Prevention-Oriented Cervical Cancer Triage Using a Nigerian AI Screening Dataset

Ayebare Vicky*[1], Mudhasi Andrew[2], Melisa Grace Namuhubembe[3], Steven Ziritwawula Matovu[4], Mayinja Bob Williams[1], Confidence Raymond[5], Adaobi Chiazor Emegoakor[5], Patience Bwire[5], Aondona Moses Iorumbur[5]

[1] Mbarara University of Science and Technology, Mbarara, Uganda
[2] Makerere University, Kampala, Uganda
[3] OLK Health Tech Consults, Kampala, Uganda
[4] Ernest Cook University, Kampala, Uganda
[5] SPRINT AI Training for African Medical Imaging Knowledge Translation (SPARK) Program, Montreal, Canada
vvayebare@gmail.com

Abstract. Cervical cancer is the fourth most common cancer among women worldwide, with a disproportionate burden on low- and middle-income countries (LMICs), particularly in Sub-Saharan Africa, where Nigeria records the highest incidence and mortality [1]. Limited access to specialist colposcopists and the subjectivity of manual colposcopy make prevention-oriented, automated triage tools critically needed. Most existing deep learning models largely focus on distinguishing advanced cancer from normal tissue rather than detecting the three precancerous stages, the clinically critical window for intervention[2]. This study investigates two complementary self-supervised and domain-adaptive deep learning approaches for cervical cancer triage from colposcopy images, evaluated on a patient-level dataset of 332 patients from Bowen University, Nigeria. Method 1 employs a ViT-B/16 encoder pretrained using masked autoencoder (MAE) self-supervised learning on an open-source colposcopy dataset spanning the three precancerous stages, and compares it against three baselines (ImageNet ViT-B/16, EndoViT, and EfficientNet-B2) using identical fine-tuning. A preliminary single-split evaluation showed that the domain-adaptive MAE encoder achieved a precancer AUC of approximately 0.75 and a sensitivity of approximately 0.92, compared to approximately 0.68 for ImageNet ViT-B/16 and approximately 0.65 for EndoViT. Method 2 implements a SimCLR contrastive self-supervised pretraining pipeline using an EfficientNet-B3 backbone, leveraging unlabelled images from both datasets to learn cervical visual representations before supervised fine-tuning for binary Normal/Abnormal classification, achieving an accuracy of 0.8, recall of 0.71, and AUC of 0.83 on the held-out test set.

Keywords: cervical cancer, colposcopy, self-supervised learning

1 Introduction

Cervical cancer is a critical global public health challenge, ranking as the fourth most common cancer among women worldwide. In 2022, the World Health Organization (WHO) recorded an estimated 660,000 new cases and 350,000 deaths globally [3]. While organized screening and early treatment have lowered rates in wealthy nations, LMICs face disproportionately high incidence and mortality. This burden is particularly severe in Sub-Saharan Africa, South-East Asia, and Central America due to poverty, high HIV prevalence, limited human Papillomavirus (HPV) vaccination, and limited screening access [4].

Sub-Saharan Africa bears the highest burden of this disease. Within this region, Nigeria ranks first for cervical cancer incidence and mortality. According to the International Agency for Research on Cancer (IARC), Nigeria records approximately 13,676 new cases and 7,093 deaths annually, making it the second most prevalent cancer and cause of female cancer mortality in the country [5]. This crisis is fundamentally driven by a lack of widespread, routine screening to catch the disease early.

Because cervical cancer develops through a long pre-malignant phase, it is highly preventable through HPV vaccination, timely detection, and treatment of precancerous lesions [6]. The WHO recommends HPV DNA testing as the primary screening test, with visual inspection with acetic acid, cytology (Pap smear), or colposcopy used as triage or diagnostic follow-up for women who screen positive [7]. However, Pap smear testing is a manual process that requires several histological examinations and skilled personnel, making it time-consuming and error-prone because it relies on human judgement. In resource-limited settings where standard cytology is manual, visual tools like colposcopy are vital for triage. However, manual colposcopy is highly subjective and depends on a severe shortage of specialist clinicians, which is prone to misdiagnoses and a reduction in diagnostic efficacy from the increasing workload of visual screening [8].

Many models have been developed to automate image analysis, but they focus on identifying advanced cancer rather than prevention [9], [10]. As much as this is important, a prevention-first model is more valuable for public health than one optimized for advanced malignancy. Machine learning algorithms have been proven effective in cases where medical diagnosis requires subjective judgment. With the development of artificial intelligence technology, computer-aided diagnosis (CAD) based on deep learning has made remarkable progress, improving the accuracy and stability of diagnosis while reducing the workload of medical personnel [11]. A series of achievements have been made in the computer-assisted diagnosis of colposcopy. However, in the field of deep learning, related studies mainly focused on the gross classification of cervical cancer as cancerous and normal, based on colposcopy images. There were relatively few studies on detecting precancerous cases at any stage of progression, which could provide intuitive guidance for colposcopists.

This study addresses this gap through two complementary deep learning approaches. The first is a domain-adaptive masked autoencoder (MAE) pretraining framework in which a ViT-B/16 encoder is pretrained with MAE on an open-source colposcopy dataset spanning the three precancerous stages (the Intel dataset), then fine-tuned on a

local Nigerian screening dataset from Bowen University for staged precancer classification. This approach is benchmarked against three encoder baselines, ImageNet ViT-B/16, EndoViT, and EfficientNet-B2 using identical fine-tuning. The second is a SimCLR-based self-supervised learning (SSL) pipeline using EfficientNet-B3 as the backbone encoder, which leverages both labelled and unlabelled images from both datasets to learn domain-relevant cervical feature representations, then fine-tunes on the binary Normal/Abnormal classification task using the Bowen dataset. The two methods are evaluated independently on different classification formulations, staged multi-class versus binary, and are not directly compared against one another on shared metrics; rather, each provides its own reference point for domain-adaptive SSL pretraining on colposcopy imagery. Together, the two methods explore self-supervised pretraining strategies from complementary perspectives: contrastive learning (SimCLR) versus masked reconstruction (MAE), and binary triage versus staged precancer classification. The main contributions of this work are: (1) the first evaluation, to our knowledge, of domain-adaptive MAE pretraining for staged precancer classification on colposcopy imagery; (2) an independent SimCLR-based binary triage pipeline evaluated on the same data, offering a lighter-weight reference point for low-resource deployment.

2 Related Works

Several recent studies have applied deep learning to colposcopy-based cervical cancer classification, however, most frame the task as a binary distinction between cancer and normal tissue rather than the staged detection of precancerous lesions. Various convolutional architectures, including ResNet-50, CervixNET, VGG19, and CYENET, have demonstrated high performance in classifying cancer presence [12]. Despite these successes, the focus remains on distinguishing cancer types rather than explicitly modeling precancerous stages, often collapsing those lesions into a broad "abnormal" class. Complementary lesion-focused methods, such as CLDNet, and segmentation-oriented work utilizing EfficientNet-based networks, have improved lesion localization and delineation [13]. Nevertheless, these supervised methods predominantly rely on large quantities of expert-annotated labels and struggle to distinguish between precancerous grades. Self-supervised and semi-supervised learning approaches offer a promising alternative for low-resource medical imaging scenarios. Masked autoencoders (MAE) enable vision transformers to learn rich visual representations from unlabeled images by reconstructing randomly masked patches [14]. While MAE has shown strong performance on natural image benchmarks, its potential for domain-adaptive pretraining in colposcopy remains underexplored. Research like EndoVit demonstrated that pretraining a ViT encoder on large corpora of surgical endoscopy images improves downstream performance, establishing a strong precedent for domain-adaptive pretraining in gastrointestinal and surgical imaging [15]. This study extends this line of inquiry to colposcopy, where domain-adaptive MAE pretraining on cervical images has not been previously evaluated for staged precancer detection.

3 Methodology

This study developed and compared two complementary self-supervised learning (SSL) approaches aimed at high-sensitivity cervical precancer screening. Method 1 evaluated domain-adaptive Masked Autoencoder (MAE) pretraining across four encoders, utilizing identical fine-tuning on the Bowen dataset for a four-class patient-level precancer classification (Normal versus three precancerous stages). Method 2 implemented a SimCLR-based SSL pipeline with an EfficientNet-B3 backbone for binary (Normal versus Abnormal) classification, leveraging both labeled and unlabeled images from two datasets for pretraining before supervised fine-tuning on the Bowen dataset. Two colposcopy image datasets were utilized. The Intel open-source dataset which contained 8727 images showing all the precancer stages, was used exclusively for MAE pretraining. The primary dataset from Bowen University, Nigeria (the Bowen dataset), consisted of patient-level colposcopy folders structured into Normal and Abnormal directories with clinical annotations, each containing approximately 5 to 8 images. Normal contained images of 210 patients while Abnormal contained 122 patients, bringing the total to 332 patients and 3356 patient images.

3.1 Method 1

The proposed encoder uses a ViT-B/16 backbone (16 × 16 patch size, base model size) pretrained with masked autoencoder (MAE) self-supervised learning on the Intel colposcopy dataset. Masked autoencoders, enable vision transformers to learn rich visual representations from unlabeled images by reconstructing randomly masked image patches. In MAE pretraining, a random proportion of image patches is masked and the model is trained to reconstruct the masked patches from the unmasked context. This forces the encoder to develop a rich internal representation of the cervical tissue appearance without requiring any labelled data. Pretraining was conducted on the full Intel colposcopy dataset. Figure 1, shows the MAE pretraining reconstruction loss on the Intel ODT colposcopy dataset. The experimental design for this model holds a fine-tuning constant and varies only the pretraining strategy across four encoders. This isolates the contribution of pretraining domain and method to downstream precancer classification performance, controlling for architecture and fine-tuning procedure. For MAE pretraining on the Intel dataset, each image was subjected to random resized cropping, horizontal flipping, colour jittering, tensor conversion, and normalization, consistent with standard MAE augmentation pipelines. Two independently augmented views of each image were generated for the contrastive component of pretraining. For supervised fine-tuning on the Bowen dataset, training images were resized to 224 × 224 pixels and augmented using random horizontal and vertical flips, small rotations, and colour jittering before normalization. Additional regularization during fine-tuning was applied via Mixup augmentation. Validation and test images were resized and normalized only, without augmentation, to ensure unbiased performance evaluation. The effects of the reconstruction can be seen in Figure 2.

3.2 Method 2

Method 2 employs a SimCLR contrastive self-supervised learning framework with an ImageNet-initialized EfficientNet-B3 backbone, leveraging all labeled and unlabeled images from both the Intel and Bowen datasets. During the 40-epoch pretraining phase, images undergo dual augmentations (random cropping, horizontal flipping, color jittering), and a two-layer MLP projection head uses NT-Xent loss to maximize positive pair agreement alongside an Adam optimizer and cosine annealing schedule. Notably, the pretraining pool included images from the Bowen test partition; although its unsupervised nature prevents label leakage, future iterations must utilize a fully patient-disjoint split to prevent test-set statistics from influencing learned representations. Supervised fine-tuning for binary classification uses a 70/15/15 patient-level split of the Bowen dataset via a two-phase strategy: first, a new MLP head is trained for 20 epochs using cross-entropy loss while the encoder is frozen, followed by 100 epochs of end-to-end unfreezing using a ReduceLROnPlateau scheduler and a lower learning rate for the encoder. Finally, out-of-distribution generalization across precancerous severity grades is evaluated by testing the final binary classifier on the Intel dataset.

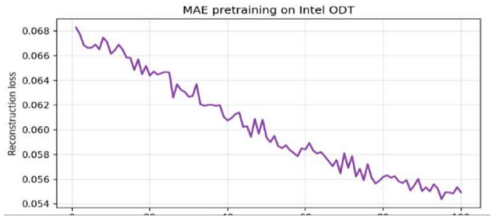

Fig. 1. MAE pretraining reconstruction loss on the Intel ODT colposcopy dataset over 100 epochs. Steady monotonic decrease confirms stable self-supervised learning of cervical tissue representations

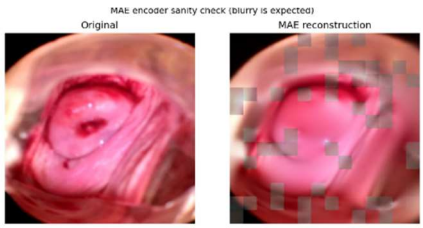

Fig. 2. MAE encoder sanity check: original colposcopy image (left) and the blurry MAE reconstruction from unmasked patches (right). The model successfully recovers cervical tissue structure and colour zones from 75% masked input, indicating meaningful

4 Results

4.1 Method 1

A single-split evaluation was conducted to guide encoder selection.

Table 1. Preliminary results on initial single split.

Encoder	Precancer AUC	Precancer Sensitivity
MAE	0.752	0.917
ImageNet ViT-B/16	0.684	0.833
EndoViT	0.655	0.250

Preliminary results indicate that domain-adaptive Masked Autoencoder (MAE) pretraining on in-domain colposcopy images outperforms natural image (ImageNet) and adjacent-domain (EndoViT) pretraining. The MAE encoder achieved the highest AUC (0.752) and demonstrated superior sensitivity, which is critical for minimizing false negatives in clinical screening. While all three models exhibited overfitting due to the small size of the Bowen dataset, the MAE encoder maintained a higher validation AUC throughout fine-tuning, confirming the robustness of domain-adaptive initialization.
Error analysis revealed that the MAE encoder effectively reduced the most clinically significant errors, misclassifying Precancerous cases as Normal, compared to the other models. However, all models struggled with ASCUS (Atypical Squamous Cells of Undetermined Significance) classification, highlighting the visual ambiguity of this category. Furthermore, UMAP dimensionality reduction of the feature embeddings showed no clear class-level separation, suggesting that the challenges in classification are inherent to the visual similarity of the tissue types. Interestingly, the data revealed two distinct clusters likely driven by imaging or acquisition variations rather than pathology, an area for future investigation.

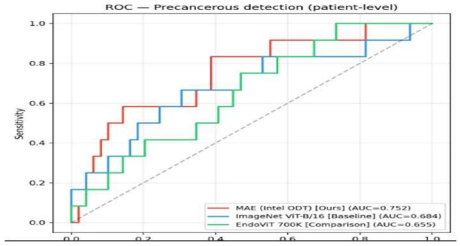

Fig. 3. Receiver operating characteristic (ROC) curves for precancerous detection at the patient level (single-split evaluation).

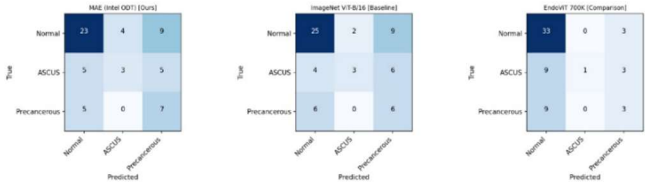

Fig. 4. Confusion matrices (single-split evaluation) for MAE (Intel ODT) [Ours] (left), ImageNet ViT-B/16 [Baseline] (centre), and EndoViT 700K [Comparison] (right)

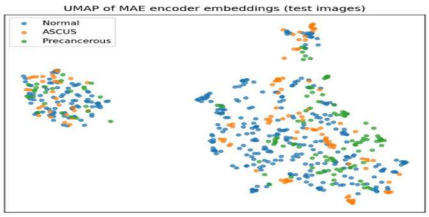

Fig. 5. UMAP projection of MAE encoder image embeddings from the test set. Normal (blue), ASCUS (orange), and Precancerous (green) points are interleaved within clusters, reflecting the genuine visual similarity between classes under colposcopy.

4.2 Method 2

Table 2 reports the performance of the SimCLR + EfficientNet-B3 binary classifier on the held-out 15% test partition of the Bowen dataset. The model correctly classified 80% of test cases, with a recall of 71% for the Abnormal class and an AUC of 0.83, indicating reasonable discriminative ability for a binary triage task on a limited dataset. The recall of 71% suggests that approximately 29% of abnormal cases were classified as Normal in the test set, which is a clinically relevant limitation for a screening application where false negatives carry significant consequences. When the fine-tuned binary classifier was applied to images from the Intel dataset, the majority of images were classified as Abnormal, consistent with the expectation that the model generalizes meaningfully to unseen precancerous severity grades beyond those present in the binary fine-tuning data. Figure 6 shows the visual representation of the predictions made by the model.

Table 2. Performance of the SimCLR + EfficientNet-B3 binary classifier on the held-out on the test data.

Metric	Value (Test Set)
Accuracy	80%
Recall (Sensitivity for Abnormal)	71%
AUC	83%

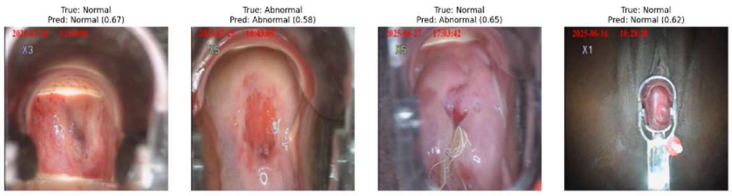

Fig. 6. Representation of model predictions. Three out of four predictions were done accurately showing good model performance

5 Discussion

This study evaluated two complementary self-supervised learning (SSL) approaches for automated cervical cancer triage using colposcopy images. Method 1 (MAE) focused on staged precancer classification, demonstrating that in-domain colposcopy pretraining significantly outperforms standard ImageNet and gastrointestinal (EndoViT) pretraining. As an immediate next step ahead of clinical translation, Method 1's initial single-split evaluation will be extended to full 5-fold patient-level stratified cross-validation with bootstrap confidence intervals to yield tighter effect-size estimates and a statistically robust comparison across all four encoders. Method 2 (SimCLR) targeted binary Normal/Abnormal triage, leveraging unlabeled data to mitigate small dataset constraints (0.83 AUC), though its 0.71 recall and binary framing limit its clinical utility compared to Method 1. Both methods face shared limitations, including high computational demands, unassessed equipment generalizability, and reliance on a small, single-institution dataset. To address these, future work will implement a fully patient-disjoint pretraining/fine-tuning separation for Method 2, evaluate standard medical-imaging CNN baselines (ResNet-50, DenseNet-121) alongside domain-pretraining models, and conduct independent external validation on larger, more diverse datasets to improve generalization and sensitivity across varying clinical protocols.

6 Impact in Resource Constrained Settings

Addressing the critical cervical cancer screening gap in resource-constrained settings like Nigeria, this study introduces an automated, self-supervised triage paradigm. By detecting three distinct precancerous stages rather than just advanced malignancy, the model enables timely clinical intervention. Leveraging unlabeled local data, our approach offers a scalable blueprint for democratizing cervical screening across Sub-Saharan Africa and reducing reliance on scarce oncology infrastructure. However, the reliance on high-performance computing which is essential for the pipeline's robustness, presents a potential barrier to direct replication in environments lacking equivalent computational resources.

Acknowledgments: This research was supported by the Lacuna Fund for Health and Equity (PI: Udunna Anazodo, 0508-S-001), the National Science and Engineering Research Council of Canada (NSERC) Discovery Launch Supplement (PI: Udunna Anazodo, DGECR-2022-00136), the Digital Research Alliance of Canada (the Alliance), the McGill University Doctoral Internship Program, and the University of Washington Azure GenAI for Science Hub. The authors also wish to extend their sincere gratitude to the SPARK Academy 2026 Instructors for their invaluable contributions, guidance, and mentorship throughout the duration of this program.
Disclosure of Interests: The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] "Control of Cancer of the Cervix in Low- and Middle-Income Countries | Annals of Surgical Oncology | Springer Nature Link." Accessed: Jun. 27, 2026. [Online]. Available: https://link.springer.com/article/10.1245/s10434-014-4344-8

[2] N. Youneszade, M. Marjani, and C. P. Pei, "Deep Learning in Cervical Cancer Diagnosis: Architecture, Opportunities, and Open Research Challenges," *IEEE Access*, vol. 11, pp. 6133–6149, 2023, doi: 10.1109/ACCESS.2023.3235833.

[3] J. Wu *et al.*, "Global burden of cervical cancer: current estimates, temporal trend and future projections based on the GLOBOCAN 2022," *J. Natl. Cancer Cent.*, vol. 5, no. 3, pp. 322–329, Jun. 2025, doi: 10.1016/j.jncc.2024.11.006.

[4] V. V. Shulhina and P. Y. Tokar, "Global disparities and trends in cervical cancer morbidity and mortality," *Ukr. J. «Health Woman»*, no. 1(182), pp. 86–90, Feb. 2026, doi: 10.15574/HW.2026.1(182).8690.

[5] "Use of Appropriate Technology to Improve Cervical Cancer Prevention and Early Detection: Experience in Jos, Nigeria | JCO Global Oncology." Accessed: Jun. 27, 2026. [Online]. Available: https://ascopubs.org/doi/10.1200/JGO.19.16000

[6] "Preventing Cervical Cancer Globally: Are We Making Progress? | Cancer Prevention Research | American Association for Cancer Research." Accessed: Jun. 27, 2026. [Online]. Available: https://aacrjournals.org/cancerpreventionresearch/article-abstract/14/12/1055/675171/Preventing-Cervical-Cancer-Globally-Are-We-Making?redirectedFrom=fulltext

[7] "Evidence based appropriate triage strategies for implementing high risk HPV as primary technology in cervical cancer screening - Minerva Ginecologica 2020 April;72(2):96-105." Accessed: Aug. 25, 2026. [Online]. Available: https://www.minervamedica.it/en/journals/minerva-obstetrics-gynecology/article.php?cod=R09Y2020N02A0096

[8] P. Vassilakos *et al.*, "Telecytologic diagnosis of cervical smears for triage of self-sampled human papillomavirus–positive women in a resource-limited setting: concept development before implementation," *J. Am. Soc. Cytopathol.*, vol. 12, no. 3, pp. 170–180, May 2023, doi: 10.1016/j.jasc.2023.02.001.

[9] Y. Kumar *et al.*, "Automating cancer diagnosis using advanced deep learning techniques for multi-cancer image classification," *Sci. Rep.*, vol. 14, no. 1, p. 25006, Oct. 2024, doi: 10.1038/s41598-024-75876-2.

[10] "Developing global image feature analysis models to predict cancer risk and prognosis | Visual Computing for Industry, Biomedicine, and Art | Springer Nature Link." Accessed: Jun. 29, 2026. [Online]. Available: https://link.springer.com/article/10.1186/s42492-019-0026-5

[11] "Computer-aided diagnosis in the era of deep learning - Chan - 2020 - Medical Physics - Wiley Online Library." Accessed: Jun. 27, 2026. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.13764

[12] "Cervical Cancer Classification Using Deep Learning Approach Using Colposcopy Images | Neural Processing Letters | Springer Nature Link." Accessed: Jun. 27, 2026. [Online]. Available: https://link.springer.com/article/10.1007/s11063-025-11770-w

[13] T. Kalimuthu, "Recent Advances in Segmentation and Classification of Renal Tumors Using EfficientNet-Based U-Net and Epistemic Neural Networks: A Systematic Review," *Int. J. Adv. Electr. Electron. Eng.*, vol. 14, no. 2, pp. 101–106, Nov. 2025, doi: 10.65521/ijaeee.v14i2.1985.

[14] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked Autoencoders Are Scalable Vision Learners," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 15979–15988. doi: 10.1109/CVPR52688.2022.01553.

[15] D. Batić, F. Holm, E. Özsoy, T. Czempiel, and N. Navab, "EndoViT: pretraining vision transformers on a large collection of endoscopic images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 19, no. 6, pp. 1085–1091, Jun. 2024, doi: 10.1007/s11548-024-03091-5.