



Article

Early Cervical Cancer Diagnosis with SWIN-Transformer and Convolutional Neural Networks

Foziya Ahmed Mohammed ^{1,2,3}, Kula Kekeba Tune ^{1,2} , Juhar Ahmed Mohammed ⁴, Tizazu Alemu Wassu ⁵ and Seid Muhie ^{6,7,*} 

- ¹ Department of Software Engineering, College of Engineering, Addis Ababa Science and Technology University, Addis Ababa 16417, Ethiopia; foziya.ahmed@aastu.edu.et (F.A.M.); kula.kekeba@aastu.edu.et (K.K.T.)
 - ² Center of Excellence for HPC and Big Data Analytics, Addis Ababa Science and Technology University, Addis Ababa 16417, Ethiopia
 - ³ Department of Information Technology, College of Computing and Health Informatics, Wolkite University, Wolkite P.O. Box 07, Ethiopia
 - ⁴ School of Pharmacy, College of Medicine and Health Science, Hawassa University, Hawassa P.O. Box 05, Ethiopia; juhar2161@gmail.com
 - ⁵ Addis Hiwot General Hospital, Addis Ababa P.O. Box 170023, Ethiopia; tizazu17@yahoo.com
 - ⁶ Enkoy LLC, Lanham, MD 20706, USA
 - ⁷ The Geneva Foundation, Silver Spring, MD 20910, USA
- * Correspondence: smuhie@genevausa.org; Tel.: +1-301-412-6171

Abstract: **Introduction:** Early diagnosis of cervical cancer at the precancerous stage is critical for effective treatment and improved patient outcomes. **Objective:** This study aims to explore the use of SWIN Transformer and Convolutional Neural Network (CNN) hybrid models combined with transfer learning to classify precancerous colposcopy images. **Methods:** Out of 913 images from 200 cases obtained from the Colposcopy Image Bank of the International Agency for Research on Cancer, 898 met quality standards and were classified as normal, precancerous, or cancerous based on colposcopy and histopathological findings. The cases corresponding to the 360 precancerous images, along with an equal number of normal cases, were divided into a 70/30 train–test split. The SWIN Transformer and CNN hybrid model combines the advantages of local feature extraction by CNNs with the global context modeling by SWIN Transformers, resulting in superior classification performance and a more automated process. The hybrid model approach involves enhancing image quality through preprocessing, extracting local features with CNNs, capturing the global context with the SWIN Transformer, integrating these features for classification, and refining the training process by tuning hyperparameters. **Results:** The trained model achieved the following classification performances on fivefold cross-validation data: a 94% Area Under the Curve (AUC), an 88% F1 score, and 87% accuracy. On two completely independent test sets, which were never seen by the model during training, the model achieved an 80% AUC, a 75% F1 score, and 75% accuracy on the first test set (precancerous vs. normal) and an 82% AUC, a 78% F1 score, and 75% accuracy on the second test set (cancer vs. normal). **Conclusions:** These high-performance metrics demonstrate the models' effectiveness in distinguishing precancerous from normal colposcopy images, even with modest datasets, limited data augmentation, and the smaller effect size of precancerous images compared to malignant lesions. The findings suggest that these techniques can significantly aid in the early detection of cervical cancer at the precancerous stage.



Citation: Mohammed, F.A.; Tune, K.K.; Mohammed, J.A.; Wassu, T.A.; Muhie, S. Early Cervical Cancer Diagnosis with SWIN-Transformer and Convolutional Neural Networks. *Diagnostics* **2024**, *14*, 2286. <https://doi.org/10.3390/diagnostics14202286>

Academic Editors: Edward J. Pavlik, Ki Hoon Ahn and Kwang-Sig Lee

Received: 15 August 2024

Revised: 5 October 2024

Accepted: 12 October 2024

Published: 14 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: cervical cancer; early diagnosis; precancerous lesions; SWIN Transformer; convolutional neural networks (CNN); colposcopy images; transfer learning; hybrid models; medical image classification; cancer screening; histopathology

1. Introduction

Cervical cancer remains a major global health issue, ranking as the fourth most common cancer among women worldwide [1]. It is primarily caused by persistent infection with high-risk human papillomavirus (HPV) types. According to the World Health Organization (WHO), there were approximately 660,000 new cases and 350,000 deaths due to cervical cancer in 2022 [1]. The burden of cervical cancer is especially high in low- and middle-income countries, where 94% of the deaths occur, reflecting disparities in access to HPV vaccination, cervical screening, and treatment services [2]. The considerable geographic variation in cervical cancer burden underlines the urgent need for accessible early diagnosis to enable timely and effective interventions [3,4].

Early diagnosis of cervical cancer is essential for improving treatment outcomes and survival rates. Detecting the disease at the precancerous stage allows for less invasive and more effective treatment options, reducing both morbidity and mortality associated with advanced stages [5,6]. Moreover, early diagnosis facilitates timely interventions, such as administering HPV vaccines, which have been shown to significantly reduce cervical cancer incidence [7,8]. The benefits of early detection extend beyond patient outcomes, enabling better resource allocation and a more efficient use of healthcare services [9]. Effective follow-up and management strategies are also important, as they prevent cervical cancer progression [10], improve prognosis, and significantly reduce treatment costs [11–13].

Despite its importance, early diagnosis of cervical cancer remains challenging, even in regions with established screening programs [14]. Traditional screening methods like Pap smears and HPV testing have limitations that affect their effectiveness. Pap smears, though widely used, are subject to variability in interpretation and often have lower sensitivity for detecting precancerous lesions [15]. While HPV testing is more sensitive, it can result in high false-positive rates, leading to unnecessary follow-up procedures and psychological distress [16]. These limitations highlight the need for more accurate, reliable, and accessible diagnostic tools to improve early detection and patient outcomes [17].

Recent advancements in deep learning, particularly with models like convolutional neural networks (CNNs) [18] and transformer-based architectures such as the SWIN Transformer [19], have revolutionized medical imaging and significantly improved early cancer detection [20,21]. These technologies can analyze complex patterns in medical images with remarkable accuracy and consistency, often surpassing traditional methods by detecting subtle differences in tissue structures that might be missed by human observers [22–25].

In this context, a colposcopy image classification using SWIN Transformer and CNN hybrid models is particularly important for improving early cervical cancer detection and treatment, as this addresses the variability and potential misdiagnosis associated with the slow and costly process of manual image interpretation. The SWIN Transformer and CNN hybrid model for a precancerous colposcopy classification combines the strengths of both architectures to deliver more accurate and interpretable results. The CNN component excels at extracting fine-grained, local features such as edges, textures, and shapes from colposcopy images, while the SWIN Transformer's mechanism of shifted window attention captures both local and global contextual information, modeling long-range dependencies across heterogeneous images. By leveraging both local feature extraction and the global context, the hybrid model generalizes better across diverse datasets, improving classification accuracy despite variability in image quality and patient factors [26]. These complementary features are then integrated and passed through a classification layer to accurately categorize the images. This dual capability is critical for distinguishing subtle differences in tissues, making it particularly effective for classifying precancerous lesions. Integration of the outputs of the hybrid components reduces false positives and negatives, leading to earlier detection and better patient outcomes, a significant advancement in cervical cancer screening [27,28].

The hybrid model's performance is further enhanced through the use of transfer learning and data augmentation, allowing it to generalize well despite limited data availability [21]. This approach is especially relevant for cervical cancer, where variations in

image quality and patient demographics can affect diagnostic outcomes [29]. Advanced techniques like contrastive learning and multi-task learning help the model focus on precancerous lesions while minimizing the impact of confounding factors [30,31]. These hybrid models offer a robust and efficient solution for early diagnosis and treatment planning in cervical cancer prevention especially in resource-limited settings, where cervical cancer is prevalent, offering reliable diagnostic support and enhancing global screening efforts [1].

However, to fully realize these advancements, certain challenges must be addressed. The need for large, annotated datasets to train deep learning models remains a significant barrier, particularly in low-resource settings [29]. Integrating these models into clinical practice also requires rigorous validation and regulatory approval to ensure their safety and effectiveness. Overcoming these obstacles is crucial for implementing AI-driven diagnostic tools in healthcare, ensuring they provide reliable and actionable insights [14]. Additionally, ethical considerations regarding data privacy and algorithmic transparency must be thoroughly addressed to gain public trust and acceptance, which is essential for the widespread adoption of these technologies [12].

The present study leverages transfer learning with SWIN Transformer and CNN hybrid models to overcome the limitations of existing diagnostic methods, further enhancing the accuracy of detecting precancerous lesions. This approach not only promises significant advancements in early cervical cancer detection, potentially leading to better patient outcomes, but also reduces the strain on healthcare systems by minimizing unnecessary follow-up procedures and treatments resulting from false positives [21]. Moreover, this approach's broader applicability extends to the early detection of other cancers and diseases [32], in addition to the automatic classification of colposcopy images.

Automatic classification in colposcopy plays a key role in enhancing diagnostic accuracy and efficiency. It assists colposcopists by analyzing images in real-time to accurately identify CIN and other precancerous conditions, reducing the risk of misclassifications and ensuring timely treatments. This reduces the likelihood of misdiagnosis and unnecessary interventions due to overdiagnosis [1,30,31]. Furthermore, by providing a standardized, objective analysis, an automatic classification addresses the variability inherent in manual image annotation, ensuring consistent and reliable interpretations through advanced machine learning models like SWIN Transformers and CNNs [19,27]. The hybrid model's ability to capture both local and global features in colposcopy images is key to achieving accurate tissue classifications and reliable diagnoses across diverse datasets and clinical settings [19,26,27,30].

Overall, cervical cancer remains a significant global health challenge, particularly in regions with limited access to healthcare services. Early diagnosis is essential for improving patient outcomes and reducing mortality. Advanced deep learning and machine learning techniques offer promising solutions to enhance the accuracy and reliability of early diagnostic methods. This study aims to contribute to these efforts by developing an accessible and reliable diagnostic system, with the potential to improve public health outcomes and the quality of life for women worldwide [33]. The findings of this study may also inform policy decisions regarding cervical cancer screening and prevention programs [34].

1.1. Related Studies

The application of SWIN Transformer and CNN hybrid models has revolutionized medical image analyses by offering superior performance in segmentation and classification tasks. The hybrid model combines CNN's local feature extraction capability with the SWIN Transformer's ability to capture long-range dependencies, making it highly effective for complex medical images like histopathology and radiology scans [19,35]. This architecture has shown remarkable success in lung cancer detection, leveraging CNN for the feature extraction of local tissues and the SWIN Transformer for capturing global contextual information, significantly improving diagnostic accuracy [28,36]. The hybrid model's ability to fuse these components enhances both the model's interpretability and its diagnostic power in multimodal datasets [37].

In brain tumor segmentation, the SWIN Transformer and CNN hybrid model was applied to accurately segment tumor boundaries in MRI scans, significantly outperforming traditional CNN-based approaches due to its advanced attention mechanism, which captures the spatial relationships between different regions of the brain [38]. By using the SWIN Transformer, this model efficiently handles the complex structure of brain tissues, leading to improved segmentation accuracy and generalizability across different datasets [39]. Furthermore, its ability to manage high-dimensional medical images without compromising on computational efficiency sets it apart from other architectures [40].

In colorectal cancer screening, the hybrid model demonstrated enhanced polyp detection by integrating CNN's capacity for fine-grained feature extraction with the SWIN Transformer's capabilities of capturing the global context. This resulted in a more accurate identification of polyps in colonoscopy images, improving early detection rates [41]. Additionally, the use of attention mechanisms enabled the model to focus on critical areas of the images, reducing false positives and improving overall diagnostic accuracy [28].

For the diagnosis of Alzheimer's disease, the SWIN Transformer and CNN hybrid model has been employed in the analysis of PET and MRI scans, effectively fusing structural and functional imaging data. This combination allows for a more comprehensive analysis, improving early detection and treatment planning for Alzheimer's disease. The SWIN Transformer's attention mechanism enhances the model's ability to capture subtle changes in brain structure, which is crucial for early-stage diagnosis [37,42]. The model's performance was validated across large-scale datasets, showing superior accuracy in classifying different stages of Alzheimer's compared to traditional CNN-based methods [37].

Moreover, the SWIN Transformer and CNN hybrid model has been applied to cardiac and liver image segmentation tasks, where it achieved state-of-the-art results. The model's ability to learn both local and global dependencies in CT and MRI scans allowed for a more precise segmentation of organs and better detection of pathological features, which is essential in clinical decision making [38,43]. Its application in these tasks underscores its versatility and robustness in handling various medical imaging challenges [35,40].

1.2. Organization of This Paper

This paper first introduces the importance of early cervical cancer detection and the limitations of traditional screening methods. It then details the methodology, focusing on the use of SWIN Transformer and CNN hybrid models for colposcopy image classification. This is followed by results that demonstrate the model's performance and a discussion of the clinical implications, challenges, and future directions for integrating these advanced machine learning techniques into healthcare practice.

2. Materials and Methods

2.1. Colposcopy Images

The colposcopy image datasets and metadata, comprising 913 images from 200 cases, were obtained from the Colposcopy Image Bank of the International Agency for Research on Cancer (IARC) [44]. The data and metadata were accessed on 16 April 2024. Of the 200 cases, three cases consisting of a total 15 images were excluded due to inconclusive colposcopy and pathology assessments. The remaining 898 colposcopy images from the 197 cases were grouped into normal ($n = 94$ cases), precancerous ($n = 77$ cases), and cancer ($n = 26$ cases) sets based on the colposcopy examination and histopathology findings (Figures 1a,b and 2).

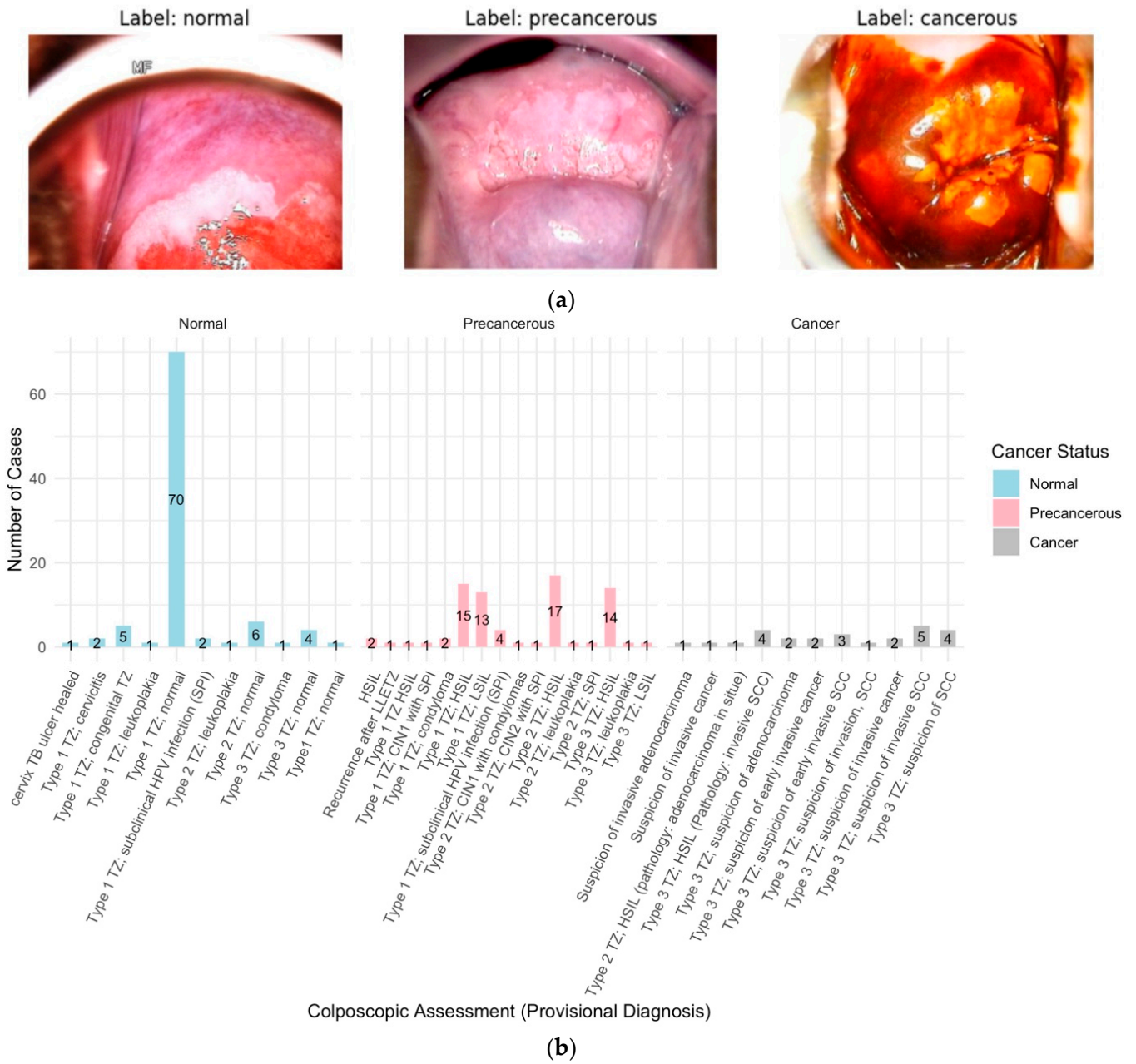


Figure 1. (a) Colposcopy images from each of the normal, precancerous, and cancer groups. (b) Provisional diagnosis vs. cancer status: The provisional diagnosis during colposcopy, which relies on the clinician’s initial visual and clinical assessment, is important for categorizing cervical lesions as normal, precancerous, or cancerous. An analysis of the provided dataset highlights specific provisional diagnoses associated with each final cancer status. For cases ultimately confirmed as normal, the most frequent provisional diagnosis was “Type 1 Transition Zone (TZ); normal.” In precancerous cases, “Types 1, 2, and 3 TZ; HSIL” and “Type 1 TZ; LSIL” were commonly noted. For cancer cases, “Type 3 TZ; suspicion of invasive squamous cell carcinoma” was the predominant provisional diagnosis.

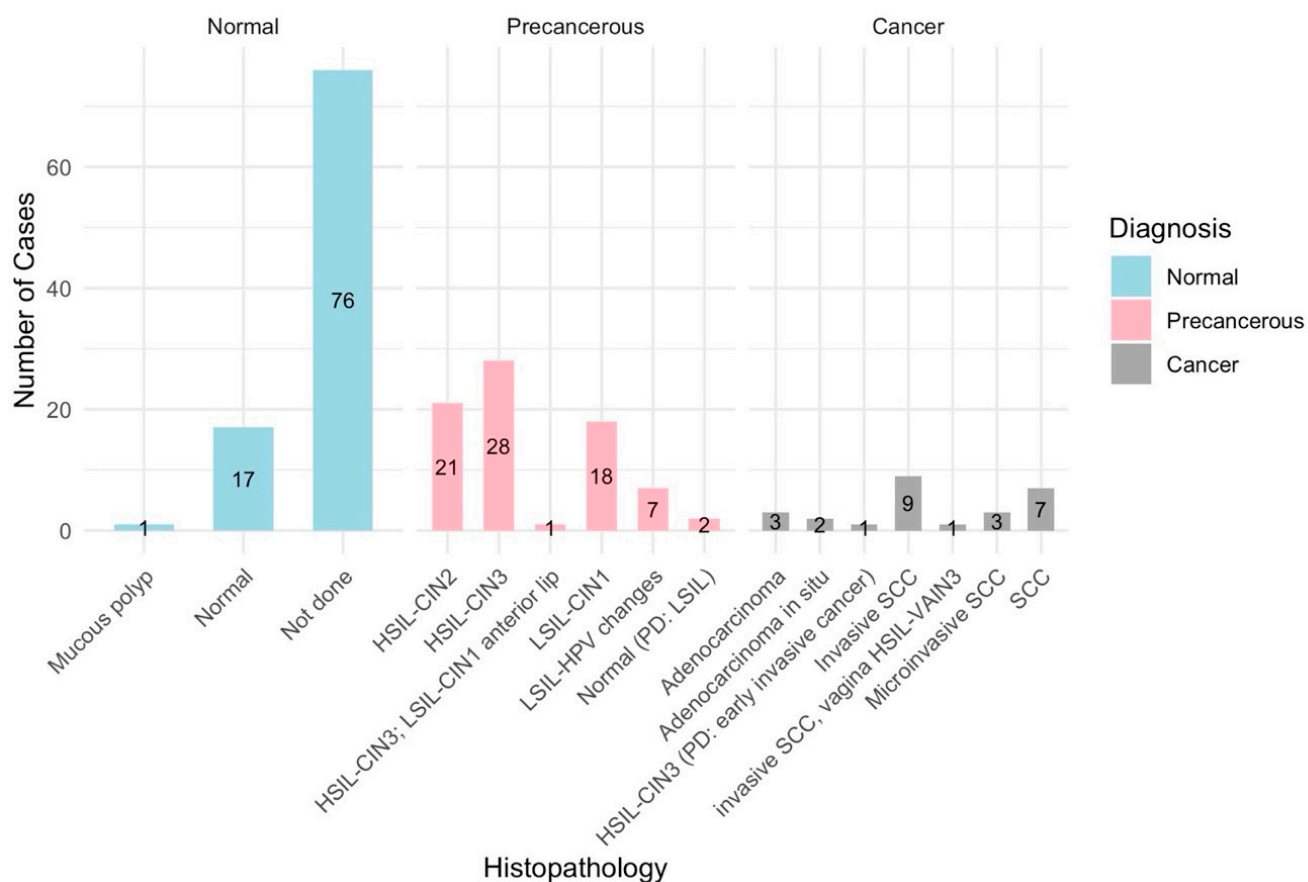


Figure 2. Histopathology vs. cancer diagnosis: A histopathological analysis was used to determine the final diagnosis of cervical lesions identified during colposcopy. In normal cases, histopathology was often not performed, suggesting that the colposcopy assessment alone was sufficient. When histopathology was performed, findings such as CIN1 or the absence of dysplasia supported the normal diagnosis. Precancerous cases were characterized by moderate to severe dysplasia (CIN2, CIN3), low-grade squamous intraepithelial lesions (LSILs), and high-grade squamous intraepithelial lesions (HSILs), indicating varying degrees of abnormality with potential progression to cancer. Cancer cases were confirmed by histopathological evidence of invasive adenocarcinoma, squamous cell carcinoma, or adenocarcinoma in situ. These findings highlight the role of histopathology in accurately diagnosing and categorizing cervical lesions, guiding appropriate patient management and treatment strategies.

2.2. Clinical Criteria for Case Grouping

Cases were meticulously categorized into normal, precancerous, and cancerous based on colposcopy assessments and histopathology findings (Figures 1b and 2), with HPV positivity, the Transformation Zone, and the SWEDE Score as corroborative metrics.

I. Normal Group:

- Cases: a total of 94 cases were characterized by normal colposcopy and histopathology findings.
- Key Features:
 - Mostly HPV-negative cases, with some HPV-positive cases.
 - Mainly Type 1 Transformation Zones, with some Type 2 and 3.
 - Adequate colposcopy samples showing original squamous epithelium, columnar epithelium, and metaplastic squamous epithelium.
 - Brown, faintly or patchy yellow, or unknown iodine staining.

- Low SWEDE scores and normal histopathology findings, such as mature squamous epithelium and nabothian cysts.
- II. Precancerous Group:
 - Cases: a total of 77 cases with colposcopy and histopathologic findings indicative of Low-grade Squamous Intraepithelial Lesions (LSILs), including CIN1, and High-grade Squamous Intraepithelial Lesions (HSILs), including CIN2 and CIN3.
 - LSILs (CIN1):
 - HPV-positive and -negative cases.
 - Type 1 or Type 2 Transformation Zone.
 - Thin acetowhite epithelium, irregular borders, fine mosaic, and fine punctation on colposcopy.
 - Moderate SWEDE scores and faintly or patchy yellow iodine uptake.
 - HSILs (CIN2 and CIN3):
 - Higher severity precancerous lesions.
 - Dense acetowhite epithelium, sharp borders, ridge sign, inner border sign, and atypical vessels on colposcopy.
 - Medium SWEDE scores and distinct yellow iodine uptake.
 - HSIL cases required medical intervention to prevent progression to invasive cancer.
- III. Cancer Group:
 - Cases: a total of 26 cases with invasive squamous cell carcinoma, adenocarcinoma, and adenocarcinoma in situ.
 - Key Features:
 - All HPV-positive cases.
 - Type 3 Transformation Zone.
 - Colposcopy findings include dense acetowhite epithelium, coarse punctation or mosaic, sharp borders, and features suspicious for invasion like irregular surfaces, erosion, or gross neoplasm.
 - High SWEDE scores and distinct yellow or non-staining iodine uptake.
 - Histopathology confirmed invasive carcinoma, with the squamo-columnar junction not visible.

The classifications of normal, precancerous, and cancerous cervical lesions are based on guidelines from the American Society for Colposcopy and Cervical Pathology (AS-CCP), which are widely recognized as the standard practice in gynecologic oncology for identifying and managing these conditions [45,46].

2.3. Preprocessing and Normalization of Images

For image preprocessing and normalization, the pandas [47] library was used to manage metadata, tensorflow [48] for image processing and model building, and numpy [49] for numerical operations. All images (800 × 600 pixels each) were resized to a standard size of 224 × 224 pixels using TensorFlow's image processing utilities, ensuring uniformity and facilitating efficient processing [50]. Each image was then converted into an array format suitable for input into the SWIN Transformer and CNN models. To optimize the performance of these models, the images were normalized by scaling the pixel values to the range [0, 1] through division by 255, an important step for speeding up the convergence of deep learning algorithms [51].

2.3.1. Image Selection and Train–Test Splitting

Images were selected based on case numbers and metadata using the pandas library for data manipulation and the os library [52] for file operations. A total of 360 precancerous

images from 77 cases, along with an equal number of randomly selected normal cases, were split into train–test sets using pandas. The split was performed at the case level, ensuring that all images from a single case, typically consisting of four or more images, were kept together. The training set included 54 normal and 54 precancerous cases, while the test set comprised 23 normal and 23 precancerous cases, resulting in a roughly 70/30 train–test split. This approach effectively prevented data leakage by ensuring that no images from the same case were split between the training and test sets. It also ensured the model was trained on a representative subset of the data and evaluated on unseen cases for accurate performance assessment [53]. The remaining normal cases, along with the Cancer group were held out as an additional test set. A cross-validation split was conducted at the case level, ensuring that all images from each case were included either in the training or the validation set, preventing overlaps and supporting a more reliable evaluation. In each fold, 11 normal and 11 precancerous cases (representing 20% of the total 54 normal and 54 precancerous cases) were used for validation.

2.3.2. Training Data Augmentation

For the data augmentation process, various techniques were applied to the training images to enhance the diversity of the dataset and improve model generalizations. The transformations included random resizing and cropping to 224×224 pixels; horizontal and vertical flipping; color jittering (adjustments to brightness, contrast, saturation, and hue); and random rotations up to 20 degrees. These augmented images were then normalized using the mean and standard deviation of the ImageNet dataset ([0.485, 0.456, 0.406] for the mean and [0.229, 0.224, 0.225] for standard deviations). For the test images, resizing to 256 pixels, center cropping to 224×224 pixels, and normalization were applied to maintain consistency during evaluation.

2.3.3. Architectural Flowchart and Mathematical Descriptions of the Hybrid Model

In the SWIN Transformer and CNN hybrid model, CNNs extract local features, while SWIN Transformers capture the global context (Figure 3). These features are integrated and fed into a binary classification layer, which outputs the probability of the image belonging to the precancerous/cancer (positive) class (Figure 3). The model is trained using binary cross-entropy loss to optimize its performance in distinguishing between two classes (“normal” and “precancerous”).

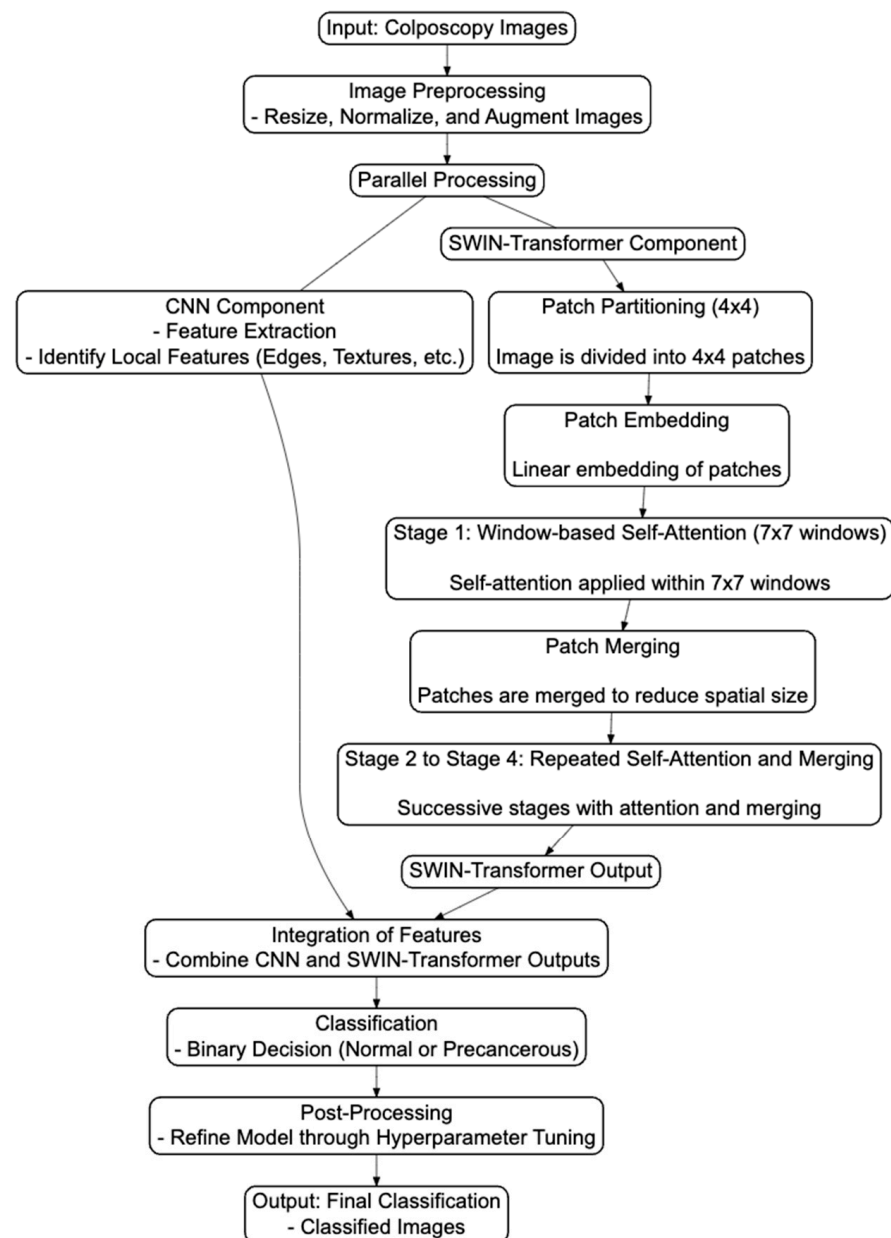


Figure 3. An architectural flowchart illustrating the integrated process where the SWIN Transformer architecture and CNN are combined into a hybrid model for the binary classification of colposcopy images. This diagram highlights the seamless interaction between the two components, demonstrating how they work together to enhance the accuracy of image classification. The CNN and SWIN Transformer processes are parallel, and both outputs are integrated before classification. This flowchart includes specific steps within the SWIN Transformer architecture (swin_base_patch4_window7_224), such as patch partitioning, embedding, window-based self-attention, and merging, before integrating with the CNN outputs. After integration, the process flows through classification, post-processing, and final output generation. Detailed steps and the Python code are available at <https://github.com/Foziyaam/SWIN-Transformer-and-CNN-for-Cervical-Cancer>.

i. Convolutional Neural Network (CNN) Component

The CNN component is responsible for extracting local features from the input image. This process typically involves convolution, activation, and pooling operations: Mathematical representation

$$\text{Convolution Operation: } X_{i+1} = f(W_1 * X_i + b_1)$$

X_l = input feature map at layer l W_l = Weight matrix at layer l
 b_l = bias at layer l * = convolution operation f = activation function (ReLU)
Pooling Operation: $X_{\text{pool}} = \text{pool}(X_{l+1})$ pool = pooling operation (max-pooling)

ii. SWIN Transformer Component:

The SWIN Transformer component captures global contextual information using a shifted window-based self-attention mechanism, which helps in understanding the overall structure and relationships within the image.

Mathematical representation

Self-Attention Mechanism: $\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$

$$Q = W^Q X(\text{Query}) \quad K = W^K X(\text{Key}) \quad V = W^V X(\text{Value})$$

$$W^Q, W^K, W^V = \text{weight matrices} \quad d_k = \text{dimension of the key vectors}$$

Shifted-Window Mechanism: Window Partition(X) → Shifted Windows

This operation divides the image into non-overlapping windows and applies self-attention within each window, enabling better global-context capturing.

iii. Integration of CNN and SWIN Transformer:

After extracting local features with the CNN and capturing the global context with the SWIN Transformer, the model integrates these features for a final classification.

Mathematical representation

Feature Integration:

$$Z = \text{Concat}(Z_{\text{CNN}}, Z_{\text{SWIN}})$$

$$Z_{\text{CNN}} = \text{features from CNN} \quad Z_{\text{SWIN}} = \text{features from SWIN - Transformer}$$

Binary Classification Layer: $\hat{y} = \sigma(W_{\text{fc}}Z + b_{\text{fc}})$

$$W_{\text{fc}} = \text{weights for the fully connected layer}$$

$$b_{\text{fc}} = \text{bias for the fully connected layer}$$

$$b_{\text{fc}} = \text{bias for the fully connected layer}$$

$$\hat{y} = \text{predicted probability for the precancerous or cancer class}$$

iv. Training Process

The model is trained to minimize the binary cross-entropy loss, which measures the difference between the predicted probability and the actual label.

Mathematical representation

Binary Cross-Entropy Loss: $\mathcal{L}(\hat{y}, y) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$

$$y = \text{actual label (0 or 1)} \quad \hat{y} = \text{predicted probability for the precancerous or cancer class}$$

Optimization: $\theta = \theta - \eta \nabla_{\theta} \mathcal{L}(\hat{y}, y)$

$$\theta = \text{model parameters} \quad \eta = \text{learning rate} \quad \nabla_{\theta} = \text{gradient with respect to parameters}$$

AdamW optimization [54]: the update rule for AdamW optimization involves additional steps. AdamW (Adaptive Moment Estimation with Weight Decay), being a more advanced optimization algorithm that improves upon traditional gradient descent, includes the following:

- a. Adaptive Learning Rates: AdamW adjusts the learning rate for each parameter individually, based on the first moment (mean) and the second moment (uncentered variance) of the gradients.

- b. **Weight Decay:** AdamW includes a weight decay term that is decoupled from the gradient-based update, helping to regularize the model by discouraging large weights.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} \mathcal{L}(\hat{y}, y) \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla_{\theta} \mathcal{L}(\hat{y}, y))^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

m_t and v_t are the first and second moment estimates, respectively

β_1 and β_2 are hyperparameters that control the decay rates of these moving averages

λ is the weight decay coefficient

AdamW is highly appropriate for training CNN and SWIN Transformer hybrid models. Its adaptive learning rates and decoupled weight decay generally lead to better convergence and improved generalizations compared to standard gradient descent.

2.4. SWIN Transformer CNN Hybrid Model Training

For the classification task, a SWIN Transformer and CNN hybrid model was designed. The SWIN Transformer architecture, specifically the `swin_base_patch4_window7_224` model [19], was chosen as a compromise between performance, computational demands, and data size, given its proven effectiveness in vision tasks and its ability to handle high-resolution images efficiently. The hybrid model combines the SWIN Transformer, which captures global contextual information, with CNN components for feature extraction, followed by a fully connected layer for classification. The output features from the SWIN Transformer undergo global average pooling, are passed through a ReLU [55] activation function, and are finally processed by a fully connected layer for binary classification. Transfer learning was achieved by incorporating pre-trained weights from large image datasets, enhancing the model's performance, given the limited size of the colposcopy image dataset. The `timm` (PyTorch Image Models) [56] library provided the pre-trained SWIN Transformer model, while the hybrid model was implemented using PyTorch [57]. This combination of techniques facilitated efficient training and reliable classification performance [58].

2.4.1. Cross-Validation Methodology for SWIN Transformer and CNN Hybrid Model

A five-fold cross-validation (CV) approach was used to optimize the training of the SWIN Transformer and CNN hybrid model. Cross-validation was implemented to ensure the model's robustness and generalizability across different subsets of the data. Here, the train dataset was split into five folds, with each fold being used once as a validation set, while the remaining folds were used for training. This process was repeated five times, allowing the model to be validated on each fold.

For each fold, the model was trained using the AdamW optimizer with a decaying learning rate, managed by a step learning rate scheduler. The performance of the model was evaluated at the end of each epoch using accuracy, F1 score, and AUC metrics on both validation and the test datasets. The best-performing model, based on test set accuracy, was saved during each epoch. This method ensures that the model is not only trained on a representative subset of the data but also validated on diverse data segments, enhancing its ability to generalize to unseen data. The cross-validation strategy employed here is critical for developing a reliable diagnostic tool that can effectively identify precancerous lesions with high accuracy and consistency.

For hyperparameter tuning, a grid search was performed over a range of learning rates, weight decay values, and batch sizes. The learning rates 0.1 , $5e-2$, $1e-2$, $5e-3$, $1e-3$, $5e-4$, $1e-4$, $5e-5$, $1e-5$, $5e-6$, and $1e-6$ were tested, along with weight decay values of $1e-2$, $1e-3$, $1e-4$, $1e-5$, and $1e-6$ and batch sizes of 16, 32, and 64. For each combination of learning rate, weight decay, and batch size, the model was trained using cross-validation, and its accuracy, F1 scores, AUCs, and other metrics were calculated. The hyperparameter

set yielding the highest accuracy was selected as the optimal configuration, with the best combination identified as the final set of hyperparameters for training.

2.4.2. Evaluation of the SWIN Transformer and CNN Models on Standard Datasets

The SWIN Transformer and CNN models were evaluated on publicly available benchmark datasets including the MNIST and the Breast Cancer Wisconsin (Diagnostic) dataset, a standard medical image data. The MNIST dataset served as a baseline for image classification tasks, allowing for a comparison with existing models. The Breast Cancer Wisconsin dataset, commonly used in medical image analyses, provided a relevant context to assess the model's performance. In both evaluations, the developed method demonstrated better performance than what has been previously reported on these datasets, indicating its effectiveness across various image classification tasks.

2.4.3. Model Performance on Test Data

The trained SWIN Transformer and CNN hybrid model was evaluated on the test dataset. Key performance metrics, including accuracy, F1 score, and Area Under the Curve (AUC), were calculated using TensorFlow and the sklearn library [59]. This evaluation provided insights into the model's ability to generalize to unseen data.

2.4.4. Evaluation of the Held-Out Cancer and Normal Data

An additional performance evaluation was conducted on the held-out dataset comprising cancer and normal images to assess the model's robustness. This step was vital for understanding the model's ability to distinguish between normal and cancerous images, particularly those it had not encountered during training. The evaluation on the held-out data using TensorFlow and sklearn demonstrated the model's effectiveness in distinguishing a normal cervix from any cancerous or other intermediate lesions (such as precancerous), indicating its potential value in diagnostic applications.

2.5. Performance Visualization

In this section, we outline key evaluation metrics, many of which are based on the confusion matrix (Table 1), which are typically structured as follows:

Table 1. Confusion matrix.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Each metric provides insights into the model's performance across different aspects, which is crucial for evaluating classification models, particularly in medical imaging contexts.

i. Sensitivity (Recall or True-Positive Rate)

Sensitivity measures the model's ability to correctly identify actual positive cases (e.g., diseased patients). A high sensitivity indicates that the model has few false negatives, meaning it is effective at capturing true positives: $sensitivity = \frac{TP}{TP+FN}$. A model with high sensitivity ensures that most of the actual positive cases are detected, reducing the risk of missing important diagnoses.

ii. Specificity (True-Negative Rate)

Specificity evaluates the model's ability to correctly identify actual negative cases. It is critical when minimizing false positives: $specificity = \frac{TN}{TN+FP}$. A high specificity means the model is effective in avoiding false positives, which is vital when a false positive can lead to unnecessary follow-ups or treatments.

iii. Positive Predictive Value (PPV or Precision)

The Positive Predictive Value (PPV), also known as precision, indicates the proportion of positive predictions that are actually correct: $PPV = \frac{TP}{TP+FP}$. High precision means that when the model predicts a positive case, it is likely to be accurate, which is particularly useful when the cost of false positives is high.

iv. Negative Predictive Value (NPV)

The Negative Predictive Value (NPV) measures the proportion of negative predictions that are correct: $NPV = \frac{TN}{TN+FN}$. A high NPV suggests the model is reliable at correctly identifying true negatives, reducing the risk of false negatives.

v. Accuracy

Accuracy represents the overall correctness of the model across all classes, considering both true positives and true negatives: $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$. Accuracy provides a general measure of how well the model is performing but may not be sufficient alone, especially in imbalanced datasets, where the model may predict the majority class correctly while failing to identify minority classes.

vi. F1 Score

The F1 score balances precision and recall, offering a single metric that takes both false positives and false negatives into account: $F1\ score = \frac{2 \times (precision \times recall)}{precision + recall}$. The F1 score is particularly useful in datasets with class imbalances, as it balances the trade-offs between precision and recall.

vii. Area Under the Curve (AUC)

The AUC-ROC score quantifies the model's ability to distinguish between positive and negative cases. A high AUC indicates that the model performs well in distinguishing between the classes across all thresholds.

Additionally, the model's performance was thoroughly examined using a variety of visualizations, including ROC curves, confusion matrices, and bar plots showing AUC and other key metrics. These visual tools offered a clear and detailed representation of the model's effectiveness, helping to convey the results comprehensively. The visualizations were generated using the matplotlib and seaborn libraries of Python [60,61] and ggplot2, along with other R packages (<https://www.r-project.org>).

Generally, the confusion matrix and related metrics provide a comprehensive understanding of the model's strengths and weaknesses across different dimensions. Sensitivity and specificity focus on the model's ability to correctly classify positive and negative cases, respectively, while PPV and NPV provide insights into the precision of predictions. Accuracy and the F1 score offer a broader measure of overall performance, and the AUC helps assess the model's discriminative power. These metrics are especially important in medical contexts, where both false positives and false negatives carry significant consequences.

2.6. Detailed Steps of Analysis and Code Availability

The Python code for model implementation and the R scripts for generating graphics are located at the GitHub repository <https://github.com/Foziyaam/SWIN-Transformer-and-CNN-for-Cervical-Cancer>. This repository contains custom scripts and detailed steps of analyses, ensuring reproducibility and transparency.

3. Results

3.1. Clinical Data Analysis

A total of 197 cases were analyzed and categorized into normal (94 cases), precancerous (77 cases), and cancerous (26 cases) (Figure 4). All samples were deemed adequate for evaluation, with the visibility of the transformation zone (TZ) varying significantly across cases: Type 1 (completely visible) in 114 cases, Type 2 (partially visible) in 33 cases, and Type 3 (not visible) in 50 cases (Figure 5). The distribution of TZ visibility within diagnostic categories showed that normal cases predominantly had Type 1 TZ (79 cases),

and precancerous cases distributed across all TZ types, while the cancer cases, except for one, had Type 3 TZ (Figure 5).

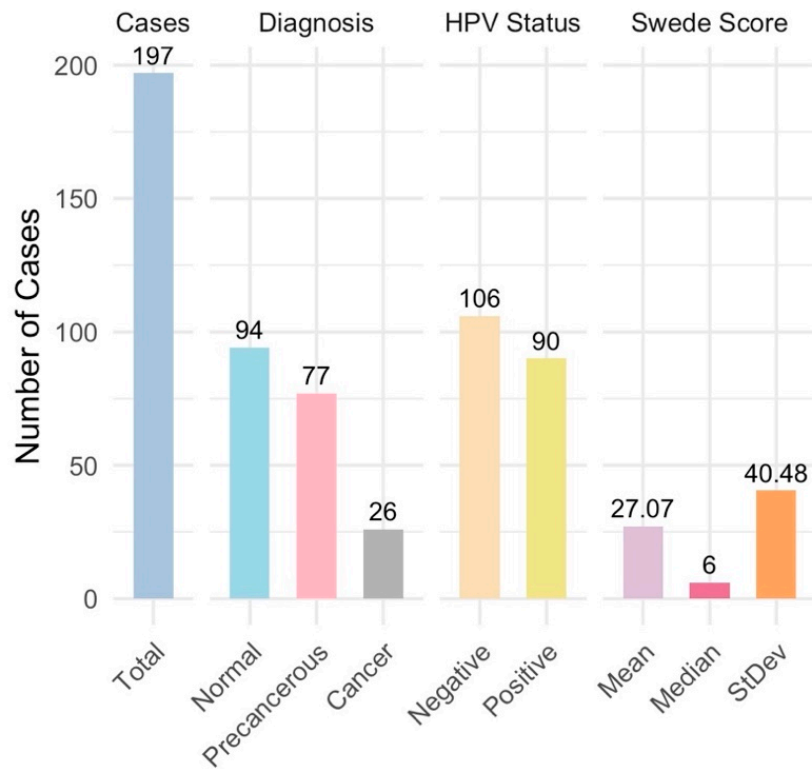


Figure 4. Summary statistics—overall case diagnosis distribution: This summary statistics provides an overview of the distribution of colposcopy cases by diagnosis, Swede score distribution, and HPV status distribution, encapsulating the key aspects of the clinical findings. The distribution of overall case diagnoses indicates a predominance of non-cancer cases, with a significant portion of precancerous and some cancer cases.

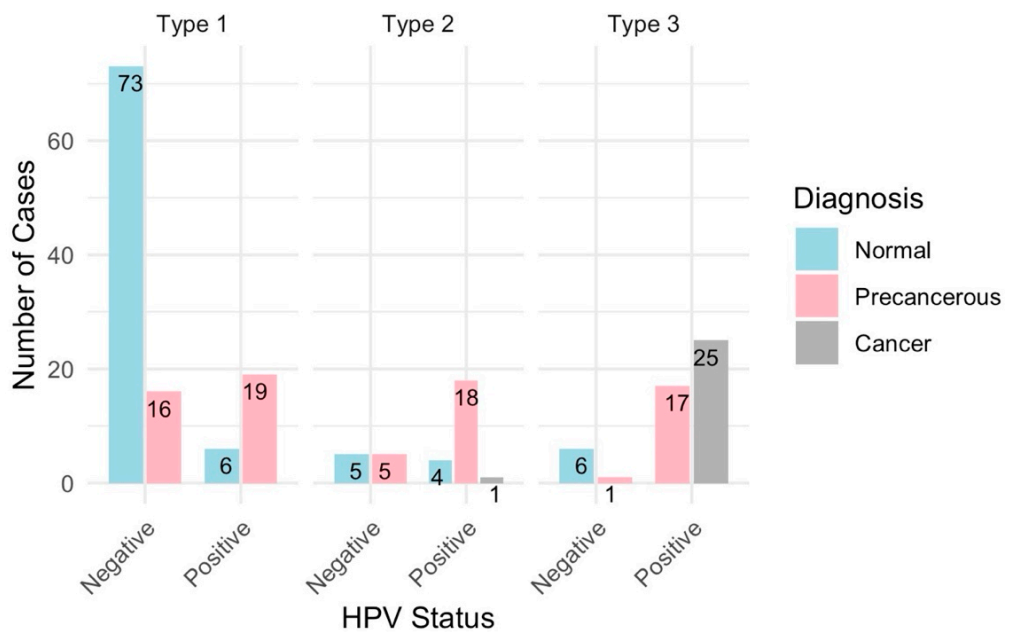


Figure 5. Distribution of cancer diagnoses by HPV status and transformation zone. These are 196 cases, since one of the cases does not have HPV test results.

The HPV status was available for 196 of 197 cases, with a higher prevalence of HPV positivity in precancerous cases (54 out of 77) and all cancer cases (26 out of 26), compared to normal cases (10 out of 94) (Figure 5). A colposcopy assessment revealed that normal findings, such as original squamous epithelium and columnar epithelium, were more common in normal cases, while abnormal findings, like dense acetowhite epithelium, punctation, and suspicious lesions, were more prevalent in precancerous and cancer cases (Figure 2). Lesions were mostly located inside the T zone, especially in the cancer cases, whereas more advanced or extensive lesions were found outside the T zone.

The Swede score analysis was an important indication of lesion severities, with the normal cases scoring between 0 and 2, precancerous cases between 3 and 7, and cancer cases > 7 (Figure 6a). Management strategies, such as LLETZ for high-grade lesions and punch biopsies for visible lesions, were determined based on lesion severity. Histopathological outcomes confirmed no significant pathological findings in the normal cases, while the precancerous cases showed varying degrees of dysplasia (LSILs, HSILs, CIN1, CIN2, and CIN3), and cancer cases were confirmed with invasive squamous cell carcinoma, adenocarcinoma, and adenocarcinoma in situ (Figure 2).

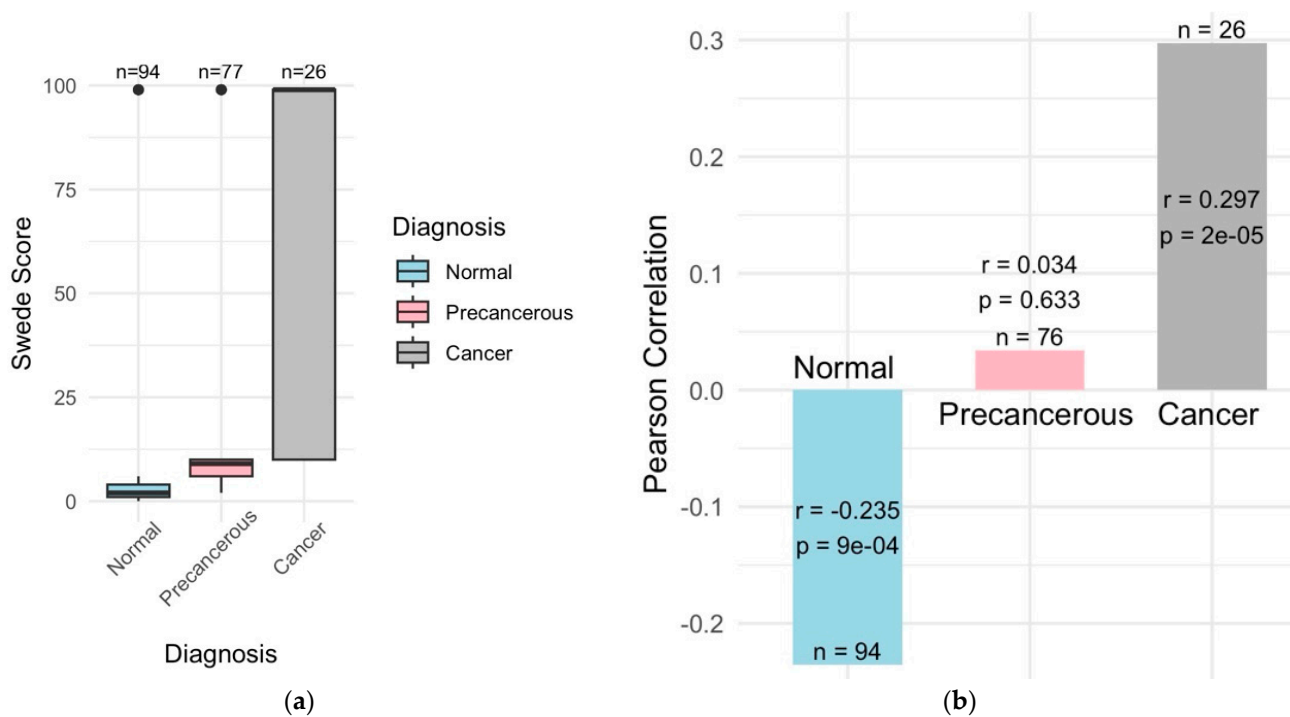


Figure 6. Correlation between Swede scores and cancer diagnosis. (a) Normal cases have the lowest-to-no Swede scores and precancerous moderate, while cancer cases have very high Swede scores. (b) Swede scores were significantly correlated with cancer diagnosis ($r = 0.3$ and $p = 2e-05$) and negatively correlated with normal diagnosis ($r = 0.2$ and $p = 9e-04$) while there was no significant correlation with the precancerous diagnosis.

3.1.1. Correlation Among HPV Positivity, Cancer Status and Transformation Zone

All the 26 cancer cases are HPV positive and predominantly have Type 3 transformation zones (Figure 5). A majority of the precancerous cases are HPV positive and exhibit Type 1, 2, and 3 transformation zones, while normal cases are primarily HPV negative and have mainly Type 1 transformation zones (Figure 5).

3.1.2. Swede Score Distributions

The Swede score distributions varied by cancer status and HPV status. Higher Swede scores were associated with cancer diagnoses and HPV positivity (Figure 6b).

3.1.3. Summary of the Clinical Findings

Key findings of the clinical data analysis include a high prevalence of HPV in both precancerous and cancer cases and significant differences in colposcopy findings across diagnostic categories. Higher Swede scores were significantly correlated with lesion severity and showed good agreement with histopathological outcomes. The visibility of TZs was related to lesion progression, with higher-grade lesions often being associated with less-visible TZs.

These findings highlight the effectiveness of colposcopy evaluations and the relevance of Swede scores in assessing lesion severity. The strong correlation between HPV status and histopathological outcomes further reinforces HPV's role as a significant risk factor for cervical neoplasia.

3.2. Classifications of Colposcopy Images

3.2.1. SWIN Transformer and CNN Hybrid Model Training with Fivefold Cross-Validation

The hybrid model, combining the SWIN Transformer and convolutional neural networks (CNNs), was trained on a dataset consisting of 486 images (from 54 normal and 54 precancerous cases). The model was trained using transfer learning techniques with pre-trained weights for 30 epochs (Figure 7a) with a batch size of 32, a learning rate of $5e-05$, a weight decay of 0.05, and a gamma of 0.8, and with a five-fold cross-validation. The model achieved a validation Area Under the Curve (AUC) of 94% (Figure 7b) and an accuracy of 87% (Figure 7c), indicating that it was able to classify most of the validating images into normal and precancerous categories, as shown by the corresponding confusion matrix (Figure 7c). Other metrics that were calculated based on the confusion matrix include validation sensitivity, 0.86; specificity, 0.90; the positive predictive value (precision), 0.92; and the negative predictive value, 0.81, which were derived from the confusion matrix. The training loss decreased sharply with increased epochs (Figure 7a), indicating efficient learning conditions (including optimal hyperparameter combinations).

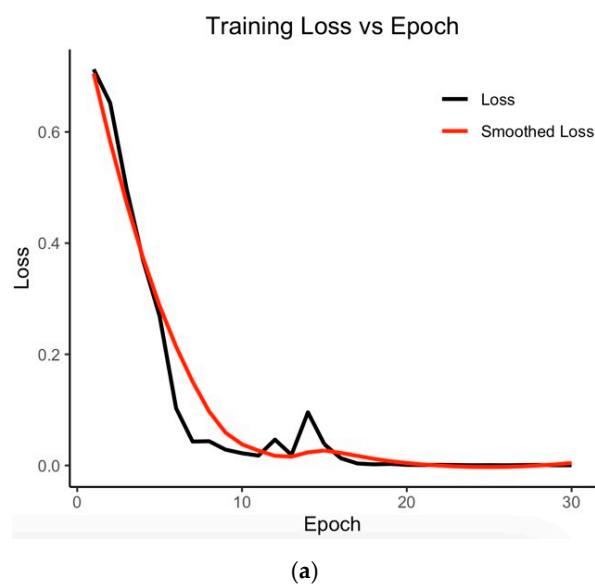


Figure 7. Cont.

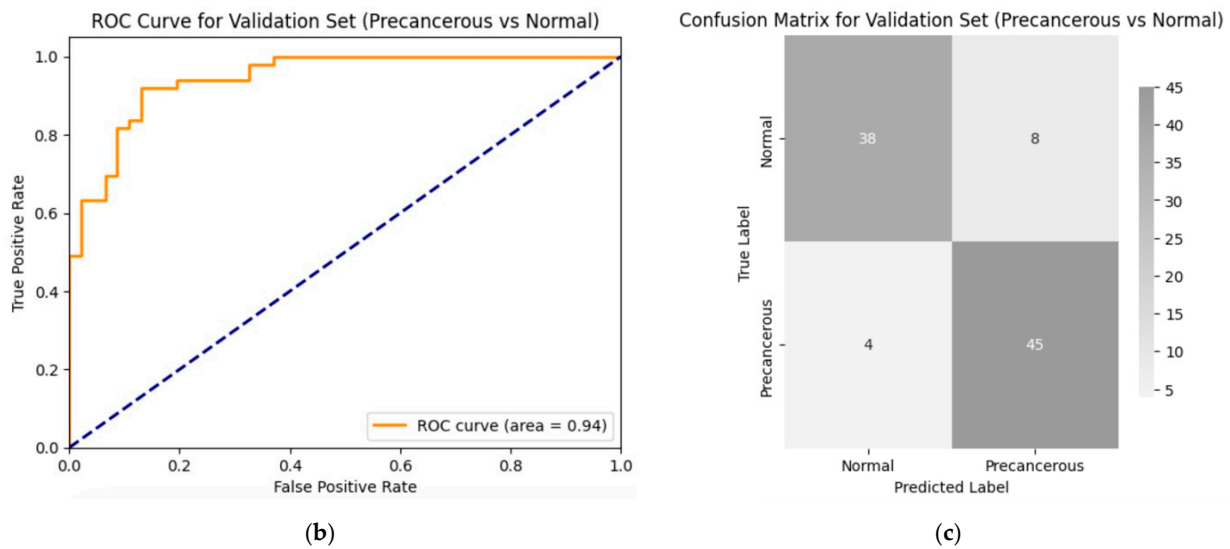


Figure 7. (a) Training loss across epochs; fivefold cross-validation metrics (red curve is the smoothing of the actual curve—the black line): (b) validation ROC curve for the validation set; (c) confusion matrix for the validation set. Validation sensitivity, 0.86; specificity, 0.90; positive predictive value (precision), 0.92; negative predictive value, 0.81; accuracy, 0.87; F1 score, 0.88; and AUC, 0.94.

3.2.2. Trained Model's Performance on the Test Set 1 (Precancerous vs. Normal)

The performance of the hybrid model was evaluated on a test set of 222 images (from 23 normal and 23 precancerous cases). The model achieved an AUC of approximately 80% (Figure 8a) with an accuracy of 75% (Figure 8b) on the test set. These results demonstrate that the model can correctly distinguish between normal and precancerous colposcopy images in four out of five cases. Additional values of the model performance metrics include sensitivity, 0.75; specificity, 0.75; the positive predictive value, 0.76; and the negative predictive value, 0.74. The performance was tested using the same hyperparameters that were used for training: batch size = 32, epochs = 30, learning rate = $5e-05$, weight decay = $5e-02$, and gamma = 0.8.

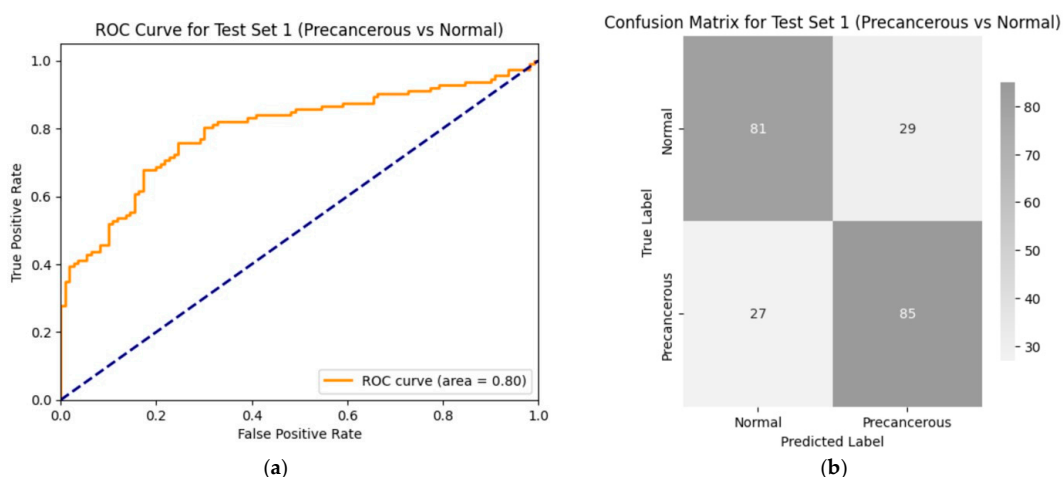


Figure 8. Performance of the trained model on the first test data (precancerous vs. normal): (a) ROC curve for test set 1 (precancerous versus normal); (b) confusion matrix for the performance of the model on test set 1 (precancerous vs. normal group). The values of the model's performance metrics include sensitivity, 0.75; specificity, 0.75; positive predictive value, 0.76; negative predictive value, 0.74; accuracy, 0.75; F1 score, 0.75; and AUC, 0.80. The performance was tested using the same hyperparameters that were used for training: batch size = 32, epochs = 30, learning rate = $5e-05$, weight decay = $5e-02$, and gamma = 0.8.

3.2.3. Performance Evaluation on Test Set 2 (Cancer vs. Normal)

To further assess the robustness and generalizability of the model, it was evaluated on the second test set consisting of 233 equivalent numbers of cancer and normal images. The trained SWIN Transformer and CNN classifier model achieved an AUC of 82% (Figure 9a) and an accuracy of 75% (Figure 9b) on the second hold-out test set. This high performance indicates the model's efficiency in distinguishing normal images from any abnormal images it was not trained on, particularly for cancer images which were not even the same type as that of the original training data. Additional values of the important metrics include sensitivity, 0.72; specificity, 0.80; positive predictive value, 0.85; negative predictive value, 0.65; and F1, 0.78. The same hyperparameters were used for this evaluation as well.

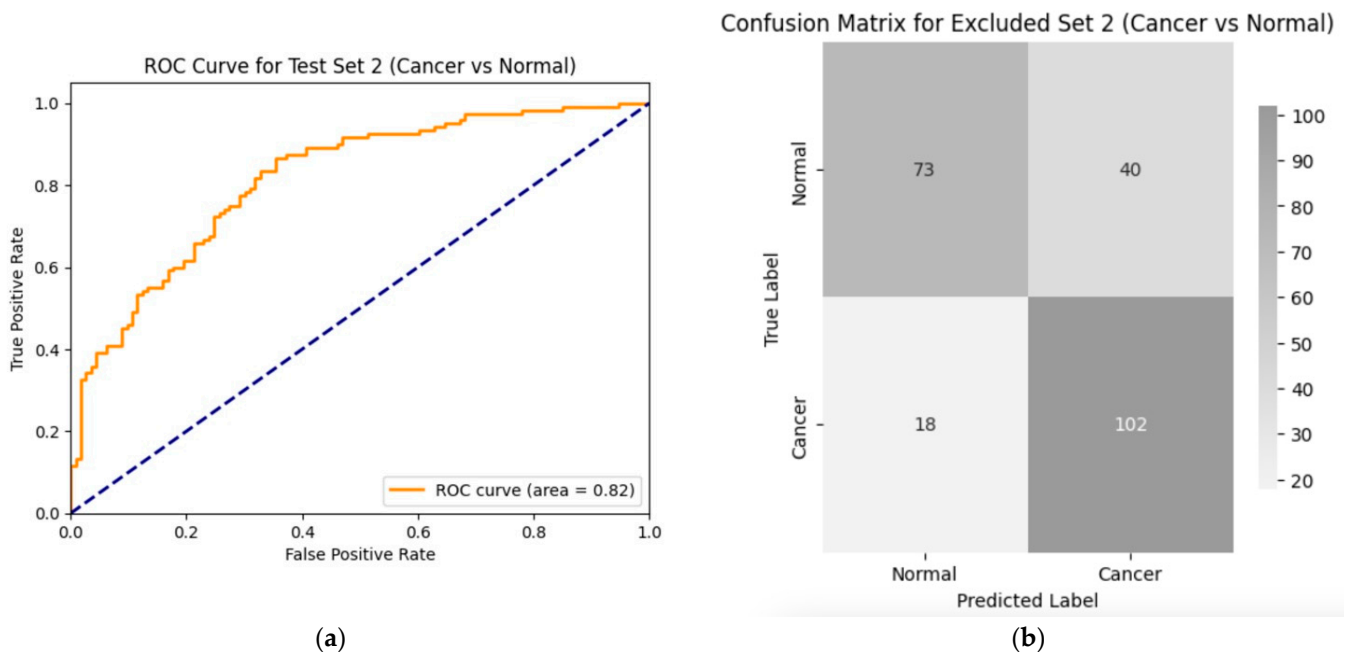


Figure 9. Performance of the second test set (images from cancer and normal cases): (a) ROC curve for test set 2 (cancer versus normal); (b) confusion matrix for the performance of the model on test set 2 (cancer vs. normal). Values of the important metrics include sensitivity, 0.72; specificity, 0.80; positive predictive value, 0.85; negative predictive value, 0.65; accuracy, 0.75; F1, 0.78; and AUC, 0.82. The same hyperparameters were used for this evaluation as well.

3.2.4. Performance of the SWIN Transformer and CNN Models on Standard Datasets

An evaluation of the SWIN Transformer and CNN models on the MNIST dataset achieved almost perfect classification, correctly classifying the digits from 0 to 9 of the test data, indicating the generalizability of the model. Additionally, an evaluation of the model on the Breast Cancer Wisconsin (Diagnostic) dataset, a modest standard medical image dataset, and the accuracy was above 96%. The high performance of the model on the Breast Cancer Wisconsin dataset provided very high confidence in efficiency and relevance of the model with regard to medical image classifications, providing a relevant context to assess the model's performance. In both evaluations, the developed method demonstrated better performance than what has been previously reported on these datasets, indicating its effectiveness across various image classification tasks.

3.2.5. Model Performance Summary

The results of this study demonstrate the potential of this hybrid model in cervical cancer diagnosis, with strong performance on validation and independent test sets. The model showed reliable generalizations to new, unseen data, particularly for cancer images not included in the training phase. By combining CNN's local feature extraction with the

SWIN Transformer's global context processing, the model successfully identified differences between normal, precancerous, and cancerous cases. While the results are promising, variability in image quality and clinical data suggests further refinements may improve the model's application in early-stage cervical cancer screening. Overall, the SWIN Transformer and CNN hybrid model offers a practical approach for medical image classifications and early diagnoses.

The SWIN Transformer and CNN hybrid model for a precancerous colposcopy classification integrates both architectures to provide accurate and interpretable results. The CNN component focuses on extracting local features, such as edges, textures, and shapes, while the SWIN Transformer captures both local and global contextual information, modeling relationships across the image. These features are combined and processed through a classification layer, effectively distinguishing subtle differences in tissues, which is important for identifying precancerous lesions. The model benefits from transfer learning and data augmentation, allowing it to perform well despite limited data. Techniques like contrastive learning and multi-task learning also help the model concentrate on disease-specific patterns while reducing the influence of confounding factors.

3.2.6. Complementing Traditional Methods

The hybrid deep learning model improves the diagnostic process by providing a consistent, accurate analysis during colposcopy, complementing primary screening tools like cytology (Pap smears) and HPV testing, which can vary in interpretation and have lower sensitivity [15,16]. A human interpretation of colposcopy images is subjective, varying due to the clinician's experience, fatigue, or bias, which can lead to inconsistent diagnoses. Subtle lesions may be overlooked, and the process is time-consuming and resource-intensive. The hybrid model can potentially address these challenges by offering a standardized, objective analysis, detecting subtle features missed by humans and providing rapid, real-time diagnostic support with better scalability and accessibility, especially in resource-limited settings. It reduces possible human errors, offering advantages in consistency, accuracy, and scalability, and ultimately enhancing the early detection of precancerous lesions.

Particularly, the SWIN Transformer's ability to model long-range dependencies and capture the global context makes it particularly well-suited for tasks like colposcopy image classification, where subtle differences in tissue structures can be critical. These results suggest that integrating AI-driven diagnostic tools into clinical practice could significantly improve the accuracy and efficiency of cervical cancer screening programs, especially in regions with limited healthcare resources.

4. Discussion

Cervical cancer remains a significant public health challenge, especially in low-resource settings where early diagnosis and timely treatment are often hindered by limited access to healthcare. The WHO has highlighted the importance of early detection in reducing cervical cancer-related morbidity and mortality, emphasizing the need for more effective diagnostic tools in these regions [62]. The results of this study, which employed a SWIN Transformer and CNN hybrid model for the classification of colposcopy images, indicate that such advanced machine learning techniques have the potential to mitigate these challenges.

The SWIN Transformer and CNN hybrid model achieved an Area Under the Curve (AUC) of approximately 80% to 82% and an accuracy of 75% on two independent test datasets, a notable achievement given the moderate size of the training data. These performance metrics indicate the model's effectiveness in distinguishing between normal and precancerous colposcopy images, as well as between normal and cancerous images that were not included in the initial training set. To contextualize this performance, it is worth noting that the US Food and Drug Administration-approved MammaPrint—a 70-gene expression profile test used to assess the risk of breast cancer recurrence in early-

stage patients—reports an AUC ranging between 68% and 75% [63–65]. This comparison highlights the significance of our model's results, particularly in the early detection of precancerous lesions, which is important for preventing the progression to invasive cervical cancer, a condition with high mortality rates [4]. The model's robust performance on unseen data further suggests its potential utility in clinical settings, where generalizability across diverse patient populations is essential.

The model's ability to generalize well, even when faced with unseen cancer images, indicates its robustness and suggests that it could be a valuable tool in clinical practice to differentiate normal from any other type of lesion abnormalities. Previous studies have shown that deep learning models can outperform traditional diagnostic methods in medical imaging, such as CNNs achieving a dermatologist-level classification in skin cancer diagnosis [9] and superior performance in detecting pneumonia from chest X-rays [66]. The success of the SWIN Transformer CNN hybrid model in this study aligns with these findings, indicating that such models could become a critical component of cervical cancer screening programs, particularly in resource-limited settings, where access to expert cytologists is scarce [67].

4.1. Clinical Implications

The improved accuracy and AUC of the hybrid model indicate its potential for clinical applications. Early detection facilitated by such models can lead to timely interventions, reducing the progression to invasive cervical cancer and lowering mortality rates [6,8]. The hybrid model supplements human interpretation with a standardized, algorithm-driven analysis, ensuring the accurate identification of LSILs and HSILs based on specific visual and textural features. This model enhances detection and classification capabilities by providing a nuanced and comprehensive analysis of colposcopy images, integrating detailed image assessments with a broader contextual understanding. The model's real-time analysis and continuous learning from diverse datasets strengthen its generalizability across different patient populations, making it a reliable tool in various clinical settings. The developed models demonstrate the potential to effectively differentiate normal cervical images from abnormal lesions. This capability extends beyond distinguishing precancerous lesions, as the models can also reasonably distinguish normal images from those of cancer images that were not the same type as that of the train set.

This study highlights the potential of combining SWIN Transformers and CNN models for the early diagnosis of cervical cancer using colposcopy images. If integrated into existing screening programs, this AI-driven approach could standardize the evaluation of colposcopy images, reducing the variability that is often faced in traditional methods like Pap smears [15] and HPV testing [16]. By providing more consistent and accurate diagnoses, such models could help reduce the number of false positives, thereby lowering the psychological and economic burden on patients and healthcare systems alike. This is particularly relevant in low- and middle-income countries, where healthcare resources are often stretched thin, and the burden of cervical cancer is disproportionately high [4].

Moreover, the model's performance suggests that similar approaches could be applied to other areas of medical imaging. The versatility of the SWIN Transformer, which captures global contextual information, combined with the CNN's strength in feature extraction, makes this hybrid approach well-suited for a wide range of diagnostic tasks. Future studies should explore the application of this model to other cancers and diseases, potentially extending its use to mammography, lung cancer screening, and beyond. The success of such models could lead to the development of comprehensive AI-driven diagnostic platforms that integrate multiple imaging modalities, further enhancing the early detection and treatment of various conditions [68].

4.2. Limitations and Future Directions

Despite these promising results, several challenges remain. One of the most significant barriers to the widespread adoption of AI in healthcare is the need for large, annotated

datasets to train and validate these models. While transfer learning and data augmentation can mitigate this issue by leveraging pre-trained models and generating new data, it is important to expose the models to a diverse range of cases to ensure generalizability [21]. For example, validating these findings with larger and more diverse datasets, and integrating additional data sources, such as multi-omics, genomics, and patient history, could significantly enhance the model's robustness and make it a more comprehensive diagnostic tool [21,29].

In addition, integrating AI tools into clinical workflows will require rigorous validation and adherence to regulatory standards to ensure their safety and effectiveness [69]. The ethical considerations surrounding the use of AI in healthcare, including issues of data privacy, algorithmic transparency, and potential biases, must also be addressed to build trust among clinicians and patients [70].

Looking forward, the integration of much larger multimodal datasets could significantly enhance the model's diagnostic accuracy and robustness. This holistic approach would provide a more comprehensive understanding of disease progression and could lead to more personalized treatment strategies [71]. The development of explainable AI models, which offer insights into the decision-making process of the algorithms, could further facilitate the adoption of these tools in clinical practice by allowing clinicians to better understand and trust the AI's recommendations [71].

5. Conclusions

This study highlights the potential of SWIN Transformer and CNN hybrid models for the early diagnosis of cervical cancer through colposcopy image classifications. The models demonstrated strong performance, with AUC values ranging between 80% and 82% and with an accuracy of 75% to 78% on independent test datasets. These results indicate the models' effectiveness in distinguishing between normal, precancerous, and cancerous colposcopy images, particularly in enhancing diagnostic accuracy after abnormal findings from traditional screening methods like cytology and HPV testing.

These models not only improve the early detection of cervical cancer but also offer the potential to reduce unnecessary follow-up procedures and treatments due to false positives, thereby reducing the psychological and economic burden on patients and healthcare systems. Their ability to generalize across diverse patient populations and perform real-time analyses makes them a valuable tool in clinical practice, especially in resource-limited settings with limited access to expert colposcopists.

To maximize the benefits of these models, future research should focus on validating these findings with larger and more diverse datasets. Incorporating additional data sources, such as multi-omics and genomics, could further enhance the models' robustness and diagnostic accuracy, paving the way for more comprehensive and personalized treatment strategies. Addressing challenges related to dataset diversity, regulatory approval, and ethical considerations, such as data privacy and algorithmic transparency, is also crucial for the successful integration of AI-driven diagnostic tools into clinical workflows.

Developing reliable and accessible diagnostic systems using SWIN Transformer and CNN hybrid models has the potential to significantly reduce the burden of cervical cancer, especially in underserved regions. The findings from this study may guide policy decisions on cervical cancer screening and prevention, ultimately leading to better public health outcomes and an enhanced quality of life for women worldwide.

Author Contributions: Conceptualization, F.A.M., K.K.T. and S.M.; methodology, F.A.M.; software, F.A.M.; formal analysis, F.A.M.; investigation, F.A.M. and S.M.; resources, F.A.M. and S.M.; data curation, F.A.M.; writing—original draft preparation, F.A.M.; writing—review, critical feedback, professional commenting, and editing, J.A.M., T.A.W. and S.M.; visualization, F.A.M. and S.M.; supervision, K.K.T. and S.M.; project administration, K.K.T. and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study used de-identified and publicly available clinical and molecular datasets.

Informed Consent Statement: Not applicable.

Data Availability Statement: Custom Python and R scripts and detailed steps of the analyses are available at the GitHub repository: <https://github.com/Foziyaam/SWIN-Transformer-and-CNN-for-Cervical-Cancer>.

Acknowledgments: FAM was financially and logistically supported by Wolkite University, Wolkite, Ethiopia, and Addis Ababa Science and Technology University, Addis Ababa, Ethiopia. The training colposcopy datasets, along with their metadata (images + metadata), were obtained from the International Agency for Research on Cancer (IARC): <https://www.iarc.who.int/cancer-type/cervical-cancer>. The IARC Colposcopy Image Bank (IARCImageBankColpo) is a resource that provides images and information related to colposcopy assessments, hosted at the IARC Cervical Cancer Image Bank: <https://screening.iarc.fr/cervicalimagebank.php>. We are deeply grateful to Eric Lucas from the IARC for providing us with a secure link to the datasets, without which this study would not have been possible. Thank you, Eric, for your invaluable support.

Conflicts of Interest: Author Seid Muhie was employed by Enkoy LLC, the remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The abbreviations used in this manuscript include AUC—Area Under the Curve, CNN—Convolutional Neural Network, HPV—Human Papillomavirus, WHO—World Health Organization, LSIL—Low-grade Squamous Intraepithelial Lesion, CIN1—Cervical Intraepithelial Neoplasia grade 1, HSIL—High-grade Squamous Intraepithelial Lesion, CIN2—Cervical Intraepithelial Neoplasia grade 2, CIN3—Cervical Intraepithelial Neoplasia grade 3, ASCCP—American Society for Colposcopy and Cervical Pathology, TZ—Transformation Zone, LLETZ—Large Loop Excision of the Transformation Zone, ROC—Receiver Operating Characteristic, CV—Cross-Validation, ReLU—Rectified Linear Unit, and AI—Artificial Intelligence.

References

1. WHO. World Health Organization Cervical Cancer: Key Facts. Available online: <https://www.who.int/news-room/fact-sheets/detail/cervical-cancer> (accessed on 29 August 2024).
2. International Agency for Research on Cancer (IARC). New Estimates of the Number of Cervical Cancer Cases and Deaths That Could Be Averted in Low-Income and Lower-Middle-Income Countries by Scale-up of Screening and Vaccination Activities. 2024. Available online: https://www.iarc.who.int/news-events/new_estimates_cervical_cancer_cases_and_deaths_lmhc_screening_vaccines/ (accessed on 24 April 2024).
3. Ferlay, J.; Ervik, M.; Lam, F.; Colombet, M.; Mery, L.; Piñeros, M.; Znaor, A.; Soerjomataram, I.; Bray, F. Global Cancer Observatory: Cancer Today 2019. Available online: <https://publications.iarc.fr/Databases/Iarc-Cancerbases/Cancer-Today-Powered-By-GLOBOCAN-2018--2018> (accessed on 10 August 2024).
4. Arbyn, M.; Weiderpass, E.; Bruni, L.; de Sanjosé, S.; Saraiya, M.; Ferlay, J.; Bray, F. Estimates of Incidence and Mortality of Cervical Cancer in 2018: A Worldwide Analysis. *Lancet Glob. Health* **2020**, *8*, e191–e203. [[CrossRef](#)] [[PubMed](#)]
5. Mohammed, F.A.; Tune, K.K.; Jett, M.; Muhie, S. Cervical Cancer Stages, Human Papillomavirus Integration, and Malignant Genetic Mutations: Integrative Analysis of Datasets from Four Different Cohorts. *Cancers* **2023**, *15*, 5595. [[CrossRef](#)] [[PubMed](#)]
6. Kim, J.J.; Burger, E.A.; Regan, C.; Sy, S. Screening for Cervical Cancer in Primary Care: A Decision Analysis for the US Preventive Services Task Force. *JAMA* **2016**, *320*, 706–714. [[CrossRef](#)] [[PubMed](#)]
7. Castle, P.E.; Maza, M. Prophylactic HPV Vaccination: Past, Present, and Future. *Epidemiol. Infect.* **2016**, *144*, 449–468. [[CrossRef](#)]
8. Saslow, D.; Solomon, D.; Lawson, H.W.; Killackey, M.; Kulasingam, S.L.; Cain, J.; Garcia, F.A.R.; Moriarty, A.T.; Waxman, A.G.; Wilbur, D.C.; et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology Screening Guidelines for the Prevention and Early Detection of Cervical Cancer. *CA Cancer J. Clin.* **2012**, *62*, 147–172. [[CrossRef](#)]
9. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]

10. Eriksson, E.M.; Lau, M.; Jönsson, C.; Zhang, C.; Risö Bergerlind, L.L.; Jonasson, J.M.; Strander, B. Participation in a Swedish cervical cancer screening program among women with psychiatric diagnoses: A population-based cohort study. *BMC Public Health* **2019**, *19*, 313. [[CrossRef](#)]
11. Tsikouras, P.; Zervoudis, S.; Manav, B.; Tomara, E.; Iatrakis, G.; Romanidis, C.; Bothou, A.; Galazios, G. Cervical Cancer: Screening, Diagnosis and Staging. *J. BUON* **2016**, *21*, 320–325.
12. Mitchell, M.F.; Schottenfeld, D.; Tortolero-Luna, G.; Cantor, S.B.; Richards-Kortum, R. Colposcopy for the Diagnosis of Squamous Intraepithelial Lesions: A Meta-Analysis. *Obstet. Gynecol.* **1998**, *91*, 626–631. [[CrossRef](#)]
13. Basu, P.; Meheus, F.; Chami, Y.; Hariprasa, R.; Zhao, F.; Sankaranarayanan, R. Management algorithms for cervical cancer screening and precancer treatment for resource-limited settings. *Int. J. Gynecol. Obstet.* **2017**, *138*, 26–32. [[CrossRef](#)]
14. Peirson, L.; Fitzpatrick-Lewis, D.; Ciliska, D.; Warren, R. Screening for Cervical Cancer: A Systematic Review and Meta-Analysis. *Syst. Rev.* **2013**, *2*, 35. [[CrossRef](#)] [[PubMed](#)]
15. Cuzick, J.; Clavel, C.; Petry, K.U.; Meijer, C.J.; Hoyer, H.; Ratnam, S.; Szarewski, A.; Birembaut, P.; Kulasingam, S.; Sasieni, P.; et al. Overview of the European and North American Studies on HPV Testing in Primary Cervical Cancer Screening. *Int. J. Cancer* **2006**, *119*, 1095–1101. [[CrossRef](#)] [[PubMed](#)]
16. Ronco, G.; Dillner, J.; Elfström, K.M.; Tunesi, S.; Snijders, P.J.F.; Arbyn, M.; Kitchener, H.; Segnan, N.; Gilham, C.; Giorgi-Rossi, P.; et al. Efficacy of HPV-Based Screening for Prevention of Invasive Cervical Cancer: Follow-up of Four European Randomised Controlled Trials. *Lancet* **2014**, *383*, 524–532. [[CrossRef](#)] [[PubMed](#)]
17. Arbyn, M.; Ronco, G.; Anttila, A.; Meijer, C.J.L.M.; Poljak, M.; Ogilvie, G.; Koliopoulos, G.; Naucler, P.; Sankaranarayanan, R.; Peto, J. Evidence Regarding Human Papillomavirus Testing in Secondary Prevention of Cervical Cancer. *Vaccine* **2009**, *27*, A88–A99. [[CrossRef](#)]
18. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
19. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *arXiv* **2021**, arXiv:210314030.
20. Mohammed, F.A.; Tune, K.K.; Assefa, B.G.; Jett, M.; Muhie, S. Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 699–735. [[CrossRef](#)]
21. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
22. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
23. Zhang, T.; Luo, Y.M.; Li, P.; Liu, P.Z.; Du, Y.Z.; Sun, P.; Dong, B.; Xue, H. Cervical precancerous lesions classification using pre-trained densely connected convolutional networks with colposcopy images. *Biomed. Signal Process. Control* **2020**, *55*, 101566. [[CrossRef](#)]
24. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410. [[CrossRef](#)]
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *Computation and Language (cs.CL); Machine Learning (cs.LG)*. *arXiv* **2017**, arXiv:1706.03762.
26. Liu, Y.; Ma, Y.; Zhu, Z.; Cheng, J.; Chen, X. TransSea: Hybrid CNN–Transformer with Semantic Awareness for 3-D Brain Tumor Segmentation. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2521316. [[CrossRef](#)]
27. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Houshy, N. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:201011929.
28. Islam, S.; Elmekki, H.; Elsebai, A.; Bentahar, J.; Drawel, N.; Rjoub, G.; Pedrycz, W. A comprehensive survey on applications of transformers for deep learning tasks. *Expert Syst. Appl.* **2024**, *241*, 122666. [[CrossRef](#)]
29. Wright, T.C.; Stoler, M.H.; Behrens, C.M.; Sharma, A.; Zhang, G.; Wright, T.L. Primary Cervical Cancer Screening with Human Papillomavirus: End of Study Results from the ATHENA Study Using HPV as the First-Line Screening Test. *Gynecol. Oncol.* **2015**, *136*, 189–197. [[CrossRef](#)]
30. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, Z. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:210204306.
31. Chandran, V.; Sumithra, M.G.; Karthick, A.; George, T.; Deivakani, M.; Elakkiya, B.; Subramaniam, U.; Manoharan, S. Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images. *BioMed Res. Int.* **2021**. [[CrossRef](#)] [[PubMed](#)]
32. Kelly, C.J.; Karthikesalingam, A.; Suleyman, M.; Corrado, G.; King, D. Key Challenges for Delivering Clinical Impact with Artificial Intelligence. *BMC Med.* **2019**, *17*, 195. [[CrossRef](#)]
33. Wentzensen, N.; Arbyn, M.; Berkhof, J.; Vinokurova, S. New Technologies for Cervical Cancer Screening. *Vaccine* **2017**, *35*, 4656–4665. [[CrossRef](#)]
34. Schiffman, M.; Castle, P.E.; Jeronimo, J.; Rodriguez, A.C.; Wacholder, S. Human Papillomavirus and Cervical Cancer. *Lancet* **2007**, *370*, 890–907. [[CrossRef](#)]

35. Park, K.-B.; Lee, J.-Y. SwinE-Net: Hybrid deep learning approach to novel polyp segmentation using convolutional neural network and Swin Transformer. *J. Comput. Des. Eng.* **2022**, *9*, 616–632. [[CrossRef](#)]
36. Chen, S.; Ma, X.; Zhang, Y. SWIN-Transformer and CNN Hybrid Model for Medical Image Classification and Interpretability. *Med. Image Anal.* **2022**, *72*, 102102. [[CrossRef](#)]
37. Yang, W.; Xu, L.; Huang, S. Multimodal Medical Image Fusion Using SWIN-Transformer and CNN Hybrid Model for Alzheimer's Disease Diagnosis. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 267–276.
38. Xie, Y.; Zhao, J.; Wang, R. SWIN-Transformer for Precise Organ Segmentation in Medical Imaging. *Med. Phys.* **2022**, *49*, 2795–2804. [[CrossRef](#)]
39. Huang, X.; Liu, J.; Xu, L. Brain Tumor Segmentation with SWIN-Transformer and CNN Hybrid Model. *Int. J. Comput. Assist. Radiol. Surg.* **2022**, *17*, 431–441. [[CrossRef](#)]
40. Kim, J.; Park, J.; Choi, H. Hybrid Transformer-CNN Models for Accurate Diagnosis of COVID-19 from Chest X-Rays. *IEEE Access* **2022**, *10*, 78542–78555. [[CrossRef](#)]
41. Zhou, X.; Zhang, Y.; Li, S. Hybrid Deep Learning Models for Polyp Detection in Colonoscopy Images. *Pattern Recognit. Lett.* **2022**, *157*, 75–83. [[CrossRef](#)]
42. Tan, M.; Le, Q.; Vu, B. Data Augmentation and Transfer Learning in SWIN-Transformer-CNN Hybrid Models for Medical Imaging. *J. Med. Artif. Intell.* **2022**, *4*, 125–133.
43. Zhang, X.; Hu, Z.; Chen, Y. Transformer-CNN Hybrid Models in Medical Image Segmentation: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 589–603.
44. International Agency for Research on Cancer Colposcopy Image Bank 2024. Available online: <https://screening.iarc.fr/cervicalimagebank.php> (accessed on 26 January 2024).
45. Massad, L.S.; Einstein, M.H.; Huh, W.K.; Katki, H.A.; Kinney, W.K.; Schiffman, M.; Solomon, D.; Wentzensen, N.; Lawson, H.W. 2012 Updated Consensus Guidelines for the Management of Abnormal Cervical Cancer Screening Tests and Cancer Precursors. *Obstet. Gynecol.* **2013**, *121*, 829–846. [[CrossRef](#)]
46. Perkins, R.B.; Guido, R.S.; Castle, P.E.; Chelmow, D.; Einstein, M.H.; Garcia, F.; Huh, W.K.; Kim, J.J.; Moscicki, A.B.; Nayar, R.; et al. 2019 ASCCP Risk-Based Management Consensus Guidelines. *J. Low. Genit. Tract Dis.* **2020**, *24*, 102–131. [[CrossRef](#)] [[PubMed](#)]
47. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, TX, USA, 28 June–3 July 2010; pp. 51–56.
48. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation (OSDI '16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
49. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)] [[PubMed](#)]
50. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. *arXiv* **2016**, arXiv:160304467.
51. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
52. Python Software Foundation The Python Standard Library: Os—Miscellaneous Operating System Interfaces 2023. Python v3.13.0 Documentation. Available online: <https://docs.python.org/3/library/os.html> (accessed on 29 August 2024).
53. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009.
54. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the 7th International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
55. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
56. Wightman, R. PyTorch Image Models (Timm). 2019. Available online: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C21&q=Wightman%2C+R.+PyTorch+Image+Models+%28Timm%29&btnG= (accessed on 10 August 2024).
57. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8024–8035.
58. Wightman, R. PyTorch Image Models. 2019. Available online: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C21&q=Wightman%2C+R.+PyTorch+Image+Models&btnG= (accessed on 10 August 2024).
59. Pedregosa, F. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
60. Waskom, M.L. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [[CrossRef](#)]
61. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
62. World Health Organization (WHO). *Global Strategy to Accelerate the Elimination of Cervical Cancer as a Public Health Problem 2020*; WHO: Geneva, Switzerland, 2020.
63. Van't Veer, L.J.; Dai, H.; Van De Vijver, M.J.; He, Y.D.; Hart, A.A.; Mao, M.; Peterse, H.L.; Van Der Kooy, K.; Marton, M.J.; Witteveen, A.T.; et al. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. *Nature* **2002**, *415*, 530–536. [[CrossRef](#)]
64. Bueno-de-Mesquita, J.M.; Linn, S.C.; Keijzer, R.; Wesseling, J.; Nuyten, D.S.A.; Van Krimpen, C.; Meijers, C.; De Graaf, P.W.; Bos, M.M.E.M.; Hart, A.A.M.; et al. Validation of 70-gene prognosis signature in node-negative breast cancer. *Breast Cancer Res Treat.* **2009**, *117*, 483–495. [[CrossRef](#)]

65. Cardoso, F.; Van't Veer, L.J.; Bogaerts, J.; Slaets, L.; Viale, G.; Delaloge, S.; Pierga, J.Y.; Brain, E.; Passalacqua, R.; Neijenhuis, P.A.; et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N. Engl. J. Med.* **2016**, *375*, 717–729. [[CrossRef](#)] [[PubMed](#)]
66. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ng, A.Y. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-rays with Deep Learning. *arXiv* **2018**, arXiv:171105225.
67. Mitchell, M.J.; Billingsley, M.M.; Haley, R.M.; Wechsler, M.E.; Peppas, N.A.; Langer, R. Engineering Precision Nanoparticles for Drug Delivery. *Nat. Rev. Drug Discov.* **2019**, *20*, 101–124. [[CrossRef](#)]
68. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
69. European Society of Radiology (ESR). What the Radiologist Should Know about Artificial Intelligence—An ESR White Paper. *Insights Imaging* **2019**, *10*, 44. [[CrossRef](#)] [[PubMed](#)]
70. Wang, H.; Zheng, B.; Yoon, S.W.; Ko, H.S. A Support Vector Machine-Based Ensemble Algorithm for Breast Cancer Diagnosis. *Eur. J. Oper. Res.* **2017**, *267*, 687–699. [[CrossRef](#)]
71. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.Z. XAI—Explainable Artificial Intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.