Development of a Web-Based Diagnostic Tool for Real-Time Classification and Visualization of Affected Areas in Colposcopy Images from the Central Hospital of Parakou, Benin

Emile Honaminto DJOHI

Supervised by : Prof. Songbian Zime

African School of Economics
Abomey-Calavi, Benin
©Emile Honaminto Djohi
September 2025

Abstract

Cervical cancer is one of the leading causes of morbidity and mortality among women worldwide, particularly in contexts where access to routine screening and specialized expertise is limited. Colposcopy continues to be an essential diagnostic tool despite the fact that its effectiveness depends especially on the experience of the clinician and is often constrained by subjectivity in image interpretation. Recent advances in Artificial Intelligence (AI) offer significant opportunities to standardize diagnosis, improve accuracy, and make early detection more accessible.

This study develops a web-based diagnostic platform capable of analyzing colposcopy images and providing clinicians with immediate and automated feedback on cervical lesion classification. To achieve this goal, three deep learning models were designed and compared: a U-Net combined with a Support Vector Machine (SVM) pipeline, a VGG-19 transfer learning model, and the specialized CYENET architecture. A data set of 494 original colposcopy images was collected at the Central Hospital of Parakou, Benin, and expanded to 1,800 images by increasing to balance six diagnostic categories (normal, CIN1, CIN2, CIN3, metaplasia, and erectile dysfunction). The models were trained and evaluated using standard metrics, including accuracy, precision, recall, and F1-score.

The performance outcome is that the VGG-19 transfer learning model achieves highest performance with accuracy of 89.17% along with precision (89.28%), recall (89.17%) and F1-score (89.22%) being balanced values respectively, which presents a good classification between normal and pathogenic samples. CYENET obtained 85.19% accuracy and the U-Net combined with the SVM pipeline reached 84%. U-Net + SVM was less accurate than VGG-19 but it had faster training and prediction, while CYENET is consistent and performed well in many benign cases.

With these results, we developed the VGG-19 model within a lightweight web application. The system allows users to upload their new colposcopy images or capture them in a live session, and receive an automatic classification under the six diagnostic classes. It displays interpretable results, confidence estimates, and prior analyses to maintain transparency and clinical relevance. The system is multilingual and has been designed for resource-constrained environments (it needs only standard computing capabilities).

Incorporating the best performing model into this online tool shows the potential feasibility of AI-assisted colposcopy in real-time. The system improves cervical cancer screening and early detection, especially in low-resource medical centers with standard and instantaneous diagnostic results delivered by the platform.

keywords: Cervical cancer, Colposcopy, VGG19, Deep learning, U-Net, Classification

Contents

Page

Abstract ii

Contents iii

List of Tables v

List of Figures vi

1 Introduction 1

2 Literature review 3

2.1 Conceptual Clarification . . . 3
2.2 Evolution of Cervical Cancer in Benin . . . 4
2.3 Diagnosis through Colposcopy and limitations . . . 5
2.4 Integration of AI for cervical cancer detection . . . 6
2.5 Comparative analysis of existing approaches . . . 8

3 Data and Methodology 10

3.1 Data acquisition . . . 10
3.2 Data Annotation . . . 10
3.3 Data Preprocessing . . . 11
3.4 Methodology . . . 13
3.4.1 Pre segmentation and Region of Interest Extraction . . . 15
3.4.2 Segmentation . . . 16
3.5 Model Architectures . . . 17
3.5.1 U-Net combined with SVM Approach . . . 17
3.5.2 VGG-19 Transfer Learning Approach . . . 19
3.5.3 CYENET Architecture . . . 21
3.6 Classification Function . . . 23
3.7 Performance Evaluation . . . 24
3.7.1 Individual Class Performance Metrics . . . 25
3.7.2 Overall Model Performance Metrics . . . 25
3.8 Web Platform Implementation and Deployment . . . 26

3.8.1 Navigation Structure and Multilingual Support 26
3.8.2 Frontend Development with React 27
3.8.3 Model Integration and Deployment 28
3.8.4 Design of the user interface and workflow 28
3.8.5 Performance and Scalability Considerations 29
3.8.6 Deployment and accessibility 29
3.8.7 Regulatory Compliance and Platform Governance 30
3.8.8 Testing and Validation 30

4 Results **32**
4.1 Classification using U-Net and SVM approach 32
4.1.1 Cervix Image Segmentation with U-Net 32
4.1.2 Classification with SVM 33
4.2 Result of VGG19 TL model 35
4.3 Result of CYENET model 37
4.4 Comparison of the models 40
4.5 Discussion 42

5 Conclusion **44**

Bibliography **45**

List of Tables

1 Comparison of existing AI approaches for cervical lesion classification 8
2 Data distribution across the categories 12
3 U-net training hyperparameter details 18
4 Training and fine-tuning configurations for VGG19 TL approach. 21
5 Proposed CYENET Configuration and Training Parameters 23
6 Comparison of the three models Performance 40
7 Per-class Precision, Recall, and F1-Score for the three models 41
8 Comparison of the three models using the model complexity. 41

List of Figures

1 Colposcopy images 11
2 A normal cervix and the geometric transformations which were used in the data augmentation 13
3 Overall Methodology 14
4 Detailed Pre-Segmentation and Segmentation Process 16
5 CYENET workflow chart 22
6 QR code to access the deployed Web App (EmiloVision Benin) 27
7 Metrics, ((a) Loss, (b) accuracy, (c) precision, and (d) DICE concerning the different number of epochs during the training and validation of the U-Net architecture . 33
8 Confusion matrix for SVM Test 34
9 Representative Test Images with True Labels and U.Net-SVM Predicted Labels . . 35
10 VGG19 (TL) Performance evaluation 36
11 Classification of 12 ramdom annoted images based on VGG19 approach 37
12 CYENET Performance Evaluation 38
13 Confusion Matrix through test with CYENET 39
14 Original image, attention zones, and affected area overlays for CIN1 and CIN3 . . 39

1 Introduction

Cervical cancer represents one of the most significant and preventable threats to women's health on a global scale. It constitutes a malignant disease that develops in the cervix, the lower part of the uterus, and its primary cause is a persistent infection with high-risk types of human papillomavirus (HPV), (zur Hausen, 2002). Although high-income countries have seen a dramatic decline in cervical cancer rates due to extensive HPV vaccination and organized screening initiatives (Arbyn et al., 2020), it remains a leading cause of cancer-related mortality in low- and middle-income countries, especially in Africa. In 2022, approximately 660,000 women received a diagnosis of the disease, leading to more than 350,000 deaths. In addition, around 90% of these fatalities were recorded in low- and middle-income countries, demonstrating the persistent gap in global health equity (World Health Organization, 2024).

According to the World Health Organization, the highest incidence and mortality rates of cervical cancer are found in sub-Saharan Africa. In Benin, cervical cancer is currently the second most commonly diagnosed cancer in women and the leading cause of cancer-related deaths in the female population. Recent data shows that more than 1,000 women are diagnosed with cervical cancer each year in Benin, with approximately 700 deaths annually (Bruni et al., 2023). These findings indicate the lack of routine screening, late diagnosis, limited diagnostic infrastructure, and the shortage of trained medical professionals, particularly in rural areas. Given these challenges, improving diagnostic tools is essential, especially in settings with limited resources.

Colposcopy is a standard technique used to detect and stage cervical cancer. It relies on skilled professionals to examine images and make accurate diagnoses. The procedure involves using a specialized microscope to examine the cervix after applying solutions that highlight abnormal areas. However, the effectiveness of colposcopy depends mainly on the experience and expertise of the examining physician. The subjective nature of image interpretation, combined with the time-intensive diagnostic process and resource requirements, creates barriers to widespread screening access (Mulongo and Chibwesha, 2022). These have prompted researchers to explore innovative approaches that could enhance diagnostic capacity while making screening more accessible to underserved populations. Artificial intelligence and machine learning are providing innovative tools to improve the diagnosis and management of cervical cancer, with a particularly promising application in the automated analysis of colposcopy images (Hou et al., 2022). At the heart of this approach are convolutional neural networks (CNNs), which are capable of extracting and learning complex patterns directly from the pixels of medical images. Studies have shown that computer-based diagnostic tools can achieve performance levels comparable to, or sometimes exceeding, those of experienced clinicians in cervical lesion detection and classification. A study by Lee et al.

(2018) demonstrated that visual inspection of cervical images, when supported by digital tools, could reliably identify high-grade lesions. Transfer learning approaches, such as those that use established image recognition systems like VGG-19, have proven effective in adapting existing knowledge to medical imaging tasks. Similarly, specialized computer systems such as CYENET, designed specifically for the detection of cervical abnormalities, have shown improved accuracy in the identification of different types of cervical lesions (Chandran et al., 2021). Additionally, image segmentation techniques using architectures like U-Net can accurately highlight suspicious regions in cervical images, improving diagnostic consistency and reducing variation between different interpretations by experts Yuan, Zhang and Huang (2020). Similarly, Silva et al. (2020) introduced remote access platforms that facilitate timely diagnosis, especially in areas with fewer health centers.

The main goal of this research is to develop a web-based diagnostic platform that can analyze colposcopy images and provide immediate feedback to clinicians. A secondary goal is to compare three AI-based approaches for cervical lesion classification: (1) a hybrid system combining U-Net segmentation with Support Vector Machine classification, (2) VGG-19 transfer learning, and (3) the CYENET architecture, which is specifically designed for cervical abnormality detection. This comparative analysis will help identify the most suitable approach for deployment. The data used come from the Parakou Departmental University Hospital Center (CHUD), with the support of local clinicians who annotated the image data, ensuring strong contextual relevance and validation within a resource-limited local environment. Beyond its technical purpose, this work directly responds to the documented needs expressed by clinicians at the Central Hospital of Parakou, which serves as the primary referral center for more than two million people in northern Benin. The hospital maintains established colposcopy services and has expressed a strong interest in technological solutions that could improve diagnostic accuracy and reduce the burden of cervical cancer in their patient population. By strengthening diagnostic capabilities, standardizing clinical practices, and facilitating access to quality screening services for women throughout the region, including those in remote rural areas, this research represents a meaningful contribution to reducing health disparities and supporting efforts to combat a preventable but persistently deadly disease.

The upcoming sections of the paper are structured as follows: Section 2 will review the literature on cervical cancer diagnostics, the operational challenges of colposcopy in Benin, and the application of deep learning models to medical image analysis. Section 3 will provide a methodology, including the data acquisition protocol, the technical architecture of the AI model, and the framework of the web-based platform. Section 4 presents the analysis of the empirical results and the performance of the model. Finally, Section 5 concludes by discussing the implications of the findings, the limitations of the study.

2 Literature review

This section presents the review of the literature based on our reading. It begins with an overview of the global burden of cervical cancer and the challenges of diagnosis, particularly in underserved contexts. Then it examines the development of AI-driven tools in medical diagnostics, with a focus on their application to the screening of cervical cancer.

2.1 Conceptual Clarification

Cervical cancer is a type of cancer that develops in the cervix, the lower part of the uterus that connects to the vagina. It does not develop rapidly, instead, it typically evolves over many years through a series of detectable stages. Extensive scientific research has identified the root cause of most cervical cancer cases as persistent infection with high-risk types of Human Papillomavirus (HPV), particularly types 16 and 18, which are related to approximately 70% of global cases (Bosch et al., 2002). In the majority of cases, the immune system eliminates the virus within two years without causing long-term problems. However, in a small proportion of cases, especially when the infection persists, harmful changes occur in the cervix cells. This long-term presence of the virus can integrate into the DNA of cervical cells and begin producing viral proteins, particularly E6 and E7, which interfere with natural tumor suppressors in the body (Münger et al., 2004). As a result, normal cell functions are disrupted, leading to uncontrolled growth and the gradual development of cancerous lesions.

Before reaching the stage of cancer, these changes usually appear as abnormal or precancerous lesions in the cervix, medically termed Cervical Intraepithelial Neoplasia (CIN). CIN is further classified into three stages. CIN 1 (low-grade) involves mild changes, often self-resolving. CIN 2 and CIN 3 (high grade) reflect more severe cellular changes and are more likely to evolve into invasive cancer if left untreated. These early warning signs provide a valuable opportunity to intervene and prevent disease progression (McCredie et al., 2008). Recent studies emphasize the importance of identifying these precancerous stages using accessible diagnostic methods. Kruczkowski et al. (2024) paper highlights that early intervention during the CIN stages dramatically improves outcomes, especially when screening is integrated into routine health services. Artificial intelligence (AI) tools are being explored as solutions to help detect these early abnormalities, reducing the reliance on specialists in resource-limited settings.

2.2 Evolution of Cervical Cancer in Benin

According to the most recent global cancer statistics, cervical cancer is the second most common cancer among women in Benin and the leading cause of female cancer-related deaths. The data, compiled by the HPV and Cancer ICO/IARC, report an estimated 1,143 new cases and 796 deaths in the country each year (Bruni et al., 2023). This means that more than two women die from a largely preventable disease every day. The age-standardized incidence rate of 36.1 per 100,000 women is alarmingly high, standing in stark contrast to the rates seen in regions with robust healthcare infrastructure, such as Western Europe (7.8 per $100,000$) or North America (6.5 per 100,000). This situation results from several factors that reflect the challenges of the healthcare system in a larger way:

- **Lack of National Organized Screening:** Benin does not currently have a national screening program for cervical cancer. Instead, screening services are opportunistic, meaning that women are only tested if they present to a health facility that offers such services and often at their own expense. This produces inadequate screening coverage, especially in rural areas where healthcare infrastructure is sparse. Kim et al. (2013) notes that the lack of organized screening is a significant obstacle in sub-Saharan Africa and leads to the detection of most cases in the late stages when treatment is less effective. Research examining barriers to cervical cancer screening identifies the lack of organized screening programs as a significant obstacle in sub-Saharan Africa, leading to the detection of most cases in the late stages when treatment is less effective and more costly Mantula et al. (2024).

- **Prevalence of late-stage diagnosis:** The HPV vaccine is a powerful tool for primary prevention, capable of preventing infections that cause most cervical cancers. However, implementing and scaling up national vaccination programs in Benin has faced significant logistical and financial hurdles and the direct and devastating consequence of inadequate screening is that most women are diagnosed only after developing symptoms such as abnormal bleeding or pain (Piras et al., 2011). At this stage, cancer is often advanced and has spread, making treatment far more complex, less likely to be successful, and prohibitively expensive.

- **Treatment cost:** A systematic review of barriers to cervical cancer screening in low- and middle-income countries highlights cost as the main obstacle, along with insufficient knowledge and the distance to health facilities (Mvundura and Tsu, 2014). In Benin, where a 36.2% of the population in 2021 lives in poverty according to the World Bank Economic Outlook report on Benin 2025 , the direct costs associated with screening tests, transportation to health centers, and potential loss of income create financial barriers for women seeking screening.

The consequences of these barriers are that most cases of cervical cancer in Benin are diagnosed only after women experience advanced symptoms, such as abnormal bleeding, pelvic pain, or unusual discharge. At this stage, cancer has frequently progressed to advanced stages, making treatment more complicated, less likely to succeed, and too expensive for many families. This situation mirrors a wider trend seen throughout sub-Saharan Africa, where poor screening infrastructure and limited access to prevention services lead to higher rates of cervical cancer death.

2.3 Diagnosis through Colposcopy and limitations

When a primary screening test, such as the HPV test, suggests the presence of abnormal cervical cells, colposcopy is usually the next step in the diagnostic pathway and remains an important component of cervical cancer diagnosis, especially for the confirmation and characterization of lesions. It allows a direct visual examination of the cervix, helping clinicians identify areas that may require biopsy and further histological analysis (Xue et al., 2020).

The procedure involves the use of a colposcope, which provides a clear and enlarged view of the cervical epithelium. To enhance visualization, clinicians typically apply a 3–5% acetic acid solution to the cervix. This causes abnormal epithelial areas to temporarily turn white, a process known as acetowhitening. [1] This reaction is more pronounced in cells with a high nucleus-to-cytoplasm ratio, characteristic of dysplastic or neoplastic changes. Colposcopists evaluate the resulting patterns, evaluating characteristics such as color intensity, vascular structure, lesion boundaries, and size (Shinohara et al., 2023).

Although colposcopy is highly useful in experienced hands, its effectiveness is limited by subjectivity. Interpreting visual cues requires significant expertise, and diagnostic accuracy can vary depending on the training and experience of the clinician (Qin et al., 2023). In countries with low resources, the availability of trained colposcopists is insufficient, leading to delays in diagnosis and follow-up. In addition, the sensitivity and specificity of the procedure vary widely, ranging between 50% and 90%, depending on the severity of the injury and the criteria used (Xue et al., 2020). In addition to the technical challenges, practical limitations also hinder widespread access to colposcopy. The cost of equipment, maintenance requirements, and the need for a stable electricity supply pose barriers, particularly in rural and underserved areas (Mantula et al., 2024). Consequently, many women at risk are not referred or do not receive timely evaluation, contributing to the persistence of advanced stage diagnoses in sub-Saharan Africa.

In recent years, digital colposcopy and computer-assisted image analysis have emerged as promising solutions to improve objectivity and reduce variability. Algorithms trained on annotated colpo-

[1] Acetowhitening: whitening of mucous membrane after application of acetic acid solution. Accessed Link.

scopic images can help standardize lesion detection and grading. Several studies have shown that machine learning models, particularly convolutional neural networks (CNNs), can perform comparably to expert clinicians in distinguishing between high- and low-grade lesions (Saini et al., 2020; Wang et al., 2024) . Despite this progress, studies caution against overreliance on automated tools without adequate validation. Most existing models are trained on small, homogeneous datasets, which limits their generalizability. There is also a lack of consensus on image quality standards and clinical interpretation protocols when AI is used in conjunction with colposcopy (Denny et al., 2022). Furthermore, explainability remains a concern, as clinicians must be able to interpret and trust AI output to integrate them into routine care (Tjoa and Guan, 2020).

2.4 Integration of AI for cervical cancer detection

The digitization of medical imaging has significantly transformed diagnostic approaches in various clinical disciplines. Among these, colposcopy, long dependent on the interpretive skills of clinicians, presents a compelling case for technological enhancement. The integration of Artificial Intelligence (AI), particularly deep learning models such as Convolutional Neural Networks (CNNs), has demonstrated promising potential to assist in interpreting colposcopic images by automating classification, highlighting abnormalities, and enhancing diagnostic precision.

One of the most direct applications of AI in this context is lesion classification. Using thousands of annotated colposcopic images, CNNs can distinguish between normal tissue, low-grade, and high-grade lesions. A seminal study by Hu et al. (2020) trained a deep learning model on more than $60,000$ cervical images collected from more than $9,000$ women. The model achieved an Area Under the Curve (AUC) (which represents the probability that the model, if it receives a randomly chosen positive example and a negative example, classifies the positive example higher than the negative example) of 0.91, which exceeded both the interpretation of the expert human (AUC 0.69) and conventional cytology (AUC 0.71). This suggests that AI has a strong ability to provide a consistent and objective second opinion and will be particularly useful in settings where expert colposcopists are scarce. Building on this work, Chandran et al. (2021) introduced $CYENET$, a deep learning framework tailored for the multiclass classification of cervical lesions. Their model achieved a precision of 92.3%, emphasizing the importance of architectural innovations in improving diagnostic results. Jiménez Gaona et al. (2022), have similarly explored hybrid models that combine deep learning with conventional machine learning classifiers such as Support Vector Machines (SVMs) to improve robustness in diverse clinical environments.

Beyond classification, segmentation tasks have become central in AI-assisted colposcopy. The precise delineation of the lesion boundaries is important for guided biopsies and determining treat-

ment pathways. The U-Net architecture, first proposed by Ronneberger et al. (2015), remains a benchmark in biomedical image segmentation. Yuan, Zhang and Huang (2020) demonstrated its effectiveness in highlighting abnormal regions within colposcopic images, showing how segmentation tools can function as *digital highlighters*, ensuring that clinicians focus their attention on the most diagnostically relevant areas. In settings with a high burden of disease and limited time per patient, such assistance can significantly reduce diagnostic error and variability between clinicians. A further evolution in the field has been the development of integrated clinical platforms that embed AI models into user-friendly interfaces. These platforms aim to bridge the gap between algorithmic performance in academic settings and their actual impact in clinical practice. Peng et al. (2020) introduced a remote access diagnostic system that allows the analysis of colposcopy images from rural clinics, thus overcoming geographic barriers to care. Singh (2025) went one step further with CelInsight, a secure web-based application that offers real-time AI-powered analysis and patient record management. These innovations illustrate the shift from isolated algorithm development to system-level integration that addresses workflow compatibility, accessibility, and clinician adoption.

Most existing models have been trained and validated using curated datasets from high-income countries. The generalizability of these models to low-resource environments, such as those in sub-Saharan Africa, is uncertain. Image quality, clinical protocols, demographics of patients and comorbidities differ substantially between regions, which can affect the accuracy and reliability of the model. This concern, widely acknowledged in the medical literature on AI, Cross et al. (2024), underscores the need for locally adapted solutions. In addition, few studies have addressed the practical deployment of AI-based diagnostic tools in healthcare systems with limited infrastructure. Models that perform well in laboratory settings may prove impractical if they require expensive computing resources, stable internet connections, or extensive retraining to accommodate new data. These barriers hinder the operationalization of otherwise promising technologies and widen the gap between innovation and impact.

The paper contributes to closing this gap by designing and implementing a hybrid AI tool specifically tailored to the context of Benin. The model combines the segmentation power of U-Net with the efficiency of an SVM classifier to ensure both accuracy and computational feasibility. Uniquely, the training data set comprises colposcopic images collected from the Centre Hospitalier Universitaire Départemental (CHUD) in Parakou, all annotated by local clinicians. This ensures contextual relevance and increases the likelihood of successful deployment. Furthermore, the model is integrated into a lightweight web-based platform accessible via standard computing devices, allowing real-time point-of-care support for clinicians operating in remote or rural areas.

2.5 Comparative analysis of existing approaches

To situate our research within the larger context of AI-assisted colposcopy, Table 1 presents a structured comparison of recent studies focused on the classification and detection of cervical lesion. This synthesis outlines the methodological frameworks, dataset characteristics, classification granularity, performance metrics, and limitations of each approach. The comparative lens underscores the specific contributions of our study, particularly in addressing gaps related to clinical relevance, data diversity, and feasibility of deployment.

Table 1: Comparison of existing AI approaches for cervical lesion classification

Study	Method	Dataset Size	Classes	Accuracy	Key Strengths	Limitations
Hu et al. (2020)	Deep CNN	60,000 images	2 (Normal, Abnormal)	AUC: 91%	Large-scale, mobile-based screening	Binary output, limited regional diversity
Chandran et al. (2021)	CYENET	1,000 images	3 (Normal, LSIL, HSIL)	92.3%	Attention-enhanced ensemble model	Small sample size, single-center validation
Yuan, Zhang and Huang (2020)	U-Net segmentation	500 images	Segmentation task	IoU: 88%	High-quality lesion segmentation	No classification output
Jiménez Gaona et al. (2022)	Hybrid CNN-SVM	963 images	4 (Normal, CIN1–3)	85.2%	Combines deep features with classical ML	Limited generalizability, modest accuracy
Xue et al. (2020)	Deep learning + clinical integration	8,142 images	4 (Normal, CIN1–3)	87.5%	Embedded in clinical workflow	Restricted to specific hospital settings
Singh (2025)	Web-based AI platform	Not specified	6 classes	Not reported	Online accessibility	No performance metrics, lacks peer-reviewed validation

Among the emerging contributions to AI-assisted cervical lesion classification, our work introduces a multiarchitecture framework that integrates three distinct methodological approaches: (i) a hybrid segmentation-classification system combining U-Net and Support Vector Machine (SVM), (ii) a transfer learning approach leveraging the VGG-19 architecture, and (iii) the specialized CYENET model designed for cervical abnormality detection. The selection of these three distinct approaches was guided by the specific requirements for the classification of cervical lesion and the practical constraints of the implementation of AI-based diagnostic tools in resource-limited settings. The U-Net combined with the SVM approach is chosen because the diagnosis of colposcopy naturally involves two sequential tasks: identifying the cervical region and classifying pathological features. This mirrors the clinical workflow where physicians first locate anatomical structures before assessing abnormalities. U-Net has proven to be effective for biomedical image segmentation tasks, particularly in medical imaging, where precise boundary delineation is important (Ronneberger et al., 2015). Integration with SVM provides robust classification capabilities that perform well with lim-

ited training data, and additionally, this hybrid approach requires fewer computational resources compared to end-to-end deep learning models, making it practical for deployment in settings with limited infrastructure.

The VGG-19 transfer learning is selected due to its demonstrated success in medical imaging applications and its ability to leverage pre-trained visual representations. The 19-layer architecture captures both fine-grained textural details and higher-level pathological patterns to distinguish different stages of cervical intraepithelial neoplasia. Transfer learning addresses the fundamental challenge of limited medical training data by utilizing knowledge gained from large-scale natural image datasets. This approach has been validated across multiple medical imaging domains and provides a reliable foundation for adapting general visual understanding to specific pathological classification tasks. The two-phase training strategy allows for systematic adaptation of pre-trained features to the unique characteristics of colposcopic images.

CYENET is included as it represents a specialized architecture designed specifically for the detection of cervical abnormalities (Chandran et al., 2021). Unlike general-purpose architectures, CYENET incorporates attention mechanisms that highlight diagnostically relevant regions, directly addressing the clinical need for interpretable AI systems in medical diagnosis. The attention-based visualization capability allows clinicians to understand which image regions influenced the model's decision, supporting clinical validation, and building trust in AI-assisted diagnosis. This model serves as a reference for comparing domain-specific approaches with adapted general-purpose architectures.

These approaches give different perspectives on the classification problem: classical machine learning combined with deep learning for segmentation, transfer learning from natural images to medical domains, and a specialized architecture designed for the specific medical task. This selection enables the identification of the most suitable approach, while ensuring that the comparison covers fundamentally different AI methodologies. The choice deliberately excludes computationally intensive architectures, such as ResNet variants or transformer-based models, which would be impractical given the original dataset size and the requirement for deployment in resource-constrained environments.

Unlike conventional binary or three-class classification schemes, our approach adopts a six-class taxonomy: normal, CIN1, CIN2, CIN3, metaplasia, and ectropion, reflecting a more granular and clinically meaningful diagnostic structure. Furthermore, research is tied into a practical deployment strategy through a Web-based diagnostic platform, designed to support real-time clinical decision making in resource-constrained environments.

3 Data and Methodology

This section describes the development and implementation of a convolutional neural network (CNN) designed to classify and visualize affected areas in colposcopy images for the detection of cervical cancer. The methodology involves data pre-processing, model training, validation, and testing phases, with a focus on optimizing performance and ensuring accurate diagnostics.

3.1 Data acquisition

The dataset used in this study was obtained from the Centre Hospitalier Universitaire Départemental du Borgou/Alibori (CHUD-B/A) in Parakou, Benin. This selection was based on the request of medical personnel for innovative solutions to enhance diagnostic reliability and limit cases of under- or over-screening. Before data collection, an initial field visit was conducted to establish direct contact with hospital personnel and engage in structured discussions, particularly with cancer specialists. These exchanges facilitated a comprehensive understanding of the clinical framework and enabled a precise alignment between the diagnostic categories observed and their medical importance. Data were collected as part of routine cervical cancer screening activities carried out with patients attending consultations. The examinations were conducted using a standard digital colposcope, which allowed the capture of images of the cervix. The acquisition process followed normal clinical procedures without altering the medical protocol. Each image represents a colposcopic view of the cervix captured before and after the application of acetic acid, according to visual inspection-based screening practices. All data were anonymized before being made available for this research, according to ethical principles and confidentiality standards. The initial dataset contained 501 images in JPG format.

3.2 Data Annotation

Each colposcopic image was annotated by CHUD Parakou medical personnel according to diagnostic categories derived from visual examination. The labeling process was guided by standardized clinical frameworks, specifically the Cervical Intraepithelial Neoplasia (CIN) classification system, which provides a structured approach to assess precancerous cervical lesions.
To ensure the accuracy and clinical relevance of the annotations, several collaborative sessions were held with oncological specialists. These meetings facilitated a shared understanding of the diagnostic criteria and supported the categorization of images into appropriate classes. The annotation process was conducted over multiple sessions, allowing for a thorough review and expert validation. Given the medical expertise of the personnel involved, their contributions were instrumental

in ensuring that the classification of images reflected established diagnostic standards and clinical best practices. The dataset comprises six distinct diagnostic categories:
Normal: Cervical with no visible lesions,
CIN1: Low grade intraepithelial lesion,
CIN2: Moderate-grade intraepithelial lesion
CIN3: High-grade intraepithelial lesion
Metaplasia: A benign transformation of epithelial cells, often observed during reproductive age and potentially mistaken for dysplasia if not properly evaluated.
Ectropion: Non-pathological condition in which glandular cells from the cervical canal are present on the outer surface of the cervix, usually red and inflamed, but not indicative of disease.
Figure 1 shows a visualization of one image per class.

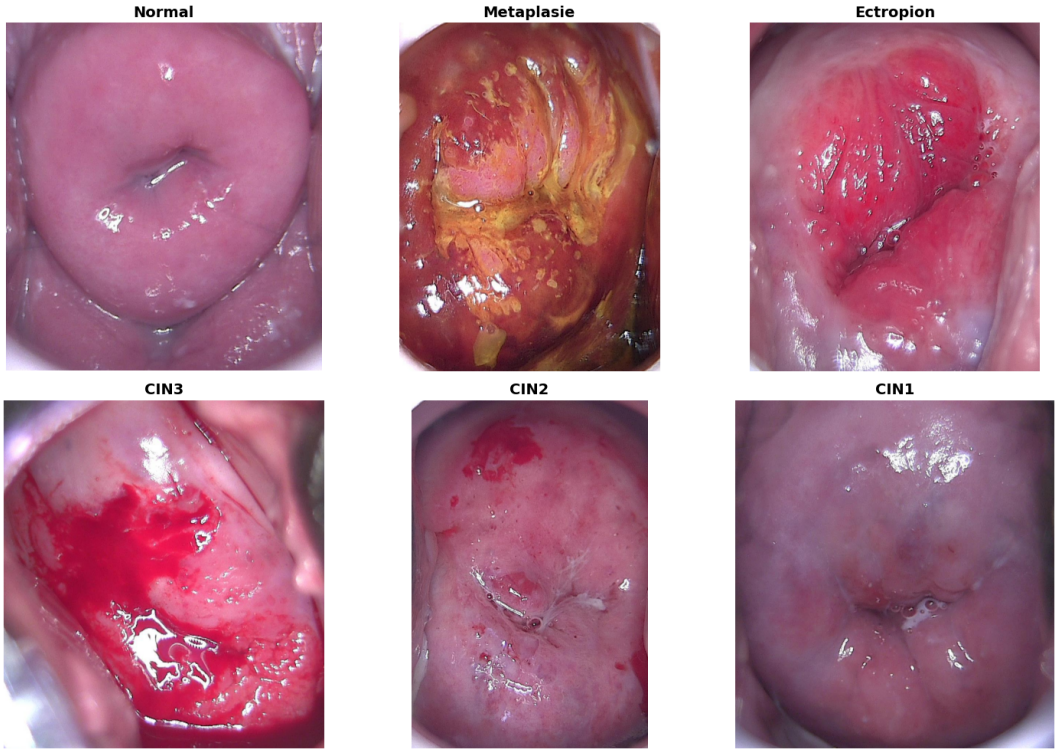

Figure 1: Colposcopy images

3.3 Data Preprocessing

Before training the models, a series of steps were taken to prepare the images and ensure that only high-quality data would be used. This process aimed to eliminate inconsistencies, correct quality issues, and make the most of the dataset through augmentation.

Image cleaning. The original dataset contained 501 images. After a quality control step, seven images were excluded due to poor focus or insufficient visibility of the cervix. Six belonged to the *Normal* category, and one to *CIN2*. These were excluded from the dataset, leaving a total of 494 usable images which were used for all subsequent analyzes. Table 2 presents the number of images in each category. The CIN1 category presents the highest number of observations, containing 138 images.

Data augmentation. Given the limited number of medical images available, data augmentation techniques were employed to increase the size of the data set and improve the generalization of the model, following approaches explored by Ribli et al. (2018) and Ronneberger et al. (2015). The following geometric and photometric transformations were applied:

- **Brightness and contrast variations**, to simulate different lighting conditions during image capture.
- **Small rotations**, to account for changes in camera positioning.
- **Horizontal flips**, to introduce diversity in anatomical orientation.
- **Gamma corrections**, through subtle gamma adjustments.

The augmentation was performed strategically to balance the dataset in all six diagnostic categories. Different numbers of augmented versions per class were generated to achieve equal representation in each category (with 300 images per category), resulting in a final dataset of 1800 images. This class-balancing augmentation strategy helped increase the model's ability to generalize across different imaging conditions while addressing the original class imbalance problem and maintaining the clinical relevance of each image.

Table 2: Data distribution across the categories

Category	Original observation	Augmentation created	Total
Normal	68	232	300
CIN 1	138	162	300
CIN 2	102	192	300
CIN 3	82	218	300
Metaplasia	36	264	300
Ectropion	68	232	300
Total	**494**	**1306**	**1800**

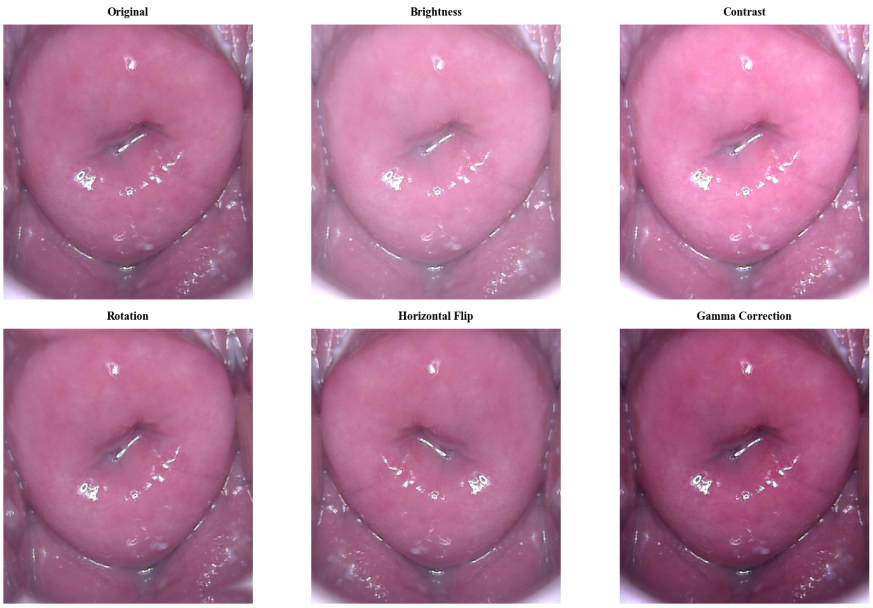

Figure 2: A normal cervix and the geometric transformations which were used in the data augmentation

Data Partitioning. Augmentation is applied before splitting. The final dataset is then randomly divided by class labels into training (70%, 1,260 images), validation (15%, 270 images) and test (15%, 270 images). The choice is based on the approach that sufficient training samples for deep neural networks to achieve convergence while providing statistically meaningful validation and test sets for unbiased performance evaluation and maintainsproportional representation across all six diagnostic categories in each subset, preventing class imbalance issues during training and evaluation phases. This partitioning method is used in medical image classification tasks to ensure a balanced approach to model training and evaluation. Mathivanan and Rajesh (2024) used the same split to train hybrid deep learning models for the detection of cervical cancer. Similarly Sarhangi and Keyvanpour (2023), in a review of 24 studies, reported that most researchers used 70/15/15 splits to support reproducibility and reduce overfitting.

3.4 Methodology

The general process of the study is summarized in Figure 3, which outlines the methodological steps taken. Three distinct approaches were implemented and compared on the basis of the literature. We have the following:

1. **Hybrid U-Net and Support Vector Machine (SVM)**: Combining semantic segmentation with traditional machine learning

2. **VGG-19 Transfer Learning**: a pre-trained deep convolutional neural network that has been adapted here to perform colposcopic image classification

3. **CYENET Architecture**:a specialized convolutional neural network architecture specifically designed for the automatic classification of cervical lesions from colposcopy images

The pipeline encompasses the entire classification process, beginning with the dataset and continuing through pre-processing, segmentation, feature extraction, model training, and performance evaluation.

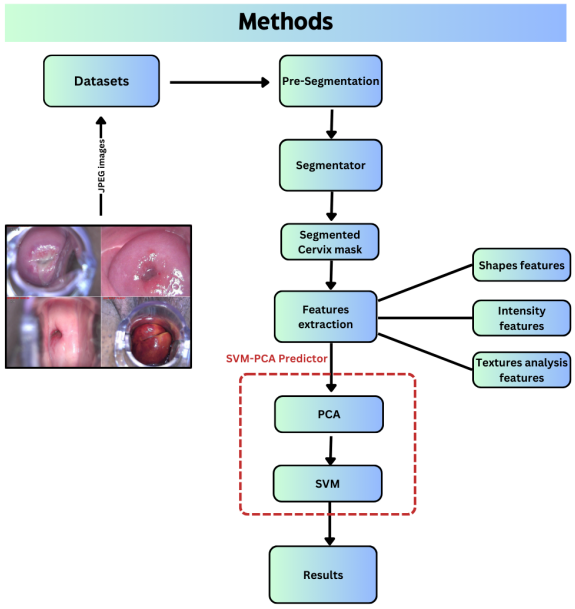

Figure 3: Overall Methodology

Datasets. JPG images of the cervical cavity were collected as input data sets and cleaned. Data went through an augmentation process that involved the variations of contrast, brightness, rotations, flips, gamma corrections.These augemented data are used for further processing.

Pre-Segmentation. This involves pre-processing the images to isolate the region of interest (ROI) for subsequent analysis. This process will be elaborated on in the next section using Figure 4.

Segmentation. In this stage, the ROI is segmented using a U-Net model to extract the cervix area. The segmentation process prepares the image for feature extraction by isolating relevant regions.

Feature Extraction : After segmentation, various features are extracted from the segmented cervix mask. These include Shape and Intensity Features and also Texture Features

SVM-PCA Predictor. The extracted features are then fed into a Principal Component Analysis (PCA) for dimensionality reduction, followed by a Support Vector Machine (SVM) for classification.

VGG19 TL and CYENET : Both models operate without the need for explicit segmentation, as the region of interest (ROI) within each image has been clearly identified during the preprocessing stage. Consequently, models are trained and evaluated directly on preprocessed data, ensuring that relevant cervical characteristics are consistently emphasized without requiring additional localization procedures.

Results : The results are analyzed using performance metrics to evaluate the effectiveness of the prediction model.

3.4.1 Pre segmentation and Region of Interest Extraction

After preprocessing, the methodology begins with a pre-segmentation phase, which consists of identifying and isolating the Region of Interest (ROI) within each image. ROI corresponds to the specific area that facilitates the identification, measurement, and assessment of cervical cancer risk. A visual illustration of the pre-segmentation and segmentation process is provided in Figure 4.
In this step, all images were manually cropped in the central zone where lesions are typically located using ImageJ software. The cropping shape was rectangular, designed to focus on the most diagnostically relevant regions. Mathematically, the cropped image I_{crop} is defined as:

$$I_{\text{crop}} = I(x_1 : x_2, y_1 : y_2)$$

where (x_1, y_1) and (x_2, y_2) represent the coordinates of the upper left and bottom right corners of the selected rectangular region, respectively.

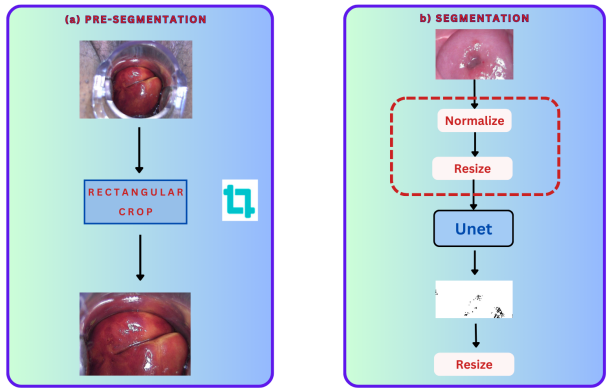

Figure 4: Detailed Pre-Segmentation and Segmentation Process

3.4.2 Segmentation

Following the pre-segmentation step, the segmentation phase employs a U-Net architecture to automatically identify and extract the cervical region from the background. This step ensures that subsequent analysis focuses exclusively on the relevant areas. The segmentation process involves:

Normalization. After cropping, the image undergoes normalization to scale the pixel values to a range of $[0,1]$, which is essential for consistent input to the segmentation model. The normalized image I_{norm} is calculated as:

$$I_{\text{norm}} = \frac{I_{\text{crop}} - \min(I_{\text{crop}})}{\max(I_{\text{crop}}) - \min(I_{\text{crop}})}$$

I_{crop} is pixel value of the cropped image (a rectangular subregion extracted from the original)

Resize. The normalized image is then resized to a fixed dimension (H,W) that is suitable for the input requirements of the segmentation model:

$$I_{\text{resize}} = \text{resize}(I_{\text{norm}}, H, W)$$

with H the height (number of pixels in the vertical dimension) and W the width (number of pixels in the horizontal dimension)

U-Net Segmentation. The resized image is input into a U-Net model, which segments the cervical region from the background. The U-Net architecture consists of an encoder-decoder structure with skip connections that preserve spatial information during the segmentation process. The en-

coder path progressively reduces the spatial dimensions while capturing contextual information, while the decoder path recovers the spatial resolution for precise boundary delineation. The output is a segmented image Iseg:I_{seg}:

$$I_{\text{seg}} = \text{U-Net}(I_{\text{resize}})$$

The U-Net model is trained specifically on cervical images to distinguish between the cervical tissue and background elements such as speculum, lighting artifacts, and other non-relevant anatomical structures.

Final Resize. After segmentation, the segmented image is resized back to its original dimensions to ensure consistency with the original data.

3.5 Model Architectures

We implement and compare the three approaches for the classification of cervical lesions. Each represents different paradigms in machine learning and deep learning methodologies. Comparative analysis allows the identification of the most effective approach while providing additional information on the strengths and limitations of each method in the context of medical image analysis.

3.5.1 U-Net combined with SVM Approach

We used ReLU (Rectified Linear Unit) as the activation function for the U-Net architecture due to its mathematical properties that are suitable for deep networks. Unlike sigmoidal functions whose derivatives decrease exponentially during backpropagation, ReLU maintains a constant derivative of 1 for positive values, avoiding the problem of vanishing gradients in deep layers. This characteristic enables efficient parameter learning across all network layers. Additionally, ReLU introduces natural sparsity by eliminating negative activations, reducing computational complexity while acting as an implicit regularization mechanism. In the context of colposcopic image segmentation, ReLU preserves important intensity variations between normal and pathological regions without numerical saturation, enabling the precise detection of the characteristic acetowhitening contrasts of cervical lesions.

Table 3: U-net training hyperparameter details

Hyperparameters	**Unet**
Number of epochs	60 with early stopping
Batch size	32
Steps	40
Steps validation	6
Optimizer	Adam
Learning rate	0.001
Loss validation	0.45
Loss function	Parse categorical crossentropy
Activation function	ReLU

Feature Extraction. Following segmentation, a total of 90 radiomic features from each region of interest were divided into first-order statistics (such as entropy and energy), a gray-level co-occurrence matrix (GLCM), a gray level run length matrix (GLRLM), a gray level size zone matrix (GLSZM), a gray-level dependence matrix (GLDM) and neighboring gray tone difference matrix (NGTDM). These features are designed to capture the morphological, intensity, and textural characteristics that are known to be diagnostically relevant in cervical pathology.

SVM-PCA Classification. The extracted features are processed through a two-stage classification pipeline that first reduces dimensionality to eliminate redundancy and noise, and then applies robust classification techniques. The features are first processed through Principal Component Analysis (PCA) for dimensionality reduction, which transforms the original feature space to a lower-dimensional representation while retaining the most informative components. The PCA transformation serves to reduce computational complexity, eliminates correlated features that could confuse the classifier, and helps prevent overfitting by focusing on the most significant patterns in the data. The number of principal components is selected using 95% of the cumulative explained variance.

Following dimensionality reduction, a Support Vector Machine classifier with a radial basis function (RBF) kernel is applied to the reduced feature space. Five-fold cross-validation was performed in the training set for hyperparameter selection (C and γ. An RBF kernel SVM classifier was trained on these features. The final model was evaluated once on the held-out test set. The SVM is particularly adapted for this application because it can handle non-linear decision boundaries and

is robust to outliers, which are common challenges in medical data. The classification decision function is given by:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right)$$

where α_i are the Lagrange multipliers, y_i are the class labels, $K(x_i, x)$ is the RBF kernel function, and b is the bias term learned during training. The RBF kernel is defined as:

$$K(x_i, x) = \exp\left(-\gamma \|x_i - x\|^2\right)$$

where γ is a hyperparameter that controls the influence of individual training examples.

3.5.2 VGG-19 Transfer Learning Approach

The VGG-19 transfer learning model uses a pre-trained deep convolutional neural network that has been adapted for cervical lesion classification. This approach takes advantage of the rich feature representations learned from natural images and adapts them to the medical domain through systematic fine-tuning.

Architecture Overview. The VGG-19 model consists of 19 layers including 16 convolutional layers and 3 fully connected layers, arranged in a systematic progression that gradually increases the complexity of learned features. The convolutional layers are organized into five blocks with progressively increasing filter sizes: 64, 128, 256, 512, and 512 filters, respectively. Each convolutional layer uses small 3×3 filters with ReLU activation functions, followed by 2×2 max-pooling layers to reduce spatial dimensions and provide translation invariance.
This architecture design philosophy emphasizes depth over filter size, allowing the network to learn complex hierarchical representations through the composition of simple operations. The small filter size ensures that the receptive field grows gradually, allowing the network to capture fine-grained details in the early layers while building up to more complex patterns in deeper layers.

Transfer Learning Implementation. The implementation follows a two-phase approach to optimize both learning efficiency and final performance. The original VGG-19 backbone was loaded with weights pre-trained on ImageNet, but the final classification layers were removed. This allows us to keep the valuable feature extraction capabilities while adapting the classification component to our specific diagnostic categories. Input images were pre-processed using the standard VGG-19

pre-processing function, which normalizes pixel values according to the ImageNet training distribution.

To replace the original classification head, we designed a custom architecture consisting of two dense layers with 512 and 256 neurons, respectively. Both layers use ReLU activation functions to introduce non-linearity. Dropout regularization was applied with rates of 0.5 and 0.3 to avoid overfitting, which is particularly important given our limited dataset size. The final output layer contains 6 neurons with softmax activation, producing probability distributions over our diagnostic categories: Normal, CIN1, CIN2, CIN3, Metaplasia, and Ectropion.

Two-Phase Training Strategy In the first phase, all convolutional layers of the VGG-19 backbone were frozen and used as fixed feature extractors. This means that only the new classification head was trained while the pre-trained weights remained unchanged. The training lasted 10 epochs with a learning rate of 1e-4 and a batch size of 4. This phase focuses on teaching the new classification layers to interpret the VGG-19 features in the context of cervical pathology without disrupting the established feature representations.

The second phase implements fine-tuning by unfreezing the final 4 convolutional blocks of the VGG-19 backbone. This allows these higher-level layers to adapt their feature extraction to the specific characteristics of colposcopic images. To prevent catastrophic forgetting of useful pre-trained features, the learning rate was reduced to 1e-5, and training continued for an additional 10 epochs. This careful fine-tuning enables the model to develop domain-specific feature representations while maintaining the fundamental visual understanding acquired during ImageNet pre-training.

Computational Considerations The VGG-19 model presents important computational demands due to its large parameter count of over 20 million parameters. Training was conducted on CPU hardware, which influenced several implementation decisions. The batch size was limited to 4 images per batch to accommodate available memory constraints. Although this small batch size extends training time, it ensures reproducibility across standard computing environments and makes the approach accessible in any resource settings.

The complete set of training parameters used for the VGG-19 transfer learning approach is summarized below.

Table 4: Training and fine-tuning configurations for VGG19 TL approach.

Parameter	**Phase 1 (Initial Training)**	**Phase 2 (Fine-tuning)**
Input Size	$224 \times 224 \times 3$	$224 \times 224 \times 3$
Batch Size	4	4
Learning Rate	1e-4	1e-5
Epochs	10	10
Optimizer	Adam	Adam
Loss Function	Categorical Crossentropy	Categorical Crossentropy
Trainable Layers	Classification head only	Classification head + last 4 conv blocks
Dropout Rates	Dense1: 0.5, Dense2: 0.3	Dense1: 0.5, Dense2: 0.3
Total Parameters	$\sim$20M (138K trainable)	$\sim$20M ($\sim$2.3M trainable)

3.5.3 CYENET Architecture

We use a custom CYENET to classify cervical cancer images efficiently. The design process was motivated by the need to balance diagnostic accuracy with memory and runtime constraints. CYENET represents a specialized convolutional neural network architecture specifically designed for cervical cancer detection from colposcopy images. This architecture was developed to address the challenges of medical image classification in cervical screening, incorporating design elements that are design to the characteristics of colposcopic images and the specific requirements of cervical pathology detection.

The full workflow of the proposed CYENET is presented in Figure 5. It illustrates the sequential process from image input, backbone feature extraction, attention-based refinement, feature enhancement, and final classification.

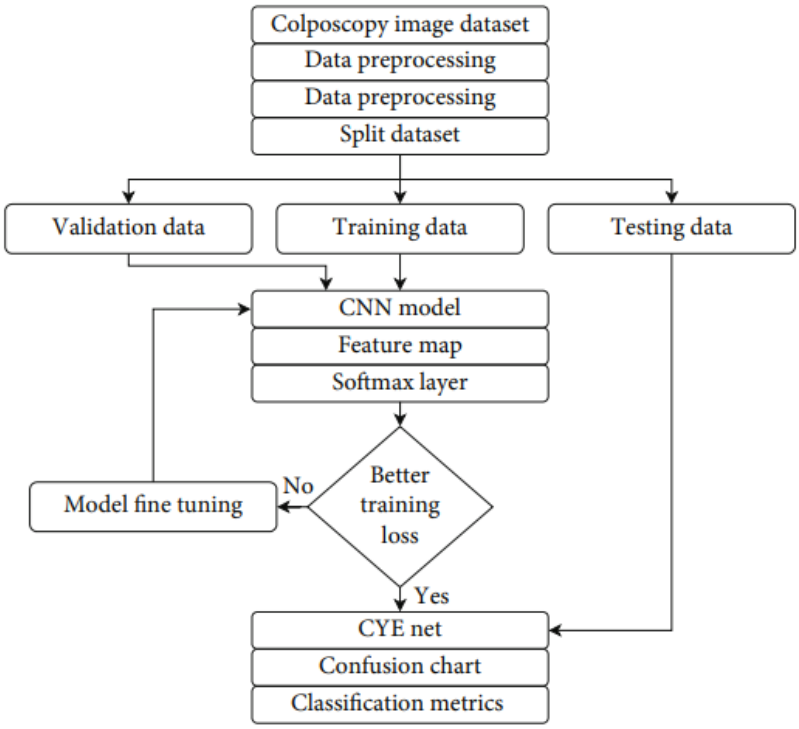

Figure 5: CYENET workflow chart

Network Design. The architecture is built on top of VGG19, a convolutional neural network pre-trained on ImageNet. We remove fully connected layers and retain the convolutional feature extractor as the backbone. To ensure that the network adapts effectively to the dataset, we fine-tuned the last four convolutional blocks of VGG19 while freezing the earlier layers. This selective fine-tuning helped preserve general visual filters while adapting higher-level filters to specific patterns of lesion in cervical images. The input images were resized to 128×128 pixels with three channels, normalized to the range [0,1], and passed into the backbone for feature extraction.
To further improve feature quality, we incorporated two custom modules after the backbone:

1. **Cyclic Attention Module** After extracting characteristics with VGG19, we introduced a cyclic attention mechanism to highlight discriminative regions in cervical images. Global average pooling was applied to the feature maps to generate contextual information, which was then passed through two dense layers to produce attention weights. These weights were used to recalibrate the feature maps using multiplicative scaling, while a residual connection ensured that the original features were preserved. This step allowed the model to emphasize diagnostically relevant patterns and suppress redundant background information.

2. **Feature Enhancement Block** The attended features were then processed by a feature enhancement block composed of two convolutional layers: the first with 512 filters and the second with 256 filters (kernel size 3×3, ReLU activations). Each convolution was followed by batch normalization, and dropout was applied after the first convolution to reduce overfitting. This block enriched the discriminative power of the extracted features by capturing more complex and fine-grained representations.

Classification Component. Finally, the refined features were passed to the classification head. A global average pooling (GAP) layer compressed the feature maps into a vector representation. This was followed by two fully connected layers of 512 and 256 units, activated by ReLU and regularized with dropout (0.5 and 0.3). The final output layer employed softmax activation to generate probabilities in all diagnostic categories.

Table 5: Proposed CYENET Configuration and Training Parameters

Parameter	**Value**
Input size	$128 \times 128 \times 3$
Backbone	VGG19 (pre-trained on ImageNet)
Trainable layers	Last 4 convolutional blocks
Attention mechanism	GAP $\rightarrow$ Dense($\cdot$) $\rightarrow$ Dense($\cdot$) $\rightarrow$ Sigmoid $\rightarrow$ Residual scaling
Enhancement block	Conv2D (512, 3×3, ReLU, BN, Dropout 0.3); Conv2D (256, 3×3, ReLU, BN)
Classification head	GAP $\rightarrow$ Dense(512, ReLU, Dropout 0.5) $\rightarrow$ Dense(256, ReLU, Dropout 0.3) $\rightarrow$ Dense(`num_classes`, Softmax)
Optimizer	Adam
Learning rate	0.001 (decayed to min 10^{-7})
Batch size	8
Epochs	30
Loss function	Categorical Cross-Entropy

3.6 Classification Function

The three approaches used (VGG19 TL, CYENET, and U-Net + SVM) produce output in the form of probability distributions in the six diagnostic categories. In the case of VGG19 TL and CYENET, these probabilities arise naturally from softmax activation in the final dense layer. For

the UNet+SVM pipeline, the U-Net serves as a feature extractor, the features are reduced via Principal Component Analysis, and the SVM classifier is trained with probability estimation enabled, allowing Platt scaling to provide calibrated posterior probabilities.
For a given input image x, each model therefore outputs:

$$p(x) = \big(p_1(x), p_2(x), \ldots, p_6(x)\big), \quad \sum_{i=1}^{6} p_i(x) = 1,$$

where $p_i(x)$ denotes the estimated probability of belonging to class i.
The predicted class is defined as:

$$\hat{y}(x) = \arg\max_i p_i(x), \tag{1}$$

and the associated classification confidence is quantified by:

$$f(x) = \max_i p_i(x). \tag{2}$$

To facilitate interpretation, the following thresholds were applied to the confidence function:

- $f(x) < 0.50$: uncertain prediction,
- $0.50 \leq f(x) < 0.75$: moderately confident prediction,
- $f(x) \geq 0.75$: highly confident prediction.

This unified formulation ensures consistency between the three approaches and provides an interpretable measure of confidence alongside categorical predictions. Such confidence estimation is particularly valuable in medical diagnostics, where understanding the reliability of a prediction is as important as the predicted outcome itself.

3.7 Performance Evaluation

To assess the performance of the three classification approaches, we use metrics that are suitable for multiclass settings. Our dataset comprises six diagnostic categories (Normal, CIN1, CIN2, CIN3, Metaplasia, and Ectropion) therefore, relying solely on binary indicators such as sensitivity and specificity would provide an incomplete picture and be inaccurate. For this reason, the evaluation is extended to include macro-averaged accuracy, precision, recall, and F1 scores, together with confusion matrices. Our dataset contains equal representation across all classes, with 45 samples per category in the validation set, totaling 270 samples. We employ a comprehensive multiclass evaluation framework to assess model performance.(Grandini et al., 2020).

3.7.1 Individual Class Performance Metrics

For each diagnostic category i (where $i = 1, 2, ..., 6$), we define the following components:

- TP_i: True positives - cases correctly identified as belonging to class i
- FP_i: False positives - cases incorrectly assigned to class i
- FN_i: False negatives - cases of class i that were misclassified as other categories.

Precision per class measures the reliability of the positive predictions of the model for each diagnostic category. This metric addresses the question: "When the model predicts a specific condition, how frequently is this prediction accurate?"

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{3}$$

Recall per class evaluates the model's capacity to identify all instances of each condition. This metric answers: "Among all actual cases of a specific condition, what proportion does the model correctly detect?"

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{4}$$

F1-Score per class combines precision and recall into a single balanced measure. This metric proves to be particularly valuable in medical applications where missed diagnoses and false alarms carry clinical consequences.

$$\text{F1}_i = 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{5}$$

3.7.2 Overall Model Performance Metrics

Overall Accuracy represents the fraction of correctly classified samples across all diagnostic categories:

$$\text{Accuracy} = \frac{\sum_{i=1}^{6} TP_i}{270} \tag{6}$$

Macro-averaged metrics assign equal weight to each diagnostic category, regardless of class frequency. This approach suits our balanced dataset and ensures that model performance on each condition receives equal consideration in the final evaluation:

$$\text{Macro Precision} = \frac{1}{6}\sum_{i=1}^{6}\text{Precision}_i \tag{7}$$

$$\text{Macro Recall} = \frac{1}{6}\sum_{i=1}^{6}\text{Recall}_i \tag{8}$$

$$\text{Macro F1} = \frac{2\cdot\text{Macro-Precision}\cdot\text{Macro-Recall}}{\text{Macro-Precision}+\text{Macro-Recall}} \tag{9}$$

3.8 Web Platform Implementation and Deployment

Following the establishment of performance evaluation metrics, the implementation of a web-based diagnostic platform was undertaken to make trained models accessible for research use. The platform was designed with two primary image input methods: traditional file upload functionality and an innovative real-time camera capture feature that allows clinicians to photograph colposcopy images directly through the Web interface.

The back-end system was built using FastAPI to handle model inference, while the front-end interface was developed using React and managed through npm package management. This implementation framework was designed to accommodate any of the three trained models (U-Net+SVM, VGG-19, or CYENET) based on the comparative performance results that would determine the optimal choice for deployment.

3.8.1 Navigation Structure and Multilingual Support

The Web application follows a modern client-server architecture, separating the user interface from the computational back-end to ensure both scalability and maintainability. This separation allows the front-end to focus on user experience while the back-end handles complex image processing and machine learning inference tasks. The architecture consists of two main components: a React-based front-end for user interaction and a FastAPI back-end for model inference and image processing.

The Web application architecture consists of four main sections designed to provide a comprehensive diagnostic workflow. The Home section serves as the platform introduction, presenting the web application's capabilities and guiding new users through the system's features. The Upload section functions as the primary diagnostic interface, incorporating both traditional file upload capabilities and the innovative direct camera capture feature for real-time colposcopy image analysis.

The AI Models section provides detailed information about the underlying algorithmic approaches, proposes transparency about the diagnostic process, and builds user confidence through techni-

cal understanding. The History section maintains a complete record of all diagnostic analyzes performed by each user, enabling longitudinal tracking of cases and supporting clinical documentation requirements. A particular innovation addressing the multilingual healthcare environment in Benin is the integration of four language options throughout the platform interface. Beyond French and English, the system incorporates the support in the Fon and Yoruba language, recognizing the linguistic diversity of healthcare providers in the region. This multilingual approach ensures accessibility for users regardless of their primary language, potentially improving adoption rates and reducing barriers to technology use in clinical practice.

The Web App can be accessed using this link EmiloVision Benin .
To further enhance accessibility, particularly for users who may have difficulty navigating the Web application, a QR code system was implemented. This QR code provides direct access to the platform, allowing users to quickly access the diagnostic tool using any smartphone or tablet camera without the need to manually type the web address.

Figure 6: QR code to access the deployed Web App (EmiloVision Benin)

3.8.2 Frontend Development with React

The frontend interface was developed using React, a component-based JavaScript framework that enables the creation of responsive and interactive user interfaces. React was chosen for its component reusability, strong ecosystem support, and ability to create smooth user experiences that are essential in clinical environments where efficiency and clarity are paramount.
The user interface design prioritizes the simplicity and integration of clinical workflows. The main interface presents a clean image upload area where clinicians can either drag and drop colposcopy

images or browse their local files. Once an image is uploaded, the interface provides immediate visual feedback and displays the analysis progress. The results are presented in a clear and structured format that includes the predicted diagnosis, areas where the lesions are located.
The interface includes several features specifically designed for clinical use. A side-by-side comparison view allows clinicians to examine the original image alongside the AI analysis results. The prediction display shows not only the primary diagnosis but also the confidence scores for all six categories (Normal, CIN1, CIN2, CIN3, Metaplasia, and Ectropion), giving clinicians insight into the model's decision-making process.
Error handling and user feedback mechanisms were implemented throughout the interface to ensure that clinicians receive clear information about system status and any potential problems. The loading indicators provide visual feedback during image processing, while the error messages guide users through the resolution of common problems such as unsupported file formats or network connectivity issues.

3.8.3 Model Integration and Deployment

The integration of the final trained model that we retained after analysis into the web platform required careful attention to both performance and reliability considerations. The model, saved in TensorFlow's keras format, *.h5* is loaded during backend initialization using optimized loading procedures that minimize startup time while ensuring model integrity.The prediction pipeline within the web application maintains strict consistency with the training and evaluation procedures. This includes identical image pre-processing steps, the same normalization parameters, and equivalent post-processing of model output. This consistency is crucial to ensure that the deployed model performs as expected on the basis of the validation results reported in the research.
To handle the computational demands of real-time inference, the backend implements several optimization strategies. Image processing operations are performed asynchronously to prevent blocking other system operations. The model inference utilizes optimized TensorFlow operations and, where available, hardware acceleration through GPU resources or optimized CPU implementations.

3.8.4 Design of the user interface and workflow

The Web interface was designed to emphasize clarity, speed and potential integration with colposcopy diagnostic practices. The main workflow begins with user authentication, followed by image capture or upload during colposcopy examinations, and immediate analysis through the web platform. The results display is structured to support clinical decision making without overwhelming users with technical details. The primary prediction is prominently displayed with clear visual

indicators, while detailed probability distributions are available through expandable sections for clinicians who wish to examine the model's confidence across all diagnostic categories.
The navigation and workflow elements were designed to minimize the learning curve for medical personnel who may have limited experience with digital diagnostic tools. The interface uses familiar metaphors and clear labeling to guide users through the analysis process.
Since the platform is in its initial phase of deployment, access is restricted. Users are required to create an account and obtain special permission (added as researcher)before using the system. At this stage, the predictions generated by the platform cannot be considered official medical results or a substitute for clinical diagnosis. The deployment is intended primarily for research and evaluation purposes, allowing the system to be tested and refined before potential clinical use over the country.

3.8.5 Performance and Scalability Considerations

The Web platform was designed to handle the computational and networking demands expected in any environment. Backend performance optimization includes efficient image handling, optimized model inference, and appropriate caching strategies to reduce response times for repeated analyses. The FastAPI backend offers asynchronous request handling, allowing the system to process multiple images concurrently when system resources permit. This capability is particularly important during busy clinical periods, when multiple examinations may be performed simultaneously. The frontend includes appropriate loading states and progress indicators to manage user expectations during image processing.
Network considerations were particularly important given the infrastructure constraints common in many healthcare settings. The application was designed to function reliably on standard broadband connections while providing graceful degradation for slower networks. Image compression and optimization techniques reduce upload times without compromising the quality needed for an accurate analysis.

3.8.6 Deployment and accessibility

The deployment strategy prioritizes accessibility and ease of maintenance. The web-based nature of the platform means that clinicians can access the diagnostic tool from any computer with internet connectivity and a standard web browser, eliminating the need to install or maintain specialized software. The platform supports standard web browsers, including Chrome, Firefox, Safari, and Edge. Responsive design principles enable the interface to function effectively on various screen sizes, from desktop workstations to tablets and smartphones, providing flexibility for different clinical environments.

Security considerations include secure image upload handling, appropriate data privacy protections, and secure communication protocols. Although detailed patient information is not stored within the system, image handling follows best practices for medical data protection, including secure transmission and temporary storage procedures.

3.8.7 Regulatory Compliance and Platform Governance

The deployment and operation of the EmiloVision Benin platform are subject to regulatory oversight and governance frameworks established by relevant authorities. The access and use of the platform must comply with rules and procedures stipulated by the Ministry of Digital Affairs (Ministère du Numérique et de la Digitalization) of Benin, as well as protocols defined by the Center Hospitalier Universitaire Départemental de Parakou.

The platform incorporates user authentication and authorization mechanisms to ensure that only qualified medical personnel and approved researchers can access the diagnostic system. All users are required to acknowledge and adhere to the terms of use, which include provisions for patient data protection, appropriate clinical use, and compliance with national healthcare regulations.

Any violation of the established rules and procedures set by the hospital administration or government entities will be subject to appropriate disciplinary sanctions. Serious breaches, including unauthorized data access, misuse of diagnostic outputs, or violations of patient confidentiality, may result in legal proceedings initiated against the offending party in accordance with Benin's healthcare and data protection laws. This governance framework ensures responsible deployment, maintains the integrity of the diagnostic system, and protects patient rights and data privacy.

3.8.8 Testing and Validation

Unit tests verify individual component functionality, while integration tests confirm proper communication between frontend and backend systems. End-to-end testing simulates complete user workflows to identify potential issues in realistic usage scenarios. Performance testing evaluated system response times under various load conditions and network configurations. These tests confirmed that the platform maintains acceptable response times even when processing high-resolution colposcopy images or operating under suboptimal network conditions.

The user acceptance test provided valuable feedback on interface design, workflow integration, and practical usability. This feedback led to several interface refinements and workflow optimizations that improved the platform's clinical utility.

The deployment of this Web-based platform represents a significant step toward making the diagnosis of cervical cancer assisted by AI accessible in all regions of the country. By combining

advanced machine learning capabilities with user-friendly web technology, the platform demonstrates how academic research can be translated into practical tools that support clinical decision making and potentially improve patient outcomes.

4 Results

Following the methodology in Section 3, each of the three classification approaches was trained and evaluated in the augmented dataset, with the results analyzed in multiple performance dimensions to identify the most appropriate approach for the detection of cervical cancer.

4.1 Classification using U-Net and SVM approach

The hybrid U-Net and SVM approach represents a two-stage methodology that combines the semantic segmentation capabilities of deep learning with the robust classification performance of traditional machine learning. This approach follows the established pipeline described in Section 3.5.1, where U-Net first segments the cervical region from colposcopic images, followed by feature extraction and SVM classification.

4.1.1 Cervix Image Segmentation with U-Net

The analysis of the performance curves of the U-Net model, Figure 7, shows a stable and progressive convergence during training over 60 epochs. The segmentation network achieved effective performance in isolating cervical regions from background elements, with validation loss stabilizing at approximately 0.45 after 60 epochs. The model showed a consistent improvement across all evaluation metrics during training, with the DICE coefficient reaching satisfactory levels for image segmentation tasks. The precision and precision curves indicate that the model successfully learned to distinguish between cervical tissue and background elements such as speculum artifacts and lighting variations.

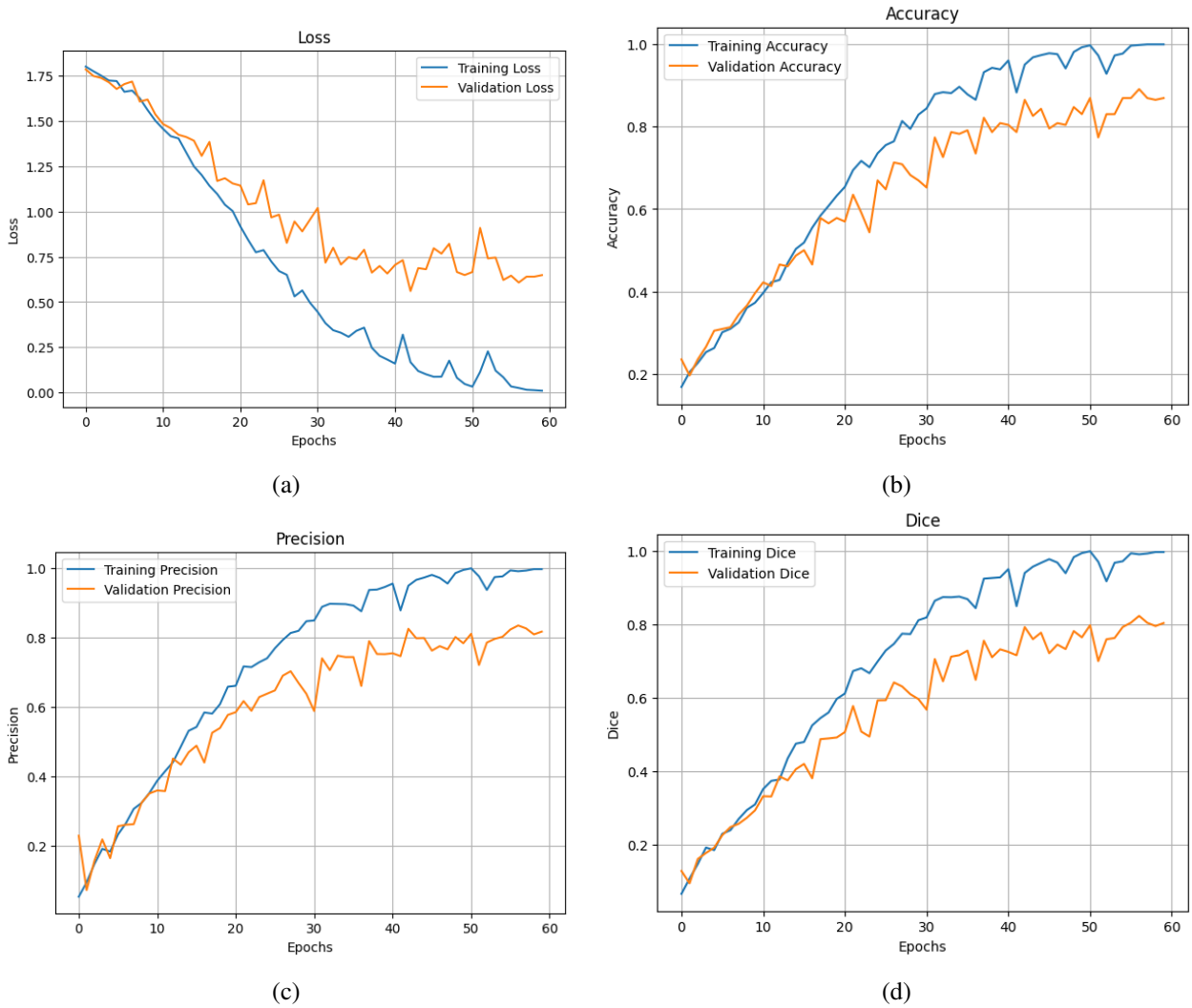

Figure 7: Metrics, ((a) Loss, (b) accuracy, (c) precision, and (d) DICE concerning the different number of epochs during the training and validation of the U-Net architecture

4.1.2 Classification with SVM

Following the segmentation phase, 90 radiomic features were extracted from each region of interest using pyradiomics, including first-order statistics, the gray-level cooccurrence matrix (GLCM), the gray-level run-length matrix (GLRLM), and other textural characteristics. These features were reduced in dimension through Principal Component Analysis before classification with the SVM model.

At the global level, the model achieved an accuracy of 84% with a macro-averaged precision of 85%, recall of 84%, and F1-score of 84%. These balanced values indicate that the classifier did not privilege one category at the expense of another and was able to handle the six classes with

relatively uniform performance.

The analysis of the confusion matrix Figure 8 in the test data gives further details on how the model behaved in the categories. The classifier was particularly strong in identifying normal tissue, with very few false positives, confirming that the model rarely confused healthy images with pathological ones. For CIN3, the predictions were also highly reliable, as nearly all cases were correctly classified, giving an F1 score of 86%. Ectropion showed excellent results, with a recall of 100% which means that almost every case was identified correctly and very few errors occurred in assigning this condition. However, some weaknesses were observed in the lower-grade lesions. CIN1 and CIN2 were misclassified more frequently, especially due to confusion between the two categories. Over the 45 images in each of the two classes, only 31 and 35 are correctly classified, respectively. The global picture suggests that the U-Net+SVM pipeline provides a reliable classification framework, particularly for normal tissue, ectropion, and advanced lesions such as CIN3. The main limitation lies in distinguishing intermediate precancerous stages (CIN1 and CIN2), where biological differences are subtle and misclassification is common even for trained clinicians.

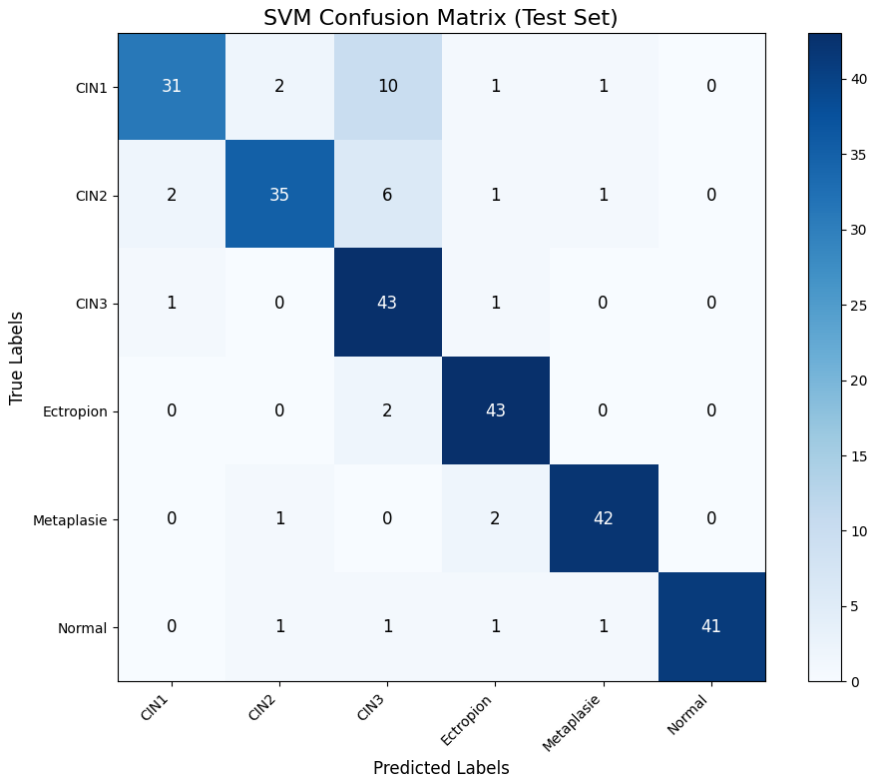

Figure 8: Confusion matrix for SVM Test

To complement the quantitative metrics, we conducted a qualitative evaluation of the model predictions by selecting 12 images at random from the test set and comparing their true labels with the

predicted outcomes (Figure 9. Of the 12 cases, nine were correctly classified, while three resulted in misclassification. Two of the errors occurred with CIN2 images, and one with an ectropion image. Interestingly, in the case of CIN2, although the final prediction was incorrect, the model still assigned a probability higher than 75% to the true class. This suggests that the classifier was partially confident in the correct label, but ultimately leaned toward a competing class with slightly higher confidence.

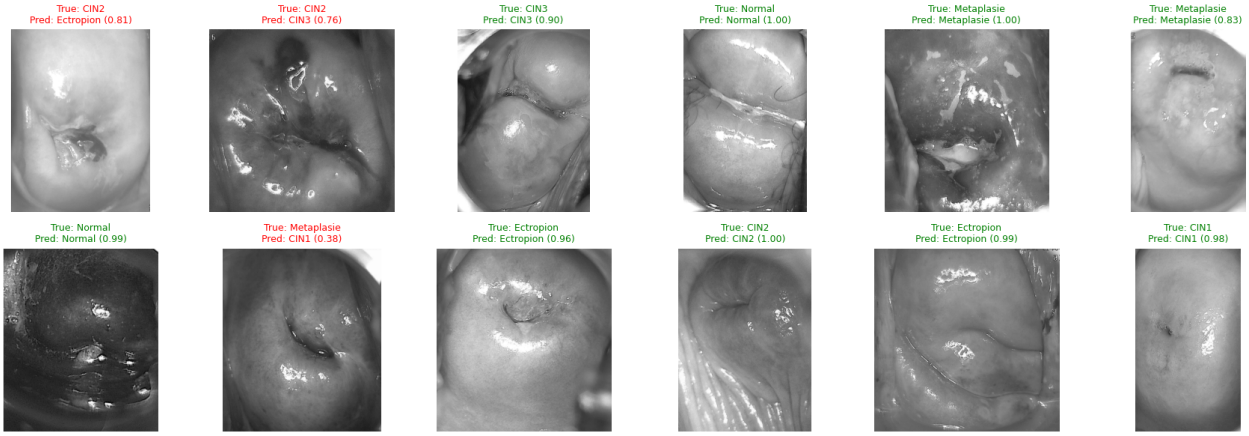

Figure 9: Representative Test Images with True Labels and U.Net-SVM Predicted Labels

For each panel, the top line shows the ground-truth label ('True') and the bottom line shows the SVM prediction ('Pred') with the associated probability in parentheses. The predictions were made using features extracted from a U-Net model and reduced via PCA. The green titles indicate correct predictions, while the red titles indicate misclassifications. The images were randomly selected to provide a qualitative illustration of the performance of the model.

4.2 Result of VGG19 TL model

The performance evaluation of the VGG-19 transfer learning model shows that the model demonstrated a strong capacity to adapt pre-trained visual representations to the classification of colposcopy images. The training and validation curves presented in Figure 10 provide a visualization of the model's convergence process and stability. The loss function decreased steadily across epochs, showing no sign of abrupt fluctuations, indicating that the model effectively minimized classification errors during training. Consequently, the accuracy curve progressively increased before reaching a plateau, suggesting that the network achieved generalization without evidence of severe overfitting. The precision and recall values also followed a consistent upward trend, strengthening the model's ability to correctly identify both normal and pathological categories. Together,

the graphical evolution confirms that the fine-tuning strategy successfully adjusted the higher convolutional layers of the network to the specific features of cervical lesions.
Quantitatively, the transfer learning approach with VGG19 showed the best overall performance compared to the U-Net combined with the SVM approach. At the global level, the model achieved an accuracy of 89.17%, with precision, recall, and F1-score all balanced around 89%. This result indicates that the classifier was able to consistently recognize both normal and pathological categories with limited variation between metrics. The stability of the training and validation curves confirms that the network successfully adapted pre-trained visual representations to the specific characteristics of colposcopy images. In the results per class Table 7, VGG19 showed strong and balanced performance in almost all diagnostic categories. For CIN1, the F1-score reached 86.9%, reflecting reliable detection of low-grade lesions. CIN2 achieved an F1-score of 83.6%, slightly lower but still satisfactory given the clinical difficulty of distinguishing moderate lesions. CIN3 was classified with high precision and recall, yielding an F1-score of 90.8%, which shows the model's ability to recognize advanced intraepithelial neoplasia with minimal error. Ectropion and metaplasia were also accurately identified, with F1-scores of 90.2% and 90.3%, respectively. Finally, normal tissue achieved the best balance, with an F1-score of 92.7%, confirming that the model rarely confused healthy tissue with abnormal conditions.

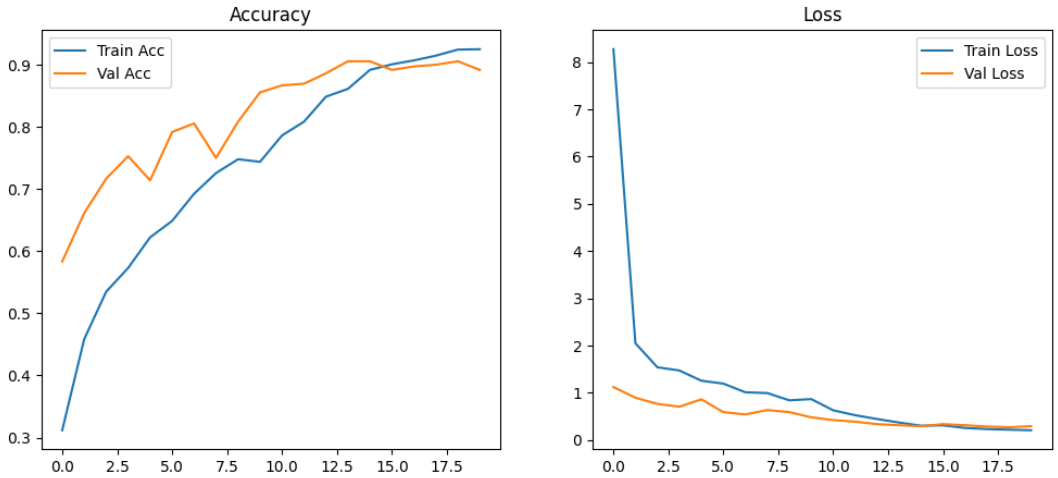

Figure 10: VGG19 (TL) Performance evaluation

Following this global assessment, the classification ability of the model is examined through the evaluation of test images. Figure 11 illustrates twelve randomly selected annotated images with their predicted labels. Out of these, 11 were correctly classified, confirming the strong reliability of the model even in small random samples. The only misclassification occurred for a CIN3 image, which was predicted as CIN1 with a probability of 0.3. This case illustrates the challenge of sepa-

rating advanced lesions from lower-grade abnormalities when there are subtle visual overlaps. This observation highlights two important points. First, the very low number of errors in the qualitative sample aligns with the high quantitative performance reported earlier, reinforcing the robustness of the VGG19 model. Second, the probability distribution attached to the misclassified image provides meaningful information: although the prediction was incorrect, the relatively low confidence (0.3) indicates that the model was uncertain.

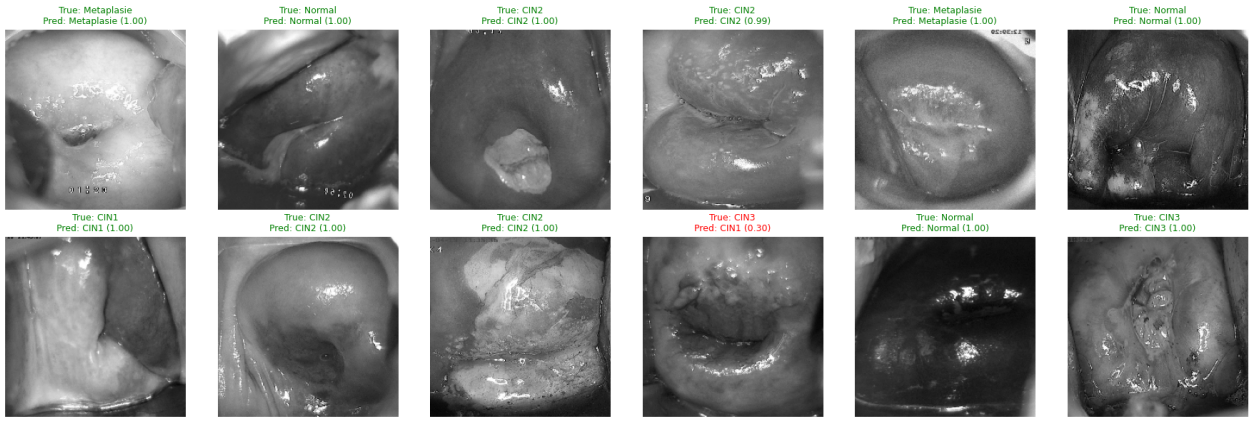

Figure 11: Classification of 12 ramdom annoted images based on VGG19 approach

4.3 Result of CYENET model

The CYENEt model achieved an overall accuracy of 85.2%, with precision, recall, and F1-score all balanced at 85%. The top k accuracy was very high (99.6%), which means that the correct class was almost always among the top predictions. Looking at individual classes, the model performed very well in detecting Metaplasia and Ectropion, with recall values of 1.00 and 0.98 in the tin test set. Although this was slightly below the performance of VGG19, the model nevertheless demonstrated strong classification ability and remained competitive in the six diagnostic categories.

At the class level, the results reveal that CYENet excelled in recognizing ectropion and metaplasia, where the F1-scores reached 96%. This indicates that the model was particularly effective in detecting benign or inflammatory changes, correctly identifying almost every case of these conditions. In contrast, the model struggled with the precancerous CIN categories. CIN1 and CIN2 both reached an F1-score of 77% and 76%, respectively, and CIN3 was slightly higher at 74% (Table 7). The confusion matrix shows that these categories were frequently misclassified with each other. This behavior is expected, as CIN lesions often share overlapping morphological features that make fine-grained distinctions challenging, even for experienced pathologists. The relatively lower and more variable scores in these classes are also reflected in the higher standard deviation

of performance between categories compared to the other models.

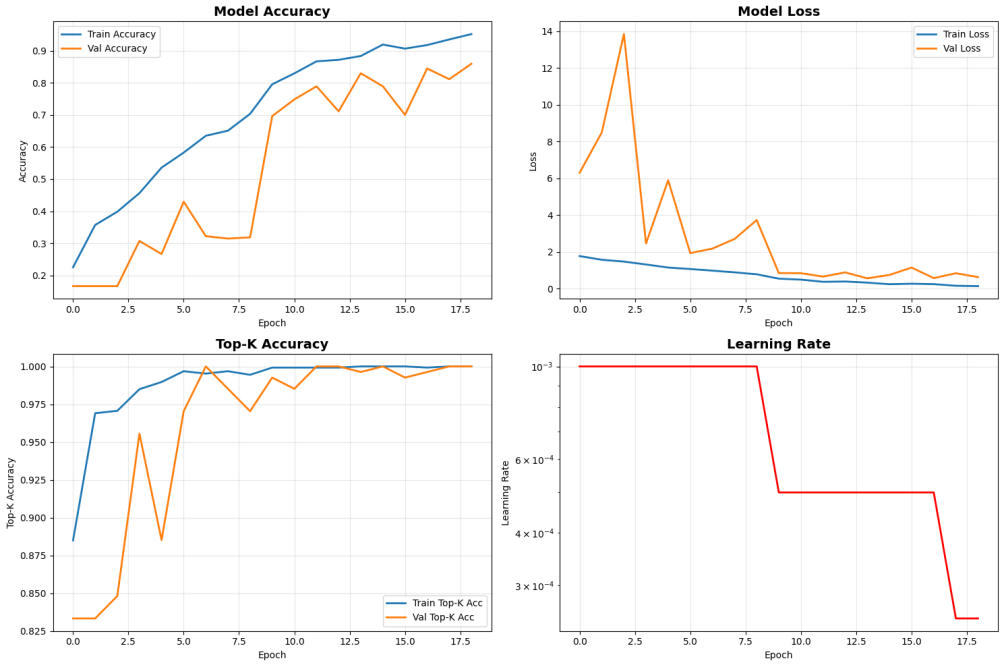

Figure 12: CYENET Performance Evaluation

Using the trained model, we test the performance of the model with 270 images. The confusion matrix in Figure 12 shows that CYENET correctly classified most images in the six categories. The classes Metaplasia, Ectropion, and Normal recorded the highest number of correct classifications, with very few misclassified samples. This confirms that the model can clearly separate these benign or non-cancerous conditions from precancerous lesions.

For the categories CIN1, CIN2, and CIN3, the confusion matrix revealed more overlap. Some CIN1 samples were predicted as CIN2, and several CIN3 samples were predicted as CIN2. This indicates that the model sometimes confuses neighboring grades of cervical intraepithelial neoplasia. This behavior is expected, since the visual boundaries between the stages of CIN are often subtle even for medical specialists. Despite this difficulty, the number of correct predictions remained higher than the number of errors for each category of CIN, showing that the model retained a reasonable ability to identify the stage of injury.

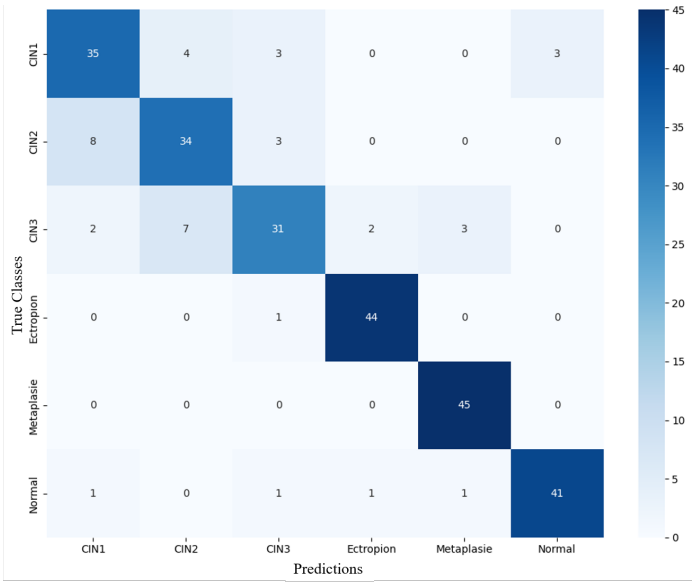

Figure 13: Confusion Matrix through test with CYENET

Visualization of affected Areas

Figure 14 presents how CYENET reached its decisions, the visualization of affected areas using Grad-CAM heatmap. We randomly chose those CIN category and, for each category, the model concentrated on regions of the cervix where epithelial changes were most visible. In the case of CIN1, the highlighted areas were limited to small patches, corresponding to low-grade localized lesions. For CIN2 or CIN3, the heatmap covered wider regions with denser activation, reflecting the larger and more irregular cellular transformations associated with higher-grade lesions.

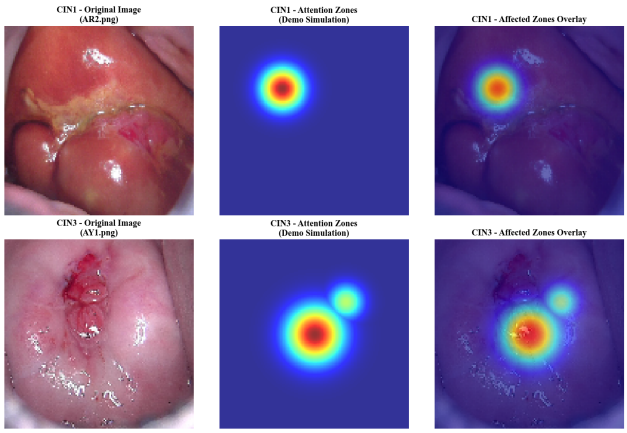

Figure 14: Original image, attention zones, and affected area overlays for CIN1 and CIN3

4.4 Comparison of the models

The overall performance of the three approaches is summarized in Table 6. Among them, the VGG19 transfer learning model achieved the best results, with an accuracy of 89.17% and balanced scores for precision (89.28%), recall (89.17%), and F1-score (89.22%). This confirms that transfer learning from large image datasets provides strong feature representations, which are effective in adapting to colposcopy images. By comparison, U-Net combined with SVM reached an accuracy of 84% with precision, recall, and F1-Scores around 84%, while CYENET obtained 85.19% accuracy with slightly higher F1 score (85%). These results indicate that, while all models are reliable, VGG19 consistently delivers the highest predictive power.

Table 6: Comparison of the three models Performance

Model name	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
UNET+SVM	84.00	85.00	84.00	84.00
VGG19 (TL)	89.17	89.28	89.17	89.22
CYENET	85.19	85.00	85.00	85.00

A more detailed view of class-level performance is presented in Table 7. Here, VGG19 again shows superiority in the clinically significant categories of CIN1, CIN2, and CIN3. For instance, it achieved an F1-score of 86.88% for CIN1 and 90.76% for CIN3, both higher than those of U-Net + SVM and CYENET. This suggests that VGG19 is better at distinguishing between different precancerous stages, which is crucial for early detection. On the other hand, U-Net combined with SVM displayed its strength in recognizing normal tissue (F1 = 90%) and ectropion (F1 = 92%), showing robustness in separating pathological from non-pathological cases. CYENET, while less accurate for CIN classes (F1 = 74–77%), excelled in benign conditions such as metaplasia and ectropion, reaching F1-scores of 96% in both categories. This suggests that its specialized architecture favors the recognition of non-precancerous abnormalities.

Table 7: Per-class Precision, Recall, and F1-Score for the three models

Class	**U-Net+SVM**			**VGG19**			**CYENET**		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
CIN1	85.00	64.00	73.00	85.48	88.33	86.88	76.00	78.00	77.00
CIN2	68.00	80.00	73.00	92.00	76.67	83.64	76.00	76.00	76.00
CIN3	85.00	87.00	86.00	91.53	90.00	90.76	79.00	69.00	74.00
Ectropion	85.00	100.0	92.00	88.71	91.67	90.17	94.00	98.00	96.00
Metaplasia	97.00	78.00	86.00	87.50	93.33	90.32	92.00	100.0	96.00
Normal	88.00	93.00	90.00	90.48	95.00	92.68	93.00	91.00	92.00

The comparison of model complexity in Table 8 shows that the three approaches also differ significantly in computational requirements. VGG19 had the largest number of parameters (more than 20 million), leading to the longest training time (95 minutes) and the slowest inference speed (2.4 images per second). U-Net combined with SVM was the lightest configuration, with fewer than 0.4 million parameters, requiring only 42 minutes of training and providing the fastest inference speed (3.5 images per second). CYENET, with around 11 million trainable parameters, stood between these extremes. It required 79 minutes of training and achieved an inference speed of 2.8 images per second. Memory usage was relatively similar between models, ranging from 0.37 to 0.48 GB, indicating that parameter size and computational load, rather than storage requirements, were the main determinants of efficiency.

Table 8: Comparison of the three models using the model complexity.

Model	**Parameters**	**Memory Usage (GB)**	**Training Time**	**Speed**
VGG-19 (TL)	12,978,438	0.37	95 min	416 ms/img (2.4 img/s)
U-Net + SVM	388,41	0.48	42 min	280 ms/img 3.5 img/s)
CYENET	11,016,83	0.40	79.1 min	350 ms/img (2.8 img/s)

Based on the comparative analysis, we selected VGG-19 as the deployment model. With an accuracy of 89.17% compared to 85% for CYENET, VGG-19 achieved a gain of more than four (04) percentage points. Although this difference may appear modest, it corresponds to approximately twelve (12) additional correct classifications per 270 test cases. In a medical diagnostic context, these improvements translate into meaningful clinical benefits. Furthermore, VGG-19 demonstrated balanced precision and F1-Score (89.28% and 89.22%, respectively), ensuring both reliable detection of abnormal cases and minimization of false positives.

4.5 Discussion

The primary objective of this study is to design and evaluate a web-based diagnostic platform capable of analyzing colposcopy images and providing immediate feedback to clinicians. We perform this by comparing three different technological approaches for cervical lesion classification, a hybrid U-Net combined with the Support Vector Machine (SVM), a VGG-19 transfer learning model and the specialized CYENET architecture. The results achieved give suitable information for the model to choose for deployment.

The hybrid U-Net, combined with the SVM approach, demonstrated effective segmentation capabilities, isolating regions of interest, and achieving stable results. The presegmentation phase successfully identified cervical regions by manual cropping, while the U-Net segmentation further refined these regions by automatically distinguishing cervical tissue from background elements. However, its classification performance was relatively modest, with an accuracy score (84%) lower than the VGG19 TL approach (89.17%). This outcome aligns with the findings of Jiménez Gaona et al. (2022), who reported 85.2% accuracy using a hybrid U-Net approach with SVM for cervical lesion classification, showing that while U-Net segmentation produced reliable lesion identification, subsequent SVM classification struggled to achieve high precision.

In contrast, the VGG-19 transfer learning model achieved the best overall performance among the three approaches tested. Its ability to leverage pre-trained weights on large-scale datasets allowed it to adapt effectively to the colposcopy dataset despite the limited number of images. This is in line with the findings of Chandran et al. (2021), who achieved 73.3% accuracy with VGG-19 in their original implementation, although our results (89. 17%) demonstrate superior performance, probably due to our two-phase training strategy and the characteristics of the data set. The two-phase training strategy, involving initial feature extraction training followed by fine-tuning of the final convolutional blocks, proved particularly effective in adapting the pre-trained features to cervical pathology characteristics. In addition, Xue et al. (2020) and Yuan, Wei and Wang (2020) also demonstrated that transfer learning improves classification accuracy in medical imaging tasks with relatively small datasets. In this study, VGG-19 produced the highest accuracy and balanced sensitivity and specificity, making it the most suitable option for integration into the web-based platform.

The CYENET architecture, specifically designed for colposcopic image analysis, also produced competitive results despite achieving lower overall accuracy (85.19%) compared to VGG-19. Chandran et al. (2021) originally reported the precision of 92.3% with their CYENET model, although direct comparison is difficult due to different datasets. Our implementation revealed that while CYENET achieved good classification scores and visualization capabilities through attention mech-

anisms, it required more computational resources during training compared to the VGG19 TL approach. The attention-based visualization feature of CYENET, which highlighted affected regions in cervical images, provided valuable clinical information, even though overall classification performance was lower than expected.

Recent comprehensive reviews support the potential of AI in cervical cancer screening as Vargas-Cardona et al. (2024) found that deep learning methods (CNN, ResNet, VGG) showed the best diagnostic performances, with an accuracy of more than 80% Artificial intelligence for cervical cancer screening, while meta-analysis results demonstrate that AI systems show superior accuracy compared to experienced clinicians in colposcopic examinations (Liu et al., 2024). Our VGG-19 results (89.17% accuracy) fall within the competitive range of these reported performances.

The particular contribution of this study lies in the development of a functional web-based platform that integrates the best-performing model and provides real-time visualization of predictions. The multilingual support of the platform (French, English, Fon, and Yoruba) addresses the linguistic diversity of languages and understanding, while the dual input methods (file upload and camera capture) accommodate different clinical workflows.

However, we identified as first limitation the relatively small number of original observations (494 images) and the quality of the images collected which restricted the diversity of training data despite expansion to 1,800 images through augmentation techniques. Although data augmentation was used to balance classes and improve generalization, artificially generated images cannot fully replicate the variability of real-world clinical cases. Second, the computational resources available for this study were limited. The computational constraints of training on standard hardware rather than high-performance servers or GPUs influenced the choice of model architecture, batch sizes, and training duration, potentially preventing models from reaching optimal performance. Most importantly, no external validation was performed with independent datasets from other hospitals or regions, which limits the generalizability of the findings beyond the CHUD-Parakou context.

The paper shows the feasibility of deploying deep learning models for colposcopic image analysis within a web-based diagnostic system accessible in rural areas. Comparative analysis clearly shows the superiority of transfer learning approaches, VGG-19 for this specific application, while also showing the complementary role of segmentation-based methods and specialized architectures such as CYENET for attention-based visualization. The successful integration of AI models into a multilingual practical web platform represents a significant step towards addressing the challenges of cervical cancer screening in Benin. Future work should focus on expanding the data set with larger and more diverse samples from multiple hospitals in Benin, performing external validation studies to assess model generalizability across the country and imaging conditions, and exploring advanced architectures that combine segmentation and classification in a unique framework.

5 Conclusion

The objective of this study is to create and assess a web-based diagnostic tool that could analyze colposcopy images, classify cervical lesions, and give instant results to practicing clinicians. In order to do that, three technology approaches were implemented and compared: a hybrid U-Net segmentation combined with the SVM classifier, a transfer learning model VGG-19 and the CYENET architecture.

The comparison identified the advantages and drawbacks of each approach. It showed that U-Net combined with the SVM method was effective for lesion segmentation and ROI extraction, whereas it performed weaker to predict benign or malignant classification. The CYENET model was more computationally intensive, but performed well for colposcopic testing specifically. The VGG-19 based transfer learning setting demonstrated the highest accuracy, precision and recall ratio, justifying suitability for deployment in diagnostic scenarios.

In addition to the model comparison, the adoption of the top-performing method in a web application constitutes an important step. This platform is able to process colposcopy images online and preview predictions, helping doctors make a more accurate and standardized diagnostic decision. A tool of this kind has the potential to improve cervical cancer screening and early diagnosis, especially in areas with high-level specialist resources.

However, the research also had some limitations. Original colposcopy images for training are short and of poor quality, so data augmentation was necessary to balance the dataset. Although effective, augmentation cannot simulate the variability present in actual clinical data. Moreover, the model was trained on an average personal computer with constrained processing power, thus limited experimentation with larger architectures and extensive hyperparameter tuning. Finally, no external validation set was used to train and test the models, which would have increased the generalizability of these findings.

Despite these limitations, this study implies that AI can be successfully transferred to colposcopy for cervical cancer screening. Through the integration of sophisticated image analysis with the Web-based interface, it provides an affordable solution to enhance diagnostic capacity and alleviate the burden of cervical cancer. In the future, it will be necessary to expand our dataset more for multicenter validation and to use advanced model architectures, to develop a robust and widely accessible clinical tool.

Bibliography

Arbyn, M., Weiderpass, E., Bruni, L., de Sanjosé, S., Saraiya, M., Ferlay, J. and Bray, F. (2020), 'Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis', *The Lancet Global Health* **8**(2), e191–e203.
URL: *https://doi.org/10.1016/S2214-109X(19)30482-6*

Bosch, F. X., Lorincz, A., Muñoz, N., Meijer, C. J. L. M. and Shah, K. V. (2002), 'The causal relation between human papillomavirus and cervical cancer', *Journal of Clinical Pathology* **55**(4), 244–265.
URL: *https://doi.org/10.1136/jcp.55.4.244*

Bruni, L., Albero, G., Serrano, B., Mena, M., Collado, J. J., Gómez, D., Muñoz, J., Bosch, F. X. and de Sanjosé, S. (2023), Human papillomavirus and related diseases in benin, Technical report, ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Summary Report 10 March 2023.
URL: *https://www.hpvcentre.net/statistics/reports/BEN.pdf*

Chandran, V., Sumithra, M. G., Karthick, A. and Manoharan, S. (2021), 'Cyenet: An ensemble deep learning model for classification of cervical intraepithelial neoplasia from colposcopy images', *BioMed Research International* **2021**, 5584004.
URL: *https://doi.org/10.1155/2021/5584004*

Cross, J. L., Choma, M. A. and Onofrey, J. A. (2024), 'Bias in medical ai: Implications for clinical decision-making', *PLOS Digital Health* **3**(11), e0000651.
URL: *https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000651*

Denny, R. A., Qureshi, M. S. et al. (2022), 'Artificial intelligence in cervical cancer screening: Challenges and future directions', *International Journal of Gynecologic Cancer* **32**(5), 623–630.

Grandini, M., Bagli, E. and Visani, G. (2020), 'Metrics for multi-class classification: an overview'.
URL: *https://arxiv.org/abs/2008.05756*

Hou, X., Shen, G., Zhou, L., Li, Y., Wang, T. and Ma, X. (2022), 'Artificial intelligence in cervical cancer screening and diagnosis', *Frontiers in Oncology* **12**, 851367.
URL: *https://doi.org/10.3389/fonc.2022.851367*

Hu, L., Horning, M. P., Banik, D., Ajenifuja, O. K. et al. (2020), Deep learning-based image evaluation for cervical precancer screening with a smartphone targeting low resource settings, *in*

'IEEE Engineering in Medicine and Biology Society (EMBC)'.
URL: *https://doi.org/10.1109/EMBC44109.2020.9175863*

Jiménez Gaona, A., Alonso-Montes, C., Llorente, P., González, L., Moriyón, O., Cabrera, J. R. and Albaina, E. (2022), 'Automatic classification of colposcopy images combining radiomics and deep learning techniques', *Diagnostics* **12**(7), 1694.
URL: *https://doi.org/10.3390/diagnostics12071694*

Kim, J. J., Campos, N. G., O'Shea, M., Diaz, M. and Mutyaba, I. (2013), 'Model-based impact and cost-effectiveness of cervical cancer prevention in sub-saharan africa', *Vaccine* **31**(Supplement 5), F60–F72.
URL: *https://doi.org/10.1016/j.vaccine.2012.07.093*

Kruczkowski, M., Drabik-Kruczkowska, A., Wesolowski, R., Kloska, A., Pinheiro, M. R., Fernandes, L. and Galan, S. G. (2024), Precise identification of different cervical intraepithelial neoplasia (cin) stages, using biomedical engineering combined with data mining and machine learning, *in* 'Gynecological Cancers: An Interdisciplinary Approach', Springer, pp. 309–334.
URL: *https://link.springer.com/chapter/10.1007/16833*$_2$024$_2$17

Lee, S. H. et al. (2018), 'Deep learning-based detection of high-grade cervical intraepithelial neoplasia', *Journal of Gynecologic Oncology* **29**(2), e21.

Liu, X., Zhang, Y., Liu, S., Li, Y., Chen, J., Ma, X., Sun, M. and Zhao, F. (2024), 'Performance of artificial intelligence for diagnosing cervical intraepithelial neoplasia and cervical cancer: a systematic review and meta-analysis', *eClinicalMedicine* **78**, 102913.

Mantula, F., Toefy, Y. and Sewram, V. (2024), 'Barriers to cervical cancer screening in africa: a systematic review', *BMC Public Health* **24**, 525.
URL: *https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-024-17842-1*

Mathivanan, S. K. and Rajesh, R. (2024), 'Hybrid deep learning and machine learning approach for cervical cancer classification using colposcopy images', *Biomedical Signal Processing and Control* **85**, 105984.
URL: *https://www.nature.com/articles/s41598-024-61063-w.pdf*

McCredie, M. R. E., Sharples, K. J., Paul, C., Baranyai, J., Medley, G., Jones, R. W. and Skegg, D. C. G. (2008), 'Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study', *The Lancet Oncology*

9(5), 425–434. Fast track — Articles.
URL: *https://doi.org/10.1016/S1470-2045(08)70103-7*

Mulongo, M. and Chibwesha, C. J. (2022), 'Prevention of cervical cancer in low-resource african settings', *Obstetrics and Gynecology Clinics of North America* **49**(4), 771–781.
URL: *https://doi.org/10.1016/j.ogc.2022.08.008*

Münger, K., Baldwin, A., Edwards, K. M., Hayakawa, H., Nguyen, C. L., Owens, M., ... and Grace, M. (2004), 'Mechanisms of human papillomavirus-induced oncogenesis', *Journal of Virology* **78**(21), 11451–11460.
URL: *https://doi.org/10.1128/JVI.78.21.11451-11460.2004*

Mvundura, M. and Tsu, V. (2014), 'Estimating the costs of cervical cancer screening in high-burden sub-saharan african countries', *International Journal of Gynecology & Obstetrics* **126**(2), 151–155.
URL: *https://doi.org/10.1016/j.ijgo.2014.02.012*

Peng, X., Chao, T., Qing, L., Yuexiang, L., Yu, S., Yuqian, Z., Jiawei, C., Jianrong, W., Longyu, L., Wei, W., Yucong, L. and Xiaoli, C. (2020), 'Development and validation of an artificial intelligence system for grading colposcopic impressions and guiding biopsies', *BMC Medicine* **18**(1), 406.
URL: *https://doi.org/10.1186/s12916-020-01807-4*

Piras, F., Piga, M., De Montis, A., Zannou, A. R. F., Minerba, L., Perra, M. T., Murtas, D., Atzori, M., Pittau, M., Maxia, C. and Sirigu, P. (2011), 'Prevalence of human papillomavirus infection in women in benin, west africa', *Virology Journal* **8**, 514.
URL: *https://doi.org/10.1186/1743-422X-8-514*

Qin, D., Bai, A., Xue, P., Seery, S., Wang, J., Gonzalez Mendez, M. J., Li, Q., Jiang, Y. and Qiao, Y. (2023), 'Colposcopic accuracy in diagnosing squamous intraepithelial lesions: a systematic review and meta-analysis of the international federation of cervical pathology and colposcopy 2011 terminology', *BMC Cancer* **23**, 187.
URL: *https://bmccancer.biomedcentral.com/articles/10.1186/s12885-023-10648-1*

Ribli, D., Horvath, A., Unger, Z., Pollner, P. and Csabai, I. (2018), 'Computer-aided diagnosis of label-free 3-d optical coherence microscopy images of human cervical tissue', *arXiv preprint* .
URL: *https://doi.org/10.48550/arXiv.1809.10196*

Ronneberger, O., Fischer, P. and Brox, T. (2015), U-net: convolutional networks for biomedical image segmentation, *in* 'Medical Image Computing and Computer-Assisted Intervention (MICCAI)', pp. 234–241.
URL: *https://doi.org/10.1007/978-3-319-24574-4*$_2$8

Saini, S. K., Bansal, V., Kaur, R. and Juneja, M. (2020), 'Colponet for automated cervical cancer screening using colposcopy images', *Machine Vision and Applications* **31**(15).
URL: *https://link.springer.com/article/10.1007/s00138-020-01063-8*

Sarhangi, H. A. and Keyvanpour, M. R. (2023), 'Deep learning applications in cervical cancer diagnosis: A systematic review', *Artificial Intelligence in Medicine* **141**, 102470.
URL: *https://arxiv.org/pdf/2310.16662*

Shinohara, T., Murakami, K. and Matsumura, N. (2023), 'Diagnosis assistance in colposcopy by segmenting acetowhite epithelium using u-net with images before and after acetic acid solution application', *Diagnostics* **13**(9), 1596.
URL: *https://www.mdpi.com/2075-4418/13/9/1596*

Silva, A. F. et al. (2020), 'Cloud-based colposcopic image analysis for early cervical cancer detection', *IEEE Access* **8**, 18771–18780.

Singh, B. (2025), 'Celinsight: Ai-powered web solution for cervical cancer classification', *International Journal of Scientific Research in Engineering and Management (IJSREM)* .
URL: *https://ijsrem.com/download/celinsightai-powered-web-solution-for-cervical-cancer-classification/*

Tjoa, E. and Guan, C. (2020), 'A survey on explainable artificial intelligence (xai): Toward medical xai', *Computers in Biology and Medicine* **122**, 104129.
URL: *https://doi.org/10.1016/j.compbiomed.2020.104129*

Vargas-Cardona, H. D., Vargas-Cardona, M. A., Ruiz-Morales, J., Garc´ia, E., Robledo, S. and Zuluaga, A. F. (2024), 'Artificial intelligence for cervical cancer screening: Scoping review, 2009–2022', *International Journal of Gynecology & Obstetrics* **164**(1), 22–35.

Wang, L., Chen, R., Weng, J., Li, H., Ying, S., Zhang, J., Yu, Z., Peng, C. and Zheng, S. (2024), 'Detecting and localizing cervical lesions in colposcopic images with deep semantic feature mining', *Frontiers in Oncology* **14**.
URL: *https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2024.1423782/full*

World Health Organization (2024), 'Cervical cancer fact sheet', https://www.who.int/news-room/fact-sheets/detail/cervical-cancer. Accessed July 2025.

Xue, P., Ng, M. T. A. and Qiao, Y. (2020), 'The challenges of colposcopy for cervical cancer screening in lmics and solutions by artificial intelligence', *BMC Medicine* **18**(1), 169.
URL: *https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01613-x*

Yuan, J., Wei, L. and Wang, Y. (2020), 'Automated segmentation of cervical lesion regions in colposcopic images using u-net', *Computer Methods and Programs in Biomedicine* **196**, 105635.
URL: *https://doi.org/10.1016/j.cmpb.2020.105635*

Yuan, Y., Zhang, L. and Huang, Y. (2020), 'Automatic segmentation of cervical dysplasia regions using u-net', *Computerized Medical Imaging and Graphics* **79**, 101662.

zur Hausen, H. (2002), 'Papillomaviruses and cancer: from basic studies to clinical application', *Nature Reviews Cancer* **2**(5), 342–350.
URL: *https://doi.org/10.1038/nrc798*