfficientNet)

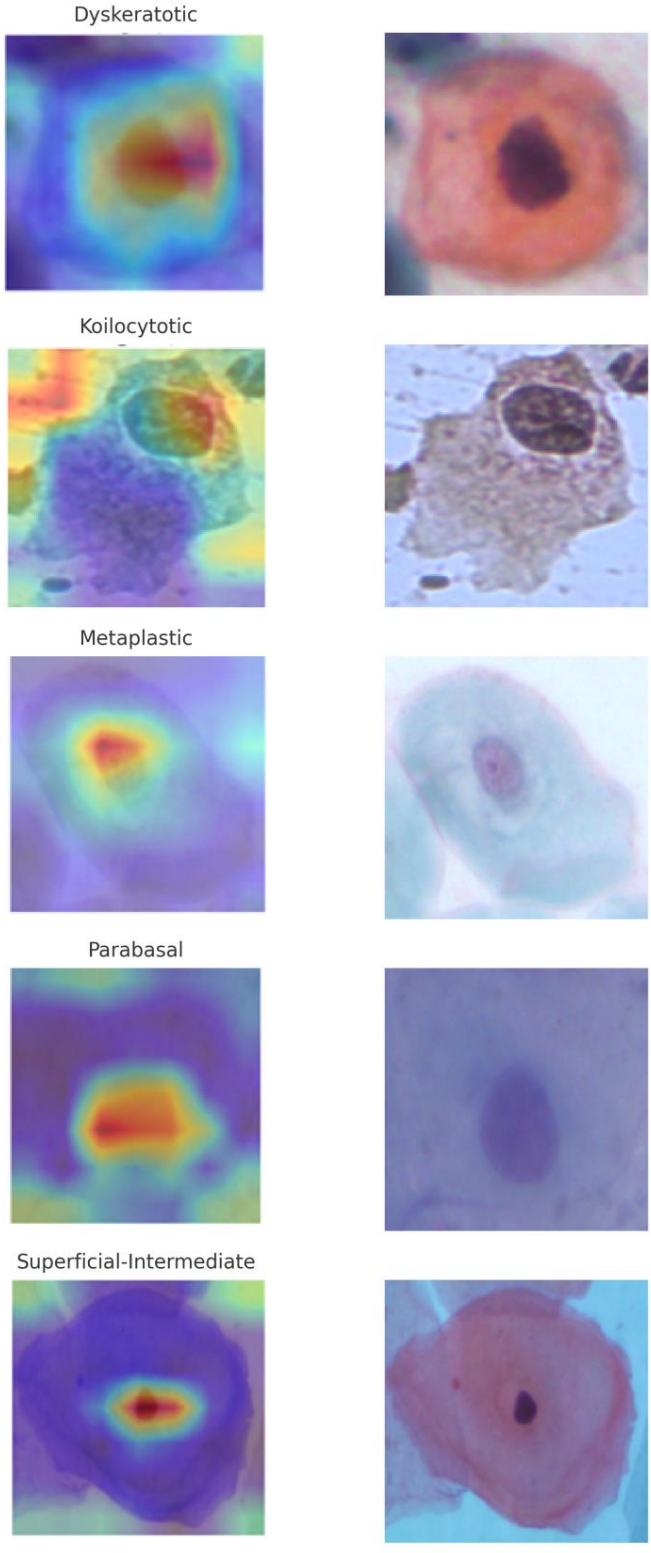

Figure 12:Grad-CAM heatmap illustrating the regions of interest focused on by the EfficientNet model in cervical cell classification.

The Visual Prompting Model delivered the cleanest and most interpretable heatmaps across all models. The activations were highly concentrated on diagnostically relevant areas, such as nuclear boundaries and abnormal regions. For example, Parabasal and Superficial-Intermediate cells demonstrated precise focus with no mislocalisations, emphasizing the Visual Prompting Model's superiority in interpretability.

Original Images and Their Grad-CAM Visualizations (Visual Prompting Model)

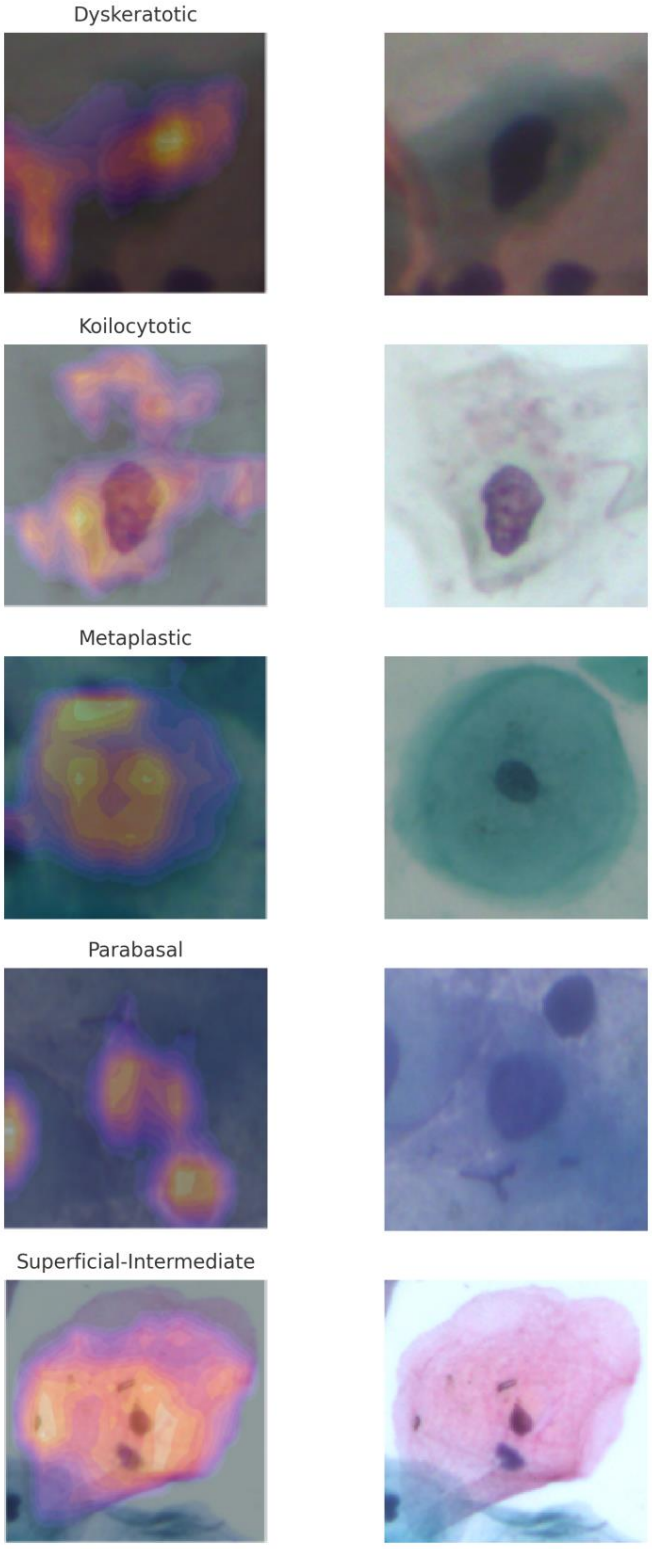

Figure 13:Grad-CAM heatmap illustrating the regions of interest focused on by the Visual Prompting Model in cervical cell classification

4.4. Comparative Summary

The Visual Prompting Model outperformed all others in both quantitative and qualitative evaluations. Its accuracy (98%), F1-score (0.99), and superior Grad-CAM heatmaps make it the most suitable model for cervical cell classification tasks. EfficientNet followed closely with robust performance and high interpretability, while SqueezeNet offered a lightweight alternative with competitive accuracy. ResNet50, though effective, was less interpretable and prone to more classification errors.

CHAPTER 5: DISCUSSION

Table 1 summarizes recent studies that employed deep learning techniques for cervical cancer detection, showcasing the diversity of methods, datasets, and reported performance metrics. These studies provide essential context for evaluating the performance and contributions of the models developed in this research. For instance, Deep Transfer Learning using CNNs with visualization schemes achieved high sensitivity (95.88%) and specificity (98.93%) when classifying cervical histological images, while a DenseNet-based model trained on the Kaggle Pap smear dataset achieved an accuracy of 98.2%. Similarly, an ensemble of VGG19, MobileNet, and DenseNet attained accuracy levels as high as 97.42%. These results align with the findings of this study, particularly regarding the performance of EfficientNet and visual prompting models.

In my study, I evaluated the effectiveness of various DL models in cervical cancer detection, specifically focusing on how CNNs – ResNet50, SqueezeNet, and EfficientNet – and visual prompting techniques enhance classification accuracy and interpretability. CNNs have been widely used in medical imaging due to their ability to automatically extract hierarchical features, while visual prompting introduces human-guided annotations to direct the model's focus toward diagnostically significant areas [58,59]. Additionally, Grad-CAM provided an explainability layer, ensuring that the model's predictions align with clinically relevant features.

Previously, CNNs have been widely used in medical imaging due to their ability to automatically extract hierarchical features and recognize complex patterns in histological images. However, despite their strong performance in various applications, challenges remain in distinguishing morphologically similar cell types.

Among the CNN models that I trained, EfficientNet achieved the highest accuracy (97%), demonstrating its strength in extracting meaningful features with optimized computational efficiency. The model was particularly effective in classifying Dyskeratotic, Metaplastic, and Parabasal cells, showing minimal misclassification errors for these classes. SqueezeNet, a lightweight architecture, achieved an accuracy of 95%, making it a viable choice for resource-constrained settings. While SqueezeNet demonstrated an improvement over ResNet50, it still exhibited some misclassifications, particularly between Koilocytotic and Superficial-Intermediate cells, highlighting the need for additional techniques to improve fine-grained classification. Park et al. (2021) reported that deep learning models combined with advanced

classification techniques significantly enhance diagnostic performance in cervical cancer, emphasizing the suitability of CNNs like ResNet50 and EfficientNet for these tasks[3].

Surprisingly, ResNet50, which is the deepest CNN that I trained in the study, achieved the lowest classification accuracy (91%), primarily due to difficulties in distinguishing Dyskeratotic and Koilocytotic cells, which share overlapping morphological features. The confusion matrices revealed that ResNet50 frequently misclassified Dyskeratotic cells as Koilocytotic and Metaplastic cells, suggesting that its feature extraction capabilities were not sufficient to resolve the nuanced differences between these categories. Although ResNet50 has demonstrated strong performance in general image classification tasks (Table 2), its limitations in this study underscore the challenges of applying standard CNNs to histological images with subtle inter-class variations.

While CNN models achieved high classification accuracy, their performance was limited by misclassification in challenging categories. I suggest that these findings highlight the need for enhanced model interpretability and targeted feature refinement to improve diagnostic accuracy, which was effectively addressed by incorporating visual prompting in the classification pipeline[60].

The Visual Prompting Model outperformed all CNN models, achieving an accuracy of 98% and an F1-score of 0.99. I suggested that robust performance of Visual Prompting Model could be attributed to the human-guided feature selection process. I specifically directed the model's attention toward diagnostically significant regions, thereby this led to minimizing misclassification errors. Most importantly, the trained Visual Prompting Model excelled in distinguishing the challenging to classify categories by the CNN models –Dyskeratotic and Koilocytotic cells.

Furthermore, by integrating annotated regions into the training process, visual prompting allowed the model to focus on clinically relevant features, such as nuclear abnormalities and cytoplasmic distortions, reducing false positives and false negatives. Unlike CNNs, which rely entirely on learned feature representations, visual prompting ensures that the model prioritizes regions identified by experts, leading to a more reliable classification.

My approach using visual prompting and Grad-CAM to advance the explainability of AI echos with recent development is AI-based medical diagnosis based on histology images. For

example, Chen et al. (2024) has proposed a Medical Visual Prompting (MVP) framework that utilizes super-pixel-guided prompting to improve segmentation and classification accuracy[61]. Their approach yield similar highly accurate model alignment observed in our Visual Prompting Model. Furthermore, Wu et al. (2024) demonstrated that visual prompting can mitigate domain shifts and reduce the reliance on extensive labelled datasets[62], a finding that reinforces the scalability and efficiency of the methods employed in this study.

The Grad-CAM heatmaps generated in this study (Figure 3) provide clear and interpretable visualizations, highlighting diagnostically significant regions. This is crucial for enhancing semantic perception in medical imaging, as noted by Liu et al. (2024), who emphasized that effective interpretability tools, like visual prompting, are essential for building clinician trust and facilitating the integration of AI-based systems into clinical workflows.

Visual Prompting demonstrates a key advantage in computational efficiency, as noted by Tian et al. (2024), who showed that adapting pre-trained models with minimal fine-tuning parameters achieves performance comparable to full fine-tuning—making it particularly suitable for deployment in resource-limited settings[63]. By reducing the computational burden, our approach offers a practical solution for scaling cervical cancer diagnostics in low-resource environments.

While this study was conducted in a controlled environment, its findings emphasize the need for multi-centre validation to ensure robustness in diverse clinical workflows. Qiao et al. (2024) highlighted the importance of standardizing data and validating AI models across diverse populations to achieve clinical applicability. This study's emphasis on Grad-CAM interpretability aligns with their call for transparency and trust in AI-assisted diagnostics.

The integration of deep learning and visual prompting in this study offers a practical means to overcome longstanding limitations in cervical cancer diagnostics. First, interpretability is substantially enhanced by highlighting diagnostically significant regions in histological images, thereby aligning model outputs with clinical decision-making. This finding aligns with previous reports emphasizing that transparent visualization tools encourage clinician trust and reduce diagnostic ambiguity[64,65]. Second, the synergy between convolutional neural networks and visual prompts drives robust classification accuracy. In particular, by directing attention

toward key features with minimal retraining, our approach surpasses traditional CNN-based algorithms that rely on extensive labelled data or sophisticated segmentation methods. Finally, the computational efficiency inherent to visual prompting expands diagnostic accessibility, especially in resource-limited settings. As supported by Bar et al. and Lei et al., minimal model modification and task-specific cues allow for rapid adaptation to new clinical contexts without necessitating high-end hardware or extensive retraining[28,64]. Taken together, these attributes highlight the promise of visual prompting as a generalizable, interpretable, and resource-efficient solution for early cervical cancer screening, especially when paired with high-performing deep learning architectures[48].

The clinical implications of this study are significant, particularly for resource-limited settings where the burden of cervical cancer is disproportionately high. Cervical cancer remains a leading cause of mortality in women, with 85% of cases occurring in LMICs due to inadequate infrastructure and a shortage of skilled professionals[15,41]. The automated and efficient nature of the Visual Prompting Model provides a scalable solution to these challenges. By achieving high diagnostic accuracy and offering interpretable outputs, the model reduces the dependency on highly skilled pathologists and ensures consistent diagnostic performance even in underserved regions.

Visual interpretability plays a pivotal role in fostering trust among clinicians in AI-assisted diagnostics. As highlighted by Garcia-Retamero et al., well-designed visual aids significantly enhance clinicians' understanding and confidence in diagnostic outputs[66]. The Grad-CAM heatmaps generated by the Visual Prompting Model not only localize diagnostically significant regions with precision but also support clinical workflows by reducing cognitive load and providing actionable insights. This transparency is critical for real-world adoption, where trust in AI systems is as important as their accuracy.

Additionally, the computational efficiency of the Visual Prompting Model addresses critical challenges in LMICs, where access to advanced hardware and training datasets is limited. Unlike traditional methods, which are resource-intensive and reliant on extensive retraining, visual prompting minimizes these requirements while maintaining high performance. This positions the model as a feasible and cost-effective solution for large-scale deployment, contributing to global efforts to eliminate cervical cancer[15] .

This study not only demonstrates the technical feasibility of combining deep learning with visual prompting for cervical cancer detection but also underscores its potential to transform

clinical practice. By addressing disparities in healthcare access, enhancing interpretability, and improving diagnostic accuracy, the proposed approach paves the way for scalable and equitable solutions in cervical cancer screening and diagnosis.

5.3. Limitations

Despite its promising findings, my study my have several limitations. First, the dataset used for model training and evaluation, while diverse, may not fully represent the variations encountered in real-world clinical settings, such as differences in imaging quality, staining techniques, and patient demographics. Second, the model's performance was assessed in a controlled environment; its robustness in dynamic clinical workflows remains untested. Additionally, while the Visual Prompting Model significantly enhanced interpretability, further studies are needed to validate whether clinicians can effectively utilize these visualizations in decision-making. Lastly, the computational efficiency of the model, while suitable for resource-limited settings, should be benchmarked against emerging lightweight architectures for more precise assessments of scalability.

5.4. Future Directions

Future research should address these limitations by validating the model's performance on larger, multi-centre datasets to ensure its generalizability across diverse populations. Incorporating real-world testing in clinical settings will also be critical for assessing usability and effectiveness under operational constraints. Moreover, expanding the model's functionality to include hybrid diagnostic approaches such as integrating histological imaging with genomic or molecular biomarker data could enhance its predictive power and applicability. Exploring alternative explainable AI techniques beyond Grad-CAM could further improve interpretability and clinician trust. Lastly, developing strategies for continuous model improvement, such as incorporating active learning to adapt to new data, will ensure its long-term relevance and scalability.

CHAPTER 6: CONCLUSION

I set out to develop and evaluate a deep learning–based approach for early cervical cancer detection, emphasizing both accuracy and interpretability through visual prompting techniques. The findings demonstrate that the proposed Visual Prompting Model surpasses established CNNs by achieving higher classification accuracy and providing superior localization of critical diagnostic features. Furthermore, the Grad-CAM heatmaps offer transparent visual explanations, aligning the model's predictions with clinical decision-making processes and reinforcing its potential for practical implementation in pathology settings.

Beyond its technical performance, the model's computational efficiency highlights its suitability for resource-constrained environments, where the global burden of cervical cancer is disproportionately high. By reducing the reliance on extensive retraining and large-scale labelled datasets, the approach is both scalable and feasible for varied clinical contexts. However, broader validation across multiple centres and populations remains essential to confirm the model's robustness and facilitate clinical translation. Future work should also explore integrating multimodal data and alternative explainable AI techniques, as well as evaluating real-world performance in operational workflows.

In conclusion, this study underscores the promise of combining deep learning and visual prompting to advance early cervical cancer diagnosis. The enhanced interpretability, high accuracy, and efficiency demonstrated here set a foundation for further innovations in AI-driven medical imaging and highlight a pathway toward improving patient outcomes on a global scale.

REFERENCES

1. Alsmariy, R., Healy, G. & Abdelhafez, H. Predicting Cervical Cancer using Machine Learning Methods. **11**, 173–184 (2020).

2. Lu, J., Song, E., Ghoneim, A. & Alrashoud, M. Machine learning for assisting cervical cancer diagnosis : An ensemble approach. *Futur. Gener. Comput. Syst.* **106**, 199–205 (2020).

3. Park, Y. R. *et al.* Comparison of machine and deep learning for the classification of cervical cancer based on cervicography images. *Sci. Rep.* **11**, 1–11 (2021).

4. Aswathy, M. A. & Jagannath, M. Informatics in Medicine Unlocked Detection of breast cancer on digital histopathology images : Present status and future possibilities. *Detect. breast cancer Digit. Histopathol. images Present status Futur. possibilities* **8**, 74–79 (2017).

5. Kriegeskorte, N. & Douglas, P. K. Cognitive computational neuroscience. **21**, 1148–1160 (2019).

6. Tanimu, J. J., Hamada, M., Hassan, M. & Ilu, S. Y. A Contemporary Machine Learning Method for Accurate Prediction of Cervi- cal Cancer. **04004**, (2021).

7. Latif, J., Xiao, C., Imran, A. & Tu, S. Medical imaging using machine learning and deep learning algorithms: A review. *2019 2nd Int. Conf. Comput. Math. Eng. Technol. iCoMET 2019* 1–5 (2019) doi:10.1109/ICOMET.2019.8673502.

8. Liao, N. *et al.* Rethinking Visual Prompt Learning as Masked Visual Token Modeling. (2023).

9. Ng, A. Visual prompting livestream with Andrew Ng. (2023).

10. Zhang, Y., Pan, B., Gu, S., Bai, G. & Qiu, M. Visual Attention Prompted Prediction and Learning. (2023).

11. Yang, C.-F. *et al.* Verbalized Representation Learning for Interpretable Few-Shot Generalization. (2024).

12. Perkins, R. B., Wentzensen, N., Guido, R. S. & Schiffman, M. Cervical Cancer Screening: A Review. *Jama* **330**, 547–558 (2023).

13. Zhang, S., Xu, H., Zhang, L. & Qiao, Y. Cervical cancer: Epidemiology, risk factors and screening. *Chinese J. Cancer Res.* **32**, 720–728 (2020).

14. Johnson, C. A., James, D., Marzan, A. & Armaos, M. Cervical Cancer: An Overview of Pathophysiology and Management. *Semin. Oncol. Nurs.* **35**, 166–174 (2019).

15. Xue, P., Tat, M., Ng, A. & Qiao, Y. The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence. 1–7 (2020).

16. Choudhary, K. *et al.* Recent advances and applications of deep learning methods in materials science. (2022) doi:10.1038/s41524-022-00734-6.

17. Malhotra, R. & Singh, P. *Recent advances in deep learning models : a systematic literature review Recent advances in deep learning models : a systematic literature review*. *Multimedia Tools and Applications* (Springer US, 2023). doi:10.1007/s11042-023-15295-z.

18. Zhao, X., Wang, L., Zhang, Y., Han, X. & Deveci, M. *A review of convolutional neural networks in computer vision*. *Artificial Intelligence Review* vol. 57 (Springer Netherlands, 2024).

19. Ankile, L., Morgan, H. & Krange, K. Deep Convolutional Neural Networks A survey of the foundations, selected improvements, and some current applications. (2020).

20. Dhillon, A. & Verma, G. K. Convolutional neural network: a review of models, methodologies and applications to object detection. *Prog. Artif. Intell.* **9**, 85–112 (2020).

21. Zhang, Q., Wang, X., Wu, Y. N., Zhou, H. & Zhu, S. C. Interpretable CNNs for Object Classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3416–3431 (2021).

22. Sharma, S. & Guleria, K. Deep Learning Models for Image Classification: Comparison and Applications. *2022 2nd Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2022* 1733–1738 (2022) doi:10.1109/ICACITE53722.2022.9823516.

23. Gorantla, R., Singh, R. K., Pandey, R. & Jain, M. Cervical Cancer Diagnosis using CervixNet - A Deep Learning Approach. *Proc. - 2019 IEEE 19th Int. Conf. Bioinforma. Bioeng. BIBE 2019* 397–404 (2019) doi:10.1109/BIBE.2019.00078.

24. Song, J., Gao, S., Zhu, Y. & Ma, C. A survey of remote sensing image classification

based on CNNs. *Big Earth Data* **3**, 232–254 (2019).

25. Bnouni, N. *et al.* Boosting CNN Learning by Ensemble Image Preprocessing Methods for Cervical Cancer Segmentation. *18th IEEE Int. Multi-Conference Syst. Signals Devices, SSD 2021* 264–269 (2021) doi:10.1109/SSD52085.2021.9429422.

26. Dongyao Jia, A., Zhengyi Li, B. & Chuanwang Zhang, C. Detection of cervical cancer cells based on strong feature CNN-SVM network. *Neurocomputing* **411**, 112–127 (2020).

27. Hu, L. *et al.* An Observational Study of Deep Learning and Automated Evaluation of Cervical Images for Cancer Screening. *J. Natl. Cancer Inst.* **111**, 923–932 (2019).

28. Bar, A. *et al.* Visual Prompting via Image Inpainting. 1–13 (2022).

29. Li, Q., Jiang, Z., Wang, B. & Li, L. Towards Robust and Accurate Visual Prompting.

30. Huang, H., Zhao, Z., Backes, M., Shen, Y. & Zhang, Y. Prompt Backdoors in Visual Prompt Learning. (2023).

31. Jin, H., Che, H., Lin, Y. & Chen, H. PromptMRG : Diagnosis-Driven Prompts for Medical Report Generation. 2607–2615 (2020).

32. Wu, H., Bremond-Martin, C., Bouaou, K. & Clouchoux, C. Tumor Segmentation on Whole Slide Images: Training or Prompting? *Proc. - Int. Symp. Biomed. Imaging* (2024) doi:10.1109/ISBI56570.2024.10635239.

33. Lei, Y., Li, J., Li, Z., Cao, Y. & Shan, H. Prompt learning in computer vision: a survey. *Front. Inf. Technol. Electron. Eng.* **25**, 42–63 (2024).

34. Garcia-Retamero, R., Cokely, E. T. & Hoffrage, U. Visual aids improve diagnostic inferences and metacognitive judgment calibration. *Front. Psychol.* **6**, 1–12 (2015).

35. Zhang, L. *et al.* DeepPap: Deep convolutional networks for cervical cell classification. *IEEE J. Biomed. Heal. Informatics* **21**, 1633–1643 (2017).

36. Cheng, S. *et al.* Robust whole slide image analysis for cervical cancer screening using deep learning. *Nat. Commun.* **12**, 1–10 (2021).

37. Chitra, B. & Kumar, S. S. Recent advancement in cervical cancer diagnosis for automated screening: a detailed review. *J. Ambient Intell. Humaniz. Comput.* **13**, 251–269 (2022).

38. Sarhangi, H. A., Beigifard, D., Farmani, E. & Bolhasani, H. Deep Learning Techniques for Cervical Cancer Diagnosis based on Pathology and Colposcopy Images. (2024).

39. Lilhore, U. K. *et al.* Hybrid Model for Detection of Cervical Cancer Using Causal Analysis and Machine Learning Techniques. *Comput. Math. Methods Med.* **2022**, (2022).

40. Ijaz, M. F., Attique, M. & Son, Y. Data-Driven Cervical Cancer Prediction Model with Outlier Detection and Over-Sampling Methods. 1–22 (2020).

41. Hou, X. *et al.* Artificial Intelligence in Cervical Cancer Screening and Diagnosis. **12**, 1–13 (2022).

42. Chandran, V. *et al.* Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images. *Biomed Res. Int.* **2021**, (2021).

43. Debelee, T. G., Schwenker, F., Ibenthal, A. & Yohannes, D. Survey of deep learning in breast cancer image analysis. *Evol. Syst.* **11**, 143–163 (2020).

44. Sahasrabuddhe, V. V., Luhn, P. & Wentzensen, N. Human papillomavirus and cervical cancer: Biomarkers for improved prevention efforts. *Future Microbiol.* **6**, 1083–1098 (2011).

45. Fernandes, K., Chicco, D., Cardoso, J. S. & Fernandes, J. Supervised deep learning embeddings for the prediction of cervical cancer diagnosis. *PeerJ Comput. Sci.* **2018**, 1–20 (2018).

46. Chen, Ying; Qin, Xiaomin; Xiong, Jingyu; Xu, Shugong; Shi, Jun; Lv, Huabing; Li, Lin; Xing, Hui; Zhang, Q. Deep Transfer Learning for Histopathological Diagnosis of Cervical Cancer Using Convolutional Neural Networks with Visualization Schemes. *J. Med. Imaging Heal. Informatics* **10**, 391-400(10) (2020).

47. Kassani, S. H., Kassani, P. H., Wesolowski, M. J., Schneider, K. A. & Deters, R. Classification of histopathological biopsy images using ensemble of deep learning networks. *CASCON 2019 Proc. - Conf. Cent. Adv. Stud. Collab. Res. - Proc. 29th Annu. Int. Conf. Comput. Sci. Softw. Eng.* 92–99 (2020).

48. Dhawan, S., Singh, K. & Arora, M. Cervix image classification for prognosis of cervical cancer using deep neural network with transfer learning. *EAI Endorsed Trans.*

Pervasive Heal. Technol. **7**, 1–9 (2021).

49. Sompawong, N. *et al.* Automated Pap Smear Cervical Cancer Screening Using Deep Learning. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* 7044–7048 (2019) doi:10.1109/EMBC.2019.8856369.

50. Kim, J., Lee, Y. S., Yu, J., Park, Y. & Lee, S. K. Deep Learning-Based Prediction Model for Breast Cancer Recurrence Using Adjuvant Breast Cancer Cohort in Tertiary Cancer Center Registry. **11**, 1–11 (2021).

51. Paszke, A. *et al.* PyTorch : An Imperative Style , High-Performance Deep Learning Library. (2019).

52. Howard, A. G. Mobilenets: Efficient convolutional neural networks for mobile vision applications. (2017).

53. Kumari, L., Mittal, N. & Modi, M. Comparative Analysis of Noise Reduction Filters for Medical Images. *2024 Fourth Int. Conf. Adv. Electr. Comput. Commun. Sustain. Technol.* 1–7 doi:10.1109/ICAECT60202.2024.10469449.

54. Adusumalli, H., Kalyani, A., Krishna Sr, R., Pratapteja, M. & Prasada, R. Face Mask Detection Using OpenCV. 1304–1309 (2021) doi:10.1109/ICICV50876.2021.9388375.

55. Selvaraju, R. R. *et al.* Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. *Grad-CAM Vis. Explan. from Deep Networks via Gradient-based Localization* **17**, 331–336 (2016).

56. Suara, S., Jha, A., Sinha, P. & Sekh, A. A. Is Grad-CAM Explainable in Medical Images? 1–13.

57. Plissiti, M. E. *et al.* SIPAKMED : A NEW DATASET FOR FEATURE AND IMAGE BASED CLASSIFICATION OF NORMAL AND PATHOLOGICAL CERVICAL CELLS IN PAP SMEAR IMAGES Dept . of Computer Science & Engineering , University of Ioannina , Greece Dept . of Anatomy-Histology and Embryology , Faculty of Medicine , University of Ioannina , Greece. 3144–3148 (2018).

58. Alagbe, E., Isewon, I. & Oyelade, J. Optimizing machine learning performance for medical imaging analyses in low-resource environments : The prospects of CNN-based Feature Extractors. 1–13 (2025).

59. Zhu, K. *et al.* Framework Design and Comprehensive Exploration of Prompt Variations. (2025).

60. Denner, S., Zimmerer, D., Paul, F. J. & Maier-hein, K. Visual Prompt Engineering for Medical Vision Language Models in Radiology. (2024).

61. Segmentation, H. M. I. *et al.* Medical Visual Prompting (MVP): A Unified Framework for Versatile.

62. Wu, C., Restrepo, D., Shuai, Z., Liu, Z. & Shen, L. Meta-driven Visual Prompt Selection.

63. Tian, H. *et al.* MoVL : Exploring Fusion Strategies for the Domain-Adaptive Application of Pretrained Models in Medical Imaging Tasks. (2023).

64. Lei, Y., Li, J., Li, Z., Cao, Y. & Shan, H. Prompt learning in computer vision: a survey ∗. **25**, 42–63 (2024).

65. Wu, T., Lucas, E., Zhao, F., Basu, P. & Qiao, Y. Artificial intelligence strengthenes cervical cancer screening – present and future Machine learning in cervical cancer. (2024) doi:10.20892/j.issn.2095-3941.2024.0198.

66. Garcia-retamero, R. Visual aids improve diagnostic inferences and metacognitive judgment calibration. **6**, 1–12 (2015).