Article

Lightweight Low-Rank Adaptation Vision Transformer Framework for Cervical Cancer Detection and Cervix Type Classification

Zhenchen Hong [1],*,† , **Jingwei Xiong [2],*,†**, **Han Yang [3]** and **Yu K. Mo [4],5**

1   Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA
2   Graduate Group in Biostatistics, University of California, Davis, CA 95616, USA
3   Department of Chemistry, Columbia University, New York, NY 10027, USA; hy2565@columbia.edu
4   Department of Computer Science, Indiana University, Bloomington, IN 47405, USA; moyu@iu.edu
5   Department of Biology, Indiana University, Bloomington, IN 47405, USA
*   Correspondence: zhenchen.hong@email.ucr.edu (Z.H.); jwxxiong@ucdavis.edu (J.X.)
†   These authors contributed equally to this work.

Abstract: Cervical cancer is a major health concern worldwide, highlighting the urgent need for better early detection methods to improve outcomes for patients. In this study, we present a novel digital pathology classification approach that combines Low-Rank Adaptation (LoRA) with the Vision Transformer (ViT) model. This method is aimed at making cervix type classification more efficient through a deep learning classifier that does not require as much data. The key innovation is the use of LoRA, which allows for the effective training of the model with smaller datasets, making the most of the ability of ViT to represent visual information. This approach performs better than traditional Convolutional Neural Network (CNN) models, including Residual Networks (ResNets), especially when it comes to performance and the ability to generalize in situations where data are limited. Through thorough experiments and analysis on various dataset sizes, we found that our more streamlined classifier is highly accurate in spotting various cervical anomalies across several cases. This work advances the development of sophisticated computer-aided diagnostic systems, facilitating more rapid and accurate detection of cervical cancer, thereby significantly enhancing patient care outcomes.

Keywords: cervical cancer; detection; vision transformer (ViT); low-rank adaptation (LoRA); cervix type classification; deep learning; computer-aided diagnosis

1. Introduction

Cervical cancer is a significant public health concern, ranking as the fourth most prevalent cancer among women globally. It trails only breast, colorectal, and lung cancers in incidence, with over 500,000 new cases reported each year [1–7]. Even more alarming are the stark geographical disparities that exist in the global burden of cervical cancer, which reflect significant inequalities in access to preventive measures and healthcare services, reflecting the availability, coverage, and quality of preventive strategies, as well as the prevalence of risk factors. Approximately 90% of cervical cancer deaths occur in low- and middle-income countries (LMICs), underscoring the pressing need for improved access to effective diagnostic and treatment options in these regions [2,3]. In developing countries, women often face numerous barriers to accessing adequate cervical cancer screening programs. These obstacles include the high costs associated with regular examinations, limited awareness about the importance of screening, and insufficient access to medical facilities. As a result, individual patients in these regions are at a considerably higher risk of developing cervical cancer compared to those in more developed nations [8]. Implementing effective screening strategies can significantly reduce deaths caused by cervical cancer.

Effective cervical cancer screening strategies have demonstrated a remarkable impact on reducing the lifetime risk of developing the disease. Studies have shown that these interventions can decrease the risk by a substantial 25% to 36%. Moreover, cervical cancer screening proves to be highly cost-effective, with estimates suggesting that the cost per year of life saved is less than $500. This highlights the significant public health benefits and economic value of implementing comprehensive cervical cancer screening programs [9]. As modern medical and computer technologies rapidly advance, numerous screening and diagnostic approaches now rely on computer-aided detection (CAD) architectures [10]. The importance of early detection and surgical intervention in the treatment of cervical cancer cannot be overstated, given that the disease often presents no symptoms in its initial stages. There are several methodologies employed for the early detection of cervical cancer, each with its own set of advantages. For example, colposcopy offers a direct visual inspection of the cervix through a specialized magnifying device, allowing for the identification of visible abnormalities or lesions that may suggest the presence of cervical cancer [11–13]; Also, the Papanicolaou (Pap) test or biopsy alongside human papillomavirus infection (HPV) typing tests, plays a crucial role in early diagnosis by examining cervical cells under a microscope to detect precancerous or cancerous modifications [14–17]. Moreover, biomarker testing on tumor samples yields insights into the tumor's specific characteristics, aiding in the tailoring of treatment strategies. The advent of sequencing techniques over the past decade, particularly in analyzing HPV genotypes and cervical cancer [18–22], has marked a significant advance. These techniques enable the identification of genetic alterations linked to cervical cancer by detecting the virus and a wide array of HPV genotypes in clinically challenging samples, and then the genotype or evolutionary analysis [23–27] enhances our understanding of the disease's genetic landscape, offering deeper insights into its mechanisms.

While the mentioned diagnostic modalities offer the potential for early-stage cervical cancer detection—thereby improving treatment efficacy, patient outcomes, and survival rates—certain methodologies present limitations in terms of cost-effectiveness and efficiency [9,28–30]. Specifically, the procurement of sequencing data for each patient needs substantial financial investment, and the evolutionary analysis of cancer samples demands significant time [31–35]. In contrast, cervix image screening via colposcopy emerges as a more cost-efficient and time-effective strategy [36–38], enabling the identification of precancerous changes or early-stage cancerous developments. Nevertheless, the early detection of cervical cancer remains a complex challenge. This complexity is compounded by instances of low-quality cervix image screening samples and the presence of subtle abnormalities that are challenging to discern, particularly during the disease's initial stages. For example, the quality of cervix screening images varies due to discrepancies in collection methodologies, sample composition, and processing techniques, presenting a challenge in obtaining uniformly high-quality diagnostic images. And the morphological changes indicative of early-stage cervical cancer in cervix images are often subtle and difficult to differentiate from normal images, complicating the task of early detection. Also, the subjective evaluation of cervix image screening outcomes by pathologists can introduce variability in the identification of cervical abnormalities, underscoring the need for standardized interpretation frameworks. Lastly, while screening tests are designed to be sensitive, the occurrence of false-negative results is not uncommon, potentially delaying cervical cancer diagnosis and adversely affecting patient outcomes.

Despite the availability of extensive qualified datasets, the evaluation of cervix types—Type 1 Cervical Intraepithelial Neoplasia (CIN), Type 2 Squamous Intraepithelial Lesion (SIL), and Type 3 Dysplasia—from cervix screening images remains a complex and time-intensive task that necessitates the acumen of experienced clinicians. The difficulty in discerning between these cervix types from screening images is notably exacerbated by the intrinsic limitations of such images and the complex nature of the morphological changes within cervix tissue structures. In some recent studies, the results of colposcopy examinations are not consistently reproducible or precise. For one instance, false-negative rates

vary widely, ranging from 13% to 69%, due to discrepancies in physician expertise and the region of the sample being examined [39]. And other studies also reported false-negative rates ranging from 25% to 57% specifically for biopsy samples identified as positive during colposcopy examinations [40]. Nonetheless, recent innovations in imaging technologies, including high-resolution microscopy and advanced digital imaging systems, have significantly improved the quality and definition of cervix screening images, thus aiding in the identification of nuanced abnormalities [41–45]. Furthermore, the progress in digital pathology platforms has enabled the digitization of histological slides, which supports remote access, facilitates image analysis, and promotes computer-aided diagnosis, thereby enhancing operational efficiency and encouraging collaboration among medical professionals. From an analytical standpoint, machine learning and deep learning models, trained on comprehensive datasets of cervix screening images, show promise in autonomously detecting and categorizing abnormal cervix types. The efficacious deployment of these sophisticated computer vision models underscores their wider applicability in biomedical research, heralding a new era of automated and precise analysis across various experimental paradigms [18,46–51]. These algorithms excel in identifying complex patterns and features imperceptible to the human eye, thereby improving the precision and reliability of cervical cancer screening efforts. Such technological advancements in artificial intelligence have markedly augmented the capabilities in disease detection, with computer-aided and AI-based methodologies revolutionizing the diagnostic processes for cervical cancer.

However, training advanced deep learning models on medical data, which is often constrained in volume, poses substantial challenges. These models, with millions of parameters, demand a considerable dataset to avert over-fitting and guarantee effective generalization. Nevertheless, the inherently limited scale of medical datasets—stemming from privacy issues, data collection hurdles, and ethical restrictions—renders the direct application of complex CNN frameworks on such data frequently untenable, potentially resulting in sub-optimal outcomes. To address the challenges of cervix type classification from cervix screening images using advanced CNN models, this study introduces a pioneering approach in digital pathology classification. It incorporates Low-Rank Adaptation (LoRA) into the Vision Transformer (ViT) architecture, designed to enhance the precision of cervix type categorization. Our methodology leverages LoRA to enable efficient model training on constrained datasets, thus exploiting the sophisticated visual representation prowess of Vision Transformers. In comparison with conventional models from the recent decade, such as VGG, GoogLeNet, ResNet, DenseNet and ResNeXt [4,52–62], our strategy achieves enhanced performance and exhibits remarkable generalization capabilities, especially in scenarios marked by limited data. Extensive experimentation and analytical scrutiny on benchmark datasets validate the efficacy of our integrated ViT with LoRA model in accurately identifying cervical cancer markers. This study marks a significant leap forward in advancing computer-aided diagnosis systems, paving new paths for the early detection and management of cervical cancer.

2. Review of Study

The availability of larger and more diverse datasets containing cervix images from varied populations and stages of disease has notably improved the performance of deep learning-based classifiers, simultaneously mitigating the risk of model over-fitting to specific traits of the training data. Nevertheless, the majority of existing literature emphasizes the accuracy and various other performance metrics on training sets, with scant attention paid to disclosing model performance on validation or test datasets as detailed in Table A1. Moreover, research involving patient biomedical images often encounters constraints in revealing actual data and trained models, thereby limiting opportunities for external validation of the models and data described in these studies [63–65]. Upon reviewing publicly available datasets, it is common to find that while training sets display high accuracy and other metrics, the performance on validation and test sets typically yields satisfactory yet variable outcomes.

The domain of cervical cancer detection has witnessed remarkable advancements through the adoption of deep learning and computer vision, offering a range of innovative strategies to improve diagnostic accuracy. Early research efforts [66] employed machine learning algorithms, notably K-NN, to distinguish between normal and pathological cervical tissues, yielding promising results in sensitivity and specificity. Subsequent studies [67] explored the utility of deep learning further in enhancing diagnostics for cervical cancer, including the development of automatic segmentation techniques for the cervical region and evaluating the performance of deep learning approaches against traditional methods, such as Pap smears.

A pivotal development in this field is the creation of the "Colposcopy Ensemble Network" (CYENET) [68], which applies a deep learning framework for the classification of colposcopy images into distinct categories to aid in cervical cancer detection. Trained on an extensive screening dataset, CYENET has exceeded the accuracy of established models like VGG16 and VGG19. Additionally, the ongoing investigation into deep convolutional neural networks (DCNNs) with a variety of optimizers signals a persistent effort to refine the accuracy in differentiating between benign and cancerous cervical images. The emergence of computer-aided diagnosis (CAD) systems, such as "CerCan·Net" [69], and novel approaches to image size optimization [70], mark significant strides towards leveraging deep learning in cervical cancer screenings. These advancements not only underscore the vast potential of deep learning in medical imaging but also pave the way for future research focused on optimizing neural networks for enhanced diagnostic accuracy and integrating machine learning innovations to improve outcomes in patient care.

3. Materials and Methods

3.1. Images Acquisition

MobileODT has implemented a Quality Assurance workflow to support remote supervision, enhancing the decision-making process for healthcare providers in rural areas. Improving this workflow to facilitate real-time assessments of patient treatment eligibility based on cervix type would significantly contribute to the early detection of cervical cancer. In a collaborative effort, MobileODT and Intel launched a classification contest on Kaggle [71]. This competition invites participants to develop an algorithm capable of accurately determining a woman's cervix type from images, aiming to minimize ineffective treatments and ensure that patients receive the correct referrals for more specialized care if necessary.
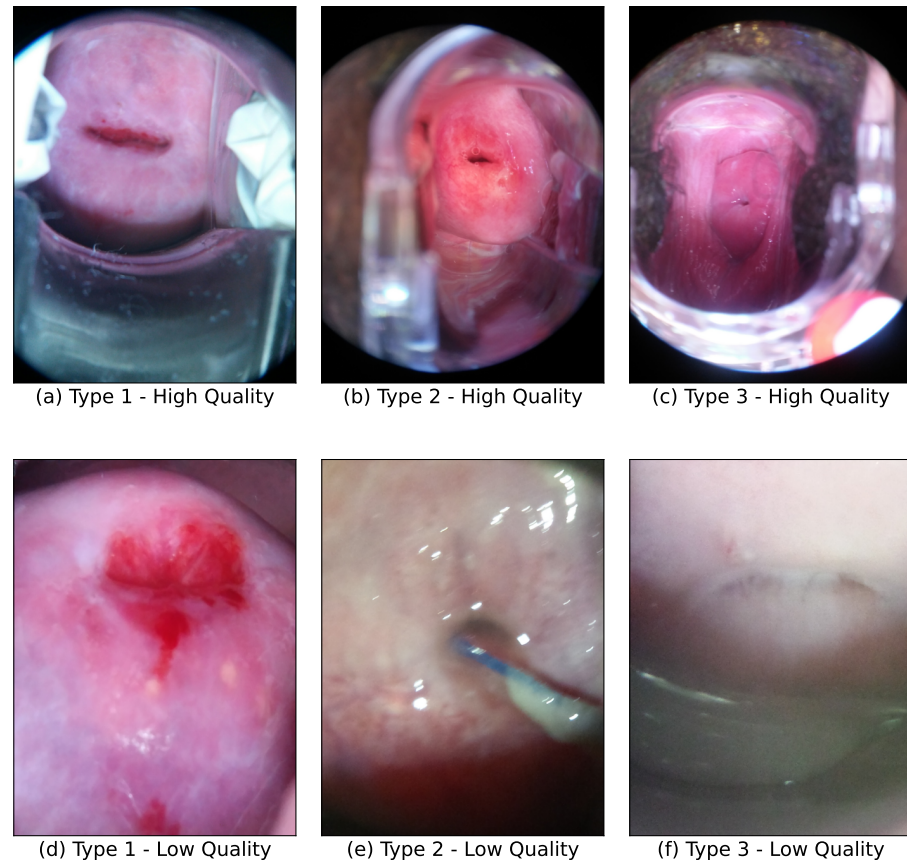
In a study involving 218,847 women in the older age group and 445,382 in the younger age group, researchers discovered a low incidence rate of cervical cancer during screening for Type 1 cervical intraepithelial neoplasia [72]. However, regular follow-up and monitoring are still crucial for women diagnosed with Type 1, as these lesions can potentially progress to higher grades of Type 1, which have a greater risk of developing into cervical cancer. Individuals with Type 2 and Type 3 cervixes require more extensive screening procedures. Detailed information about the distribution of cervix screening images in this dataset is provided in Table 1, and examples of these images are illustrated in Figure 1.

Table 1. A breakdown table of the MobileODT dataset.

Dataset	Type 1	Type 2	Type 3
Training set [1]	250	781	450
Testing set [2]	87	265	160

This set is optimized to speed up the training process [1]; Hold out data [2].

Figure 1. Sample dataset image. (**a**–**c**) are sample images of different cervix types with high quality; (**d**–**f**) are sample images of different cervix types with lower quality.

3.2. Cervix Type Classification Benchmarking Models

3.2.1. AlexNet [73]

Introduced in 2012, AlexNet [73] marked a pivotal moment in the field of computer vision, scoring the top spot in the ImageNet dataset. This architecture, featuring eight layers including convolutional, max-pooling, and fully connected layers, incorporates novel elements such as rectified linear units (ReLUs), dropout regularization, and GPU acceleration. These innovations not only cut down training time but also set a new benchmark for subsequent neural network models, catalyzing a wave of advancements in deep learning.

3.2.2. GoogLeNet [74]

GoogLeNet [74], also known as Inception-v1, introduced the inception module to the CNN landscape. This module supports the parallel use of various convolutional filter sizes within the same layer, balancing performance with computational efficiency. With 22 layers, including inception modules and a global average pooling strategy, GoogLeNet introduced auxiliary classifiers to mitigate the vanishing gradient problem, demonstrating high accuracy in image classification tasks and influencing future CNN designs.

3.2.3. VGG (Visual Geometry Group) [75]

The VGG model [75], developed by the University of Oxford's Visual Geometry Group in 2014, is celebrated for its straightforward yet effective architecture. With its series of convolutional layers followed by max-pooling, VGG exemplifies how deep and uniform structures can capture complex hierarchical features, thereby achieving remarkable accuracy in image classification challenges.

3.2.4. ResNet [76]

ResNet [76], introduced in 2015, marked a turning point in deep learning by innovatively addressing the vanishing gradient problem with residual learning. It incorporates skip connections that directly add inputs to outputs, allowing for the seamless training of networks that are significantly deeper than previously possible. This design enables the network to learn residual functions with ease, ensuring that deeper network layers can learn identity functions as a default, thereby preventing the degradation problem. The widespread adoption of ResNet across various computer vision applications can be attributed to its remarkable efficiency in learning hierarchical features, facilitating advancements in deep neural network architectures and making it a foundational model in the field.

3.2.5. DenseNet [77]

DenseNet [77], presented in 2017, offered a solution to the vanishing gradient problem by promoting feature reuse through its dense connectivity pattern, wherein each layer is connected to every other layer in a feed-forward fashion. With dense blocks and transition layers that manage parameter size, DenseNet achieves superior performance on image classification tasks, optimizing efficiency.
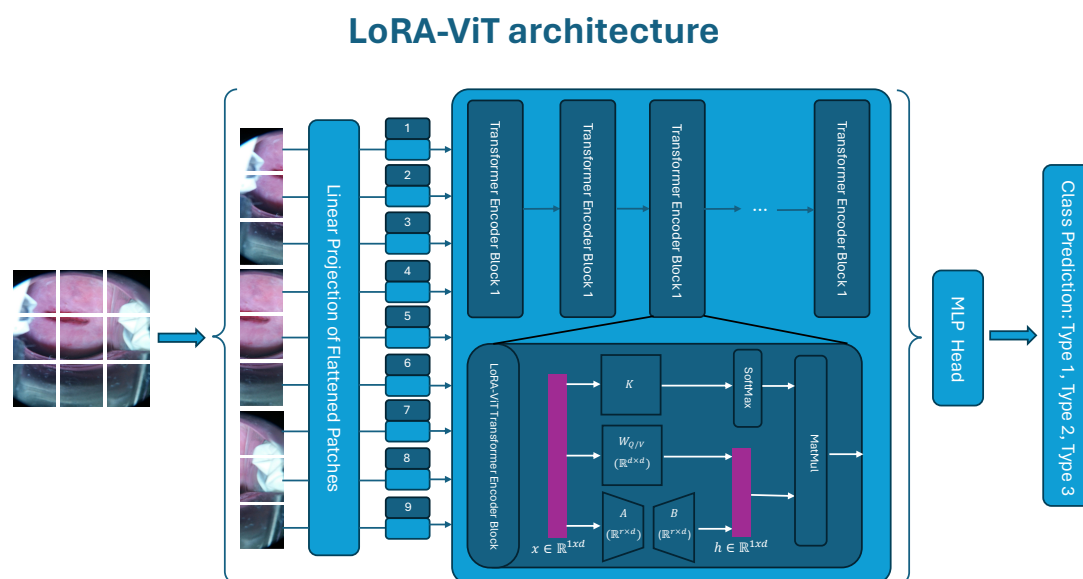
3.2.6. ResNeXt [78]

Building on the successes of ResNet, ResNeXt [78] was introduced in 2017, presenting a novel way to increase the model's capacity and performance without a substantial increase in complexity. The key innovation of ResNeXt lies in its use of "cardinality", a dimension that represents the number of independent paths within the network. This concept allows ResNeXt to capture a wide array of features by aggregating transformations from multiple paths, effectively increasing the network's robustness and efficiency without the need for a proportional rise in the parameters or computational demand.

3.3. A Cervix Type Classification Pipeline

The system flow diagram showcasing the proposed method for cervix type is illustrated in Figure 2: The architecture commences by dividing an input cervical image into patches of fixed size. Each patch undergoes linear embedding, with positional embeddings added to retain spatial information. These embedded vectors are then inputted into a standard Transformer encoder, modified to integrate Low-Rank Adaptation (LoRA). Within each self-attention layer of the Transformer encoder, low-rank decomposition matrices (referred to as A and B) are introduced into the fixed pre-trained query (W_Q) and value (W_V) projection matrices. This facilitates efficient adaptation of the pre-trained Vision Transformer to the specific task of cervical cancer classification, while preserving the majority of the model's parameters. The LoRA-ViT architecture facilitates the effective learning of task-specific representations with minimal computational overhead and mitigates the risk of over-fitting.

3.4. Low-Rank Adaptation (LoRA) and Vision Transformer (ViT)

Vision Transformer (ViT) [79] is a novel deep learning architecture that adapts the Transformer model, initially designed for natural language processing, to tasks in computer vision, notably image classification. ViT introduces a groundbreaking method by treating images as sequences of patches, mirroring the tokenization of words in language processing. This innovative approach revolutionizes image processing, leveraging the Transformer's strengths in capturing complex relationships within sequences for improved vision tasks.

LoRA-ViT architecture

Figure 2. Schematic of the LoRA-ViT architecture for cervical cancer classification. The process begins by segmenting an image into fixed-size patches, linearly embedding each patch, adding positional embeddings, and then inputting the vectors into a standard Transformer encoder. The Transformer encoder is modified by incorporating low-rank decomposition matrices (denoted as A and B), which are injected into the fixed pre-trained query (W_Q) and value (W_V) projection matrices of each self-attention layer.

While ViT models stand out for their remarkable accuracy and enhanced generalizability across various tasks [80], their application to cervical cancer classification is fraught with challenges, especially within clinical settings. This is primarily due to the fact that ViT models, with their transformer architecture, are significantly larger in terms of parameter count compared to previous CNN-based models. This architecture demands a much larger dataset and longer training times for full parameter training compared to CNNs. When it comes to the task of classifying cervical cancer images, the available datasets are exceptionally limited (fewer than 800 images per class) compared to the vast ImageNet dataset [81] typically used to train ViTs, potentially leading to training failures under full parameter training conditions. Even if a sufficient number of images are available for training, the clinical imperative for efficient storage space utilization, minimal GPU resource use, and rapid processing of medical images adds another layer of complexity to the deployment of ViT models in this context.

Given the challenges of the extensive parameter size and the high data and training requirements when applying ViT models to cervical cancer classification, there emerges a compelling need for innovative training methods. Low-Rank Adaptation (LoRA) [82] by Microsoft provides an ingenious solution to this dilemma by adapting pre-trained vision models for use in robust cervical cancer detection systems without the need for complete fine-tuning. This method involves locking the weights of the pre-trained model and integrating trainable rank decomposition matrices into each layer of the Transformer architecture. By doing so, LoRA dramatically reduces the number of trainable parameters during the fine-tuning process. This reduction not only makes the training process more feasible with the datasets typical of medical imaging but also ensures efficient use of storage and GPU resources, aligning with the critical clinical requirements for space and speed [83].

In this study, we focus on the fact that the LoRA approach to training with limited data does not compromise on performance; it maintains or even enhances the model's effectiveness compared to full-parameter training. In scenarios where extensive data are not available, which is often the case in medical contexts, the LoRA methodology is particularly

advantageous. It provides a path to leverage the superior capabilities of ViT models over traditional CNNs for medical image classification tasks, such as cervical cancer detection, without the extensive resource commitments typically associated with these models. LoRA thus stands out not just for its efficiency and reduced computational demands but for maintaining high model quality, achieving a balance that is critically needed in the medical imaging domain.

As for implementation of the method architecture, we incorporate LoRA weights into each self-attention layer of a pre-trained ViT. During fine-tuning, the pre-trained query W_Q and value projection matrices W_V in a self-attention layer have their updates restricted by the introduced LoRA weights. These weights represent them with a low-rank decomposition and are expressed as the following:

$$h = W_0 x + \Delta W x = W_0 x + BAx \tag{1}$$

where $x \in R^{1 \times d}$ is the input, and $h \in R^{1 \times d}$ is the output features. Two low-rank matrices, $B \in R^{d \times r}$ and $A \in R^{r \times d}$, compose the weight change ΔW. At the onset of training, we employ a random Gaussian initialization for matrix A and initialize matrix B with zeros. Consequently, the product of matrices B and A, denoted as ΔW, is zero initially. The ranks r of these low-rank matrices are much smaller than the model dimension d, and we empirically set $r = 8$ in our experiments. Usually, r should not be larger than 8, because the low-rank matrix amplification ability will be impaired when the rank is 64 in the experiment.

With this proposed model and pipeline architecture, our LoRA-based ViT classifier provides a solution by offering more accurate predictions of cervix types within a significantly shorter training period compared to the original ViT and other popular deep learning neural network models.

3.5. Performance Methods

The performance of the classification models was assessed using several objective evaluation metrics: accuracy, precision, recall, F1 score and Matthew Correlation Coefficient (MCC). These metrics rely on the true-positive (TP), true-negative (TN), false-negative (FN), and false-positive (FP) values of the models' predictions for each cervix type included in the confusion matrix:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \tag{6}$$

4. Results

4.1. Overall and Trainable Parameters

The landscape of deep learning models is marked by a rich diversity of architectures, each uniquely designed to address specific challenges in computer vision. State-of-the-art models such as ResNet, DenseNet, VGG, and GoogLeNet have gained widespread acclaim for their innovative methodologies and advanced architectures, establishing them as leading choices for computer-aided diagnosis applications of cervical cancer classification in recent years, details seen in the session of "Review of study". These models have revolutionized the field of deep learning by introducing novel concepts and techniques that address

critical challenges in medical image analysis. Despite the advancements these architectures offer, they share a common challenge: the increasing number of trainable parameters as models grow in complexity. This surge in parameters escalates the computational demands, affecting both training and inference phases. Therefore, finding a balance between model complexity and performance becomes paramount. Adopting optimization strategies such as parameter sharing, model pruning, and low-rank factorization is essential to manage these computational demands efficiently. For detailed insights into the dimensions of neural networks, including batch sizes and trainable parameters, Table 2 provides a comprehensive overview. The bold values are the best performance in the comparison across this paper.

Our method distinguishes itself by having the lowest count of trainable parameters among the leading deep learning architectures, with only 0.15 million trainable parameters. This figure is notably less than the 1% of parameters in the least complex model mentioned in our comparison. This dramatic reduction in trainable parameters enhances the efficiency of the training process and computational resource utilization across a wide array of datasets. Furthermore, when integrating Low-Rank Adaptation (LoRA) with the Vision Transformer (ViT), the model requires less than 0.2% of the trainable parameters found in the original ViT-base and merely 0.05% in ViT-huge. Although ViT demonstrates superior performance over all other models discussed, its substantial computational demands make the training process less feasible, particularly with limited datasets. Therefore, our ViT classifier, augmented with LoRA, presents a viable solution by enabling more precise cervix type predictions in significantly reduced training times compared to both the original ViT and other prevalent deep learning neural network models. By incorporating an optimized data loader and object-oriented programming-based image manipulation pipeline, our model achieves faster convergence rates with fewer trainable parameters and optimized architecture in Table A2, demonstrating the effectiveness of these optimizations in accelerating the model's convergence. The systematic approach to image manipulation, coupled with the modular design of the object-oriented-based processing pipeline, enables seamless integration with our model architecture. This facilitates efficient experimentation with various pre-processing techniques and parameter configurations, empowering us to fine-tune the data processing pipeline for optimal model performance. Furthermore, the refactored data loader results in a more structured and manageable code base, contributing to the overall efficiency of the data-processing tasks.

Table 2. Training parameters comparison.

Model	Number of Trainable Parameters	Model	Number of Trainable Parameters
Ours	**0.15 M**	AlexNet	60.0 M
ResNeXt-50	25.0 M	ViT-base	82.0 M
ResNet-50	25.6 M	ResNeXt-101	88.8 M
DenseNet	33.0 M	VGG-19	144 M
ResNet-101	44.6 M	ViT-huge	289.5 M

All the models listed above: Batch size = 32, Learning Rate by scheduler = CosineAnnealingLR (optimizer, cfg.epochs, 3×10^{-6}) with initial rate as 1×10^{-3}.

4.2. Confusion Matrix of the Prediction Results

A confusion matrix for classification with three classes organizes model predictions into a grid with rows representing the actual classes and columns representing the predicted classes. Each cell in the matrix shows the count (or proportion) of instances for a specific combination of actual and predicted classes. As in Figure 3, our method performs well in the classification task of three cervix types with limited data. More than 75% of images are classified correctly as the corresponding cervix type (54.0% Type 1 accuracy, 80.6% Type 2 accuracy, and 76.9% Type 3 accuracy), which is much better than the other classifiers on the

same dataset [84] with only 37.9% overall accuracy (33.0% Type 1 accuracy, 47.8% Type 2 accuracy, and 35.9% Type 3 accuracy).

Figure 3. Confusion matrix. There are three different cervix types (Type 1, Type 2 and Type 3) in the matrix. (1) The 3 by 3 matrix at the top left corner is the confusion matrix of our proposed classifier. The first row is the counts of images for each category and the second row is the corresponding count percentage. The red values in the confusion matrix are the incorrect predictions, while the green ones are the correct predictions. (2) The last row and last columns and the aggregated statistics on the sample number and corresponding proportion. The white values are the sum counts of the predicted labels or actual labels on each row or column. The green values are the percentage of correct predictions, while the red ones are the percentage of incorrect predictions.
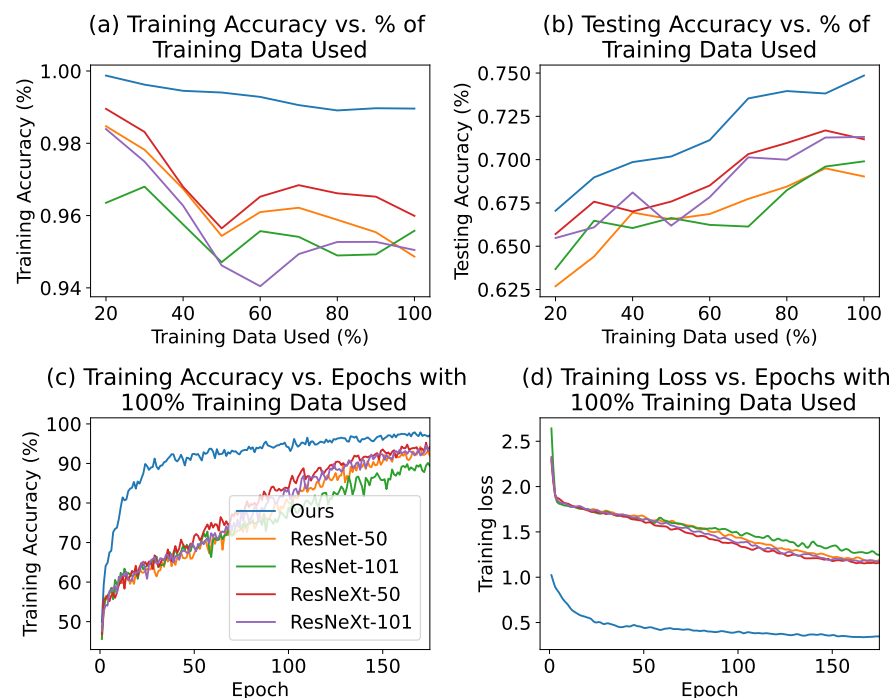
4.3. Preliminary Results, Accuracy and Loss

In classification tasks, the accuracy of both training and testing phases holds paramount importance in assessing the performance and generalization capability of classifier models. The training accuracy reflects how well the model learns from the provided data during the training phase, while the testing accuracy indicates how accurately the model can classify unseen data. In general, larger datasets tend to enhance model performance, assuming they encompass diverse and representative samples of the target population. However, the expansion of the dataset size can sometimes introduce a broader range of data quality issues. If left unaddressed, these issues may adversely affect the training accuracy. In the present study, we employed a ViT model pre-trained on the ImageNet-1K dataset [81], which consists of 1,281,167 training images. To emphasize the substantial difference in data size between our dataset, comprising only 1481 training images, and the ImageNet-1K dataset, this paper's dataset is with a "limited" size. In contrast, the term "100% training data" is used to denote the utilization of the entire available dataset during the model training process. To demonstrate the stability of the model when working with a limited dataset, we conducted an experiment by randomly selecting subsets of the data with varying proportions of the original data size. The detailed data split is presented in Table A3. As captured in Table 3 or Figure 4a, regardless of the amount of training data utilized, the proposed model in this paper always has the higher training accuracy than the other state-of-art models.

By evaluating the model's performance across these different data subsets, we aimed to assess its robustness and ability to maintain consistent results even when trained on reduced amounts of data. As illustrated in Figure 4a, we observed a decreasing trend in training accuracy with the enlargement of the training dataset. One potential remedy involves increasing the number of training epochs to accommodate the augmented data volume. Actually, upon closer examination of specific training data sizes, such as 50%, our proposed model demonstrates greater robustness compared to other models, exhibiting only minor differences in accuracy despite having the same increase in the amount of training images.

Table 3. Best accuracy and training loss of all experimental models with 100% training data used.

Model	Training Acc.	Testing Acc.	Training Loss
AlexNext	53.9%	52.5%	N/A
VGG-19	53.0%	52.9%	N/A
GooGleNet	96.8%	72.3%	N/A
DenseNet-121	97.3%	72.5%	N/A
ResNet-50	94.5%	68.9%	1.15
ResNet-101	96.0%	69.9%	1.11
ResNeXt-50	95.6%	71.1%	1.13
ResNeXt-101	94.9%	71.3%	1.14
Ours	**98.9%**	**75.0%**	**0.31**

Figure 4. Preliminary results. (**a**) The Training Accuracy vs. Percentage of Training Data Used curve is smoothed using a Gaussian kernel with a smoothness parameter of 0.8. (**b**) The Testing Accuracy vs. Percentage of Training Data Used curve is smoothed using a Gaussian kernel with a smoothness parameter of 0.5. (**c**) The Training Accuracy vs. Epochs curve, with 100% of Training Data Used, highlights our method's faster convergence speed. (**d**) The Training Loss vs. Epochs curve, with 100% of Training Data Used, demonstrates that our method achieves lower training loss.

Testing accuracy is paramount in assessing the efficacy and reliability of a classification model, where a higher testing accuracy underscores the model's capability to precisely classify unseen instances, thereby affirming its predictive prowess. As illustrated in Figure 4b, our model showcases its adeptness in identifying cervix types across diverse

training data volumes, with testing accuracy improving in tandem with the expansion of the training dataset. The minimal data size required for accurate prediction using ViT models is influenced by various factors, such as task complexity, dataset diversity, and the specific architecture and hyperparameters of the employed ViT model. While there is no universally accepted threshold for the minimal data size, it is generally recommended to utilize thousands to tens of thousands of training examples to achieve satisfactory performance with ViT models. Notably, we observed more stable performance when employing more than 70% of the available data as illustrated in Figure 4b, with around 1000 images. This finding suggests that, for the cervical colposcopy dataset used in this study and the proposed model architecture, setting the minimal data size threshold at 70% is likely to yield satisfactory results.

The velocity at which deep learning neural networks train is subject to variation, influenced by factors including model intricacy, hardware performance, and the implementation of algorithmic enhancements such as adjustments to batch size and learning rate. Our model is distinguished by its comparatively minimal trainable parameters relative to other models evaluated. This is evident in Figure 4c, where our approach reaches peak training accuracy in fewer epochs, thereby expediting model convergence. This efficiency allows for the effective use of early stopping strategies to fine-tune the model without over-fitting, with the detailed training time seen in Table A2.

Additionally, monitoring training loss is indispensable during model training. Training loss measures the discrepancy between the predicted outcomes by the model and the actual labels, serving as an indicator of the model's learning progress. A decreasing trend in training loss signifies the model's increasing accuracy in capturing underlying patterns of the data. Demonstrated by Figure 4d, our model records the most substantial reduction in training loss, highlighting its superior capacity to learn from training data effectively.

4.4. Other Related Performance Metrics

In medical imaging classification tasks, achieving high accuracy is crucial for precise diagnostic outcomes. Equally critical is the challenge presented by imbalanced datasets, characterized by a dominant prevalence of one class over others. To comprehensively evaluate a model's performance in navigating such imbalanced datasets, it is important to consider evaluation metrics beyond mere accuracy, including precision, recall, F1 score, and MCC, ensuring a thorough assessment of model performance in handling imbalanced data. By showcasing various signature performance metrics in Figure 5, our method demonstrates superior performance across all proposed metrics.

Figure 5. The performance metrics for each model with 100% of the training data. Each cell indicates the score of a specific performance metric for a given model. Models with the highest performance scores are outlined in red. The scores are computed individually for each cervix type and then averaged, weighted by the number of true instances for each type.

4.5. Cross Validation

Cross validation is a crucial technique for evaluating the performance and generalization ability of deep learning models in cervical cancer detection using imaging classification [85–88]. By partitioning the dataset of cervical images into multiple subsets and iteratively training and testing the model on different combinations of these subsets, cross validation provides a more robust assessment of the model's performance in identifying precancerous and cancerous lesions compared to a single train–test split. To test model reliability, we applied 5-fold cross validation on the complete dataset with 80% training data and 20% validation data, and our model achieved the best average performance metrics in accuracy, weighted precision, weighted recall, weighted F1 score, and Matthews Correlation Coefficient among the conventional models tested Table 4. These results demonstrate the robustness of our model and its ability to consistently outperform other approaches, validating its reliability in detecting cervical abnormalities across diverse subsets of the data.

Table 4. Testing data performance metric with 5-fold cross validation.

Model	Accuracy	Precision [1]	Recall [1]	F1-Score [1]	MCC
ResNet-50	0.682 ± **0.006**	0.686 ± 0.010	0.682 ± **0.006**	0.666 ± 0.011	0.454 ± **0.011**
ResNet-101	0.686 ± 0.012	0.688 ± 0.014	0.682 ± 0.012	0.672 ± 0.016	0.461 ± 0.023
ResNeXt-50	0.694 ± 0.007	0.690 ± **0.007**	0.695 ± 0.007	0.685 ± **0.009**	0.481 ± 0.015
ResNeXt-101	0.695 ± 0.013	0.705 ± 0.020	0.695 ± 0.013	0.681 ± 0.017	0.480 ± 0.022
Ours	**0.734** ± 0.011	**0.734** ± 0.012	**0.734** ± 0.011	**0.724** ± 0.016	**0.549** ± 0.020

Weighted metric [1]; Each cell stands for (average metric ± standard deviation).

5. Discussion

State-of-the-art models for cervical type classification have proven to be effective, with deep learning models reaching accuracy levels comparable to those of junior and even senior colposcopists in specific classification tasks [64]. Although these models have shown considerable success in terms of training accuracy, a focus solely on this metric without adequate testing accuracy or performance evaluation—particularly in scenarios involving limited datasets—may not fully capture the model's efficacy. Moreover, as models grow in complexity with more layers and trainable parameters, the task of training them becomes increasingly demanding, especially in the context of limited available data for precise classification.

Pre-trained models often require fine-tuning to achieve accurate predictions on specific tasks. In our study, we investigated the performance of various non-pre-trained models in this paper, and compared them to the LoRA-ViT architecture. The results, as depicted in Figure 4, demonstrate that when properly tuned, pre-trained models exhibit superior performance compared to conventional CNN models. Importantly, our architecture not only highlights the advantages of pre-trained models but also emphasizes an efficient approach to fine-tuning them. Without appropriate fine-tuning, pre-trained models may yield similar or even inferior performance compared to non-pre-trained CNN models, potentially due to factors such as adversarial robustness [89]. By optimizing the fine-tuning process, our architecture accelerates training and enhances the accuracy of pre-trained models in typical scenarios. This finding underscores the importance of proper fine-tuning techniques when adapting pre-trained models to specific tasks, as it can significantly improve their performance and efficiency.

Our method overcomes existing challenges by achieving unparalleled accuracy on test datasets that have not been seen during training, regardless of the amount of data used for training. Furthermore, it surpasses current models, particularly in the demanding task of Type 1 cervical classification, through the incorporation of a robust evaluation framework that ensures both reliability and effectiveness in its performance assessment. Also, our method enables the efficient training of the transformer layer without necessitating extensive computational resources. For various downstream tasks, training low-rank matrices with a reduced parameter count is sufficient, facilitating the use of pre-trained

weights across different tasks and thereby simplifying the training process. This strategy significantly speeds up the training period, obviating the need for gradients of pre-trained weights and their optimizer states, which in turn, boosts training efficiency and reduces the demand on hardware. Furthermore, the integration of trained low-rank matrices with pre-trained weights merges multi-branch architectures into a singular streamlined branch, effectively eliminating inference latency. Despite comprising only 0.05% of the trainable parameters of the original ViT, our lightweight model demonstrates its capability to efficiently manage larger datasets and conduct neural network training for medical imaging tasks with exceptional efficiency. And with cross validation, it helps to mitigate over-fitting, which is particularly important in medical imaging tasks, where the available datasets may be limited, enables the selection of optimal hyperparameters that maximize the model's accuracy and sensitivity in detecting cervical abnormalities, and ensures that the models are robust and generalizable to diverse patient populations and imaging conditions encountered in real-world clinical settings.

Moving forward, our approach maintains compatibility with parameter-efficient fine-tuning methodologies, including Adapters [90,91], Prefix-Tuning [92,93], and Visual Prompt-Tuning [94,95], supported by dimension reduction techniques, such as some previous application of Uniform Manifold Approximation and Projection [96–100] in clustering, and auto-encoder and variational auto-encoder models. These strategies are designed to fine-tune a restricted subset of parameters, introduced in a gradual manner, thereby eliminating the necessity to adjust all parameters within a pre-trained model and capture the most salient features of the data while reducing noise and computational complexity. The harmonious integration with these techniques substantially refines the tuning process. It achieves this by diminishing the requirements for computational and storage resources, all while ensuring there is no disruption to the existing framework. Also, we aim to enhance Vision Transformer adapters to facilitate multi-task learning by encapsulating task-specific knowledge and relationships. This will allow the adapters to be generalized and applied to novel tasks and domains without requiring extensive retraining or fine-tuning, thereby improving the efficiency and flexibility of the model in handling diverse computer vision challenges [101,102].

However, several critical areas demand our focus moving forward. A paramount concern lies with the dataset issue, highlighting the necessity for not only larger but also higher-quality datasets. It is notable that within all models' classification outcomes, the accuracy for Type 1 significantly lags behind that of Types 2 and 3, suggesting potential quality issues with the Type 1 data. With higher-quality Type 1 data and labels, the results are bound to significantly improve. Furthermore, investigating larger model architectures could lead to significant enhancements in model generalizability, potentially increasing accuracy even with limited training datasets. However, such advancements would require more substantial computational resources. Pursuing these avenues is essential for improving the accuracy and reliability of medical imaging classification systems. To support the training of these more sophisticated models, there is also a pressing need for better hardware, particularly more powerful GPUs. Advances in hardware technology will not only expedite the training process but also enable the handling of more complex models and larger datasets with greater ease. These enhancements are vital for pushing the boundaries of what is currently achievable in medical imaging classification, paving the way for more accurate, efficient, and reliable diagnostic tools in the future.

Author Contributions: Conceptualization, Z.H. and J.X.; methodology, Z.H. and J.X.; software, J.X.; validation, Z.H., J.X., H.Y. and Y.K.M.; formal analysis, Z.H. and J.X.; investigation, Z.H. and J.X.; resources, J.X.; data curation, Z.H. and J.X.; writing—original draft preparation, Z.H.; writing—review and editing, Z.H., J.X., H.Y. and Y.K.M.; visualization, Z.H.; supervision, Z.H. and J.X.; project administration, Z.H. and J.X. All authors have read and agreed to the published version of the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

LoRA	Low-Rank Adaptation
ViT	Vision Transformer
MCC	Matthew Correlation Coefficient
ResNet	Residual Network
Pap	Papanicolaou
HPV	Human Papillomavirus Infection
CNN	Convolutional Neural Network
GPU	Graphics Processing Unit
ReLU	Rectified Linear Unit
CAD	Computer-Aided Diagnosis
TP	True Positive
TN	True Negative
FN	False Negative
FP	False Positive

Appendix A. Additional Tables

Table A1. A summary table of the related works on the cervix colposcopy image database.

Reference	Dataset	Methods	Metrics
Mustafa and Dauda [63]	Undisclosed dataset	CNN-Adam	Accuracy = 0.90 [1], 0.74 [2]
		CNN-SGD	Accuracy = 0.83 [1]
		CNN-RMSProp	Accuracy = 0.88 [1]
Liu et al. [64]	Undisclosed data	Clinic-based model	Accuracy = 0.68 [3]
			AUC = 0.7 [3]
			Specificity = 0.71 [3]
		ResNet-50	Accuracy = 0.8 [3]
			AUC = 0.95 [3]
			Accuracy = 0.88 [3]
			Specificity = 0.87 [3]

Table A1. *Cont.*

Reference	Dataset	Methods	Metrics
Peng et al. [65]	Undisclosed data	ResNet-50	Accuracy = 0.80 [1] Specificity = 0.78 [1]
		DenseNet121	Accuracy = 0.76 [1] Specificity = 0.77 [1]
		VGG16	Accuracy = 0.86 [1] Specificity = 0.90 [1]
Cruz et al. [103]	MobileODT	CXNN	Accuracy = 0.77 [1] Accuracy = 0.664 [2]
Gorantla et al. [88]	MobileODT	CervixNet	Accuracy = 0.97 [1] Specificity = 0.98 [1] Precision = 0.98 [1] F1-score = 0.97 [1]
Aina et al. [104]	MobileODT	AlexNet SqueezeNet	Accuracy = 0.63 [2] Accuracy = 0.63 [2]
Payette et al. [105]	MobileODT	Residual CNN-32L [4] Residual CNN-53L [4] CNN-32L-Drop [4] CNN-55L-Drop [4]	Accuracy = 0.58 [2] Accuracy = 0.55 [2] Accuracy = 0.56 [2] Accuracy = 0.50 [2]
Darwish et al. [84]	MobileODT	ViT ViT-SPT/LSA	Accuracy = 0.91 [1] Accuracy = 0.91 [1] Accuracy = 0.38 [3]

Training set [1]; Validation set [2]; Testing set [3]; L = Layer [4].

Table A2. Training time comparison with entire dataset.

Model	Training Time (mins/10 epochs)	Convergence Epochs	Converged Training Time (h)
Ours	33.8	**25**	**14.0**
ViT-Large	56.3	96	90.08
ResNet-50	33.7	185	103.9
ResNet-101	**30.3**	210	106.1
ResNeXt-50	32.1	162	86.67
ResNeXt-101	31.9	149	79.2

Table A3. A breakdown table of training data size.

% of Training Data	Type 1	Type 2	Type 3
20%	50	156	90
30%	75	234	135
40%	100	312	180
50%	125	390	225
60%	150	468	270
70%	175	546	315
80%	200	624	360
90%	225	702	405
100%	250	781	450

Appendix B. Additional Figures

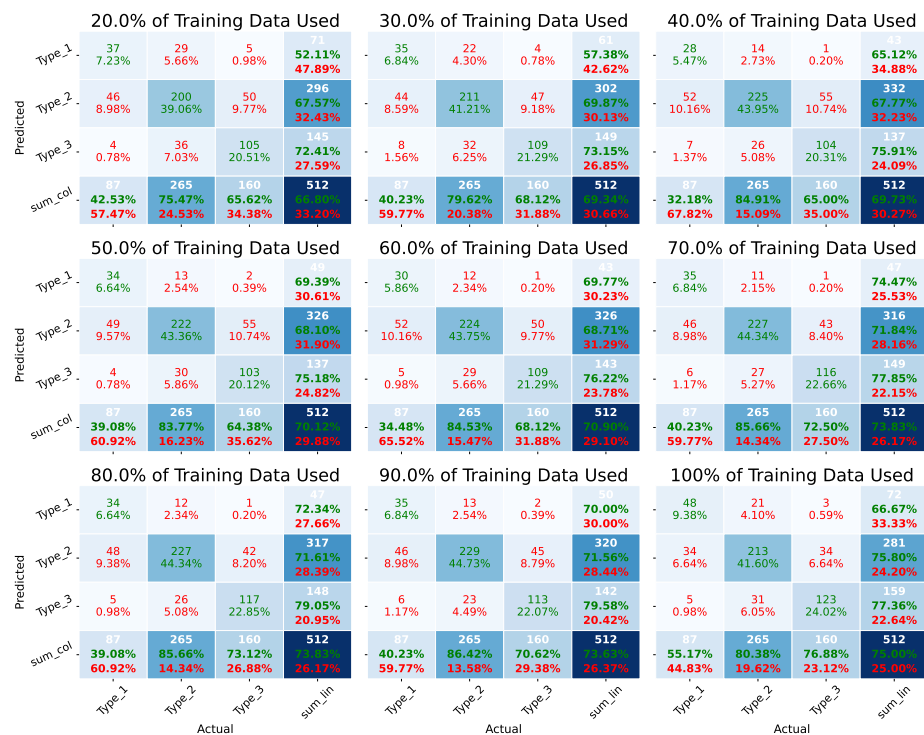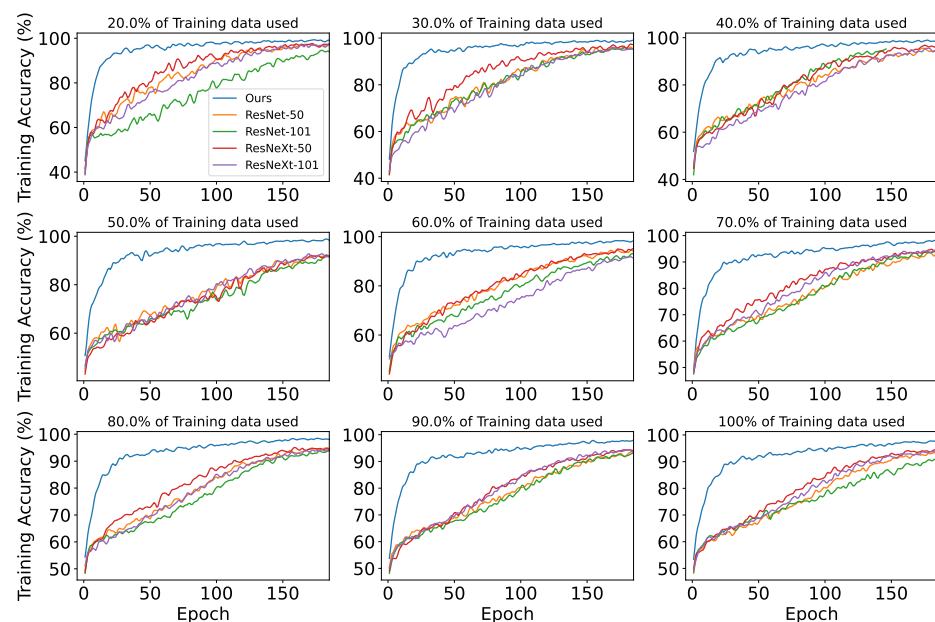**Figure A1.** Confusion matrix of our method with different % of training data used.
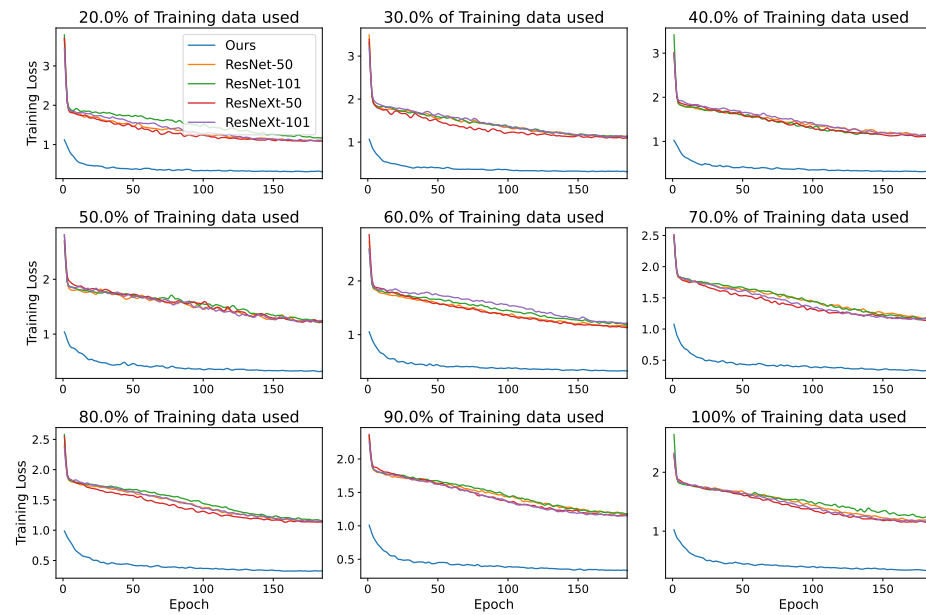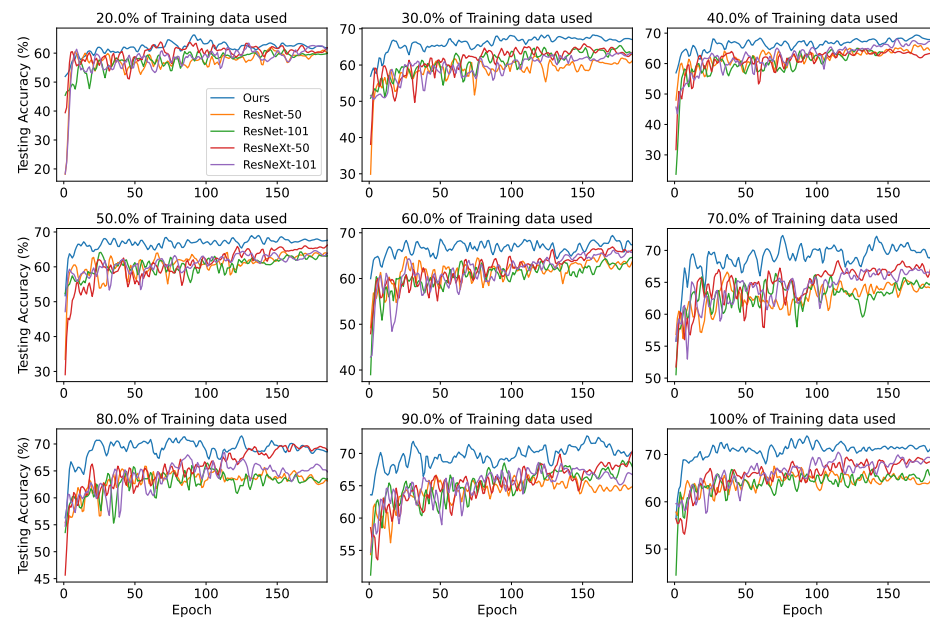

Figure A2. Training accuracy vs. epoch.

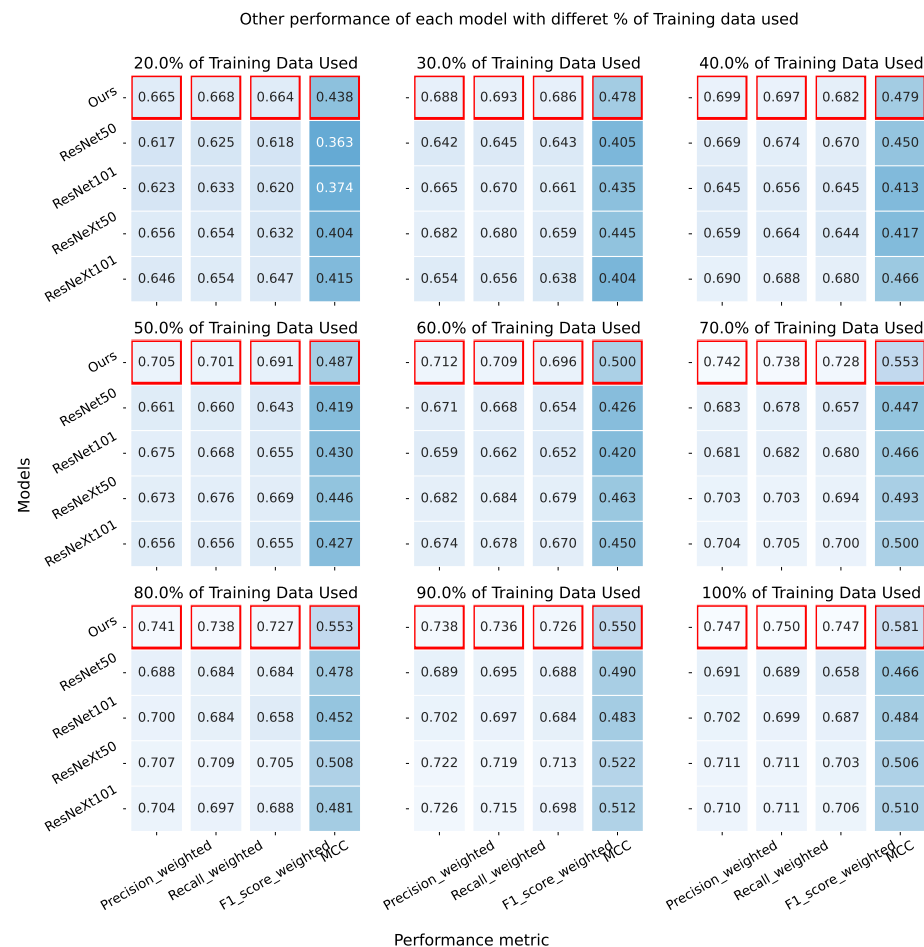*Appendix B.1. Training Loss vs. Epoch across Different % of Training Data Used*

Figure A3. Training loss vs. epoch.

Appendix B.2. Testing Accuracy vs. Epoch across Different % of Training Data Used

Figure A4. Testing accuracy vs. epoch.

Other performance of each model with differet % of Training data used

Figure A5. Other model performance metric with % training data used. Models with the highest performance scores are outlined in red. Color bar legend seen in Figure 5.

References

1. Pimple, S.; Mishra, G. Cancer cervix: Epidemiology and disease burden. *Cytojournal* **2022**, *19*, 21. [CrossRef] [PubMed]
2. 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]
3. Stelzle, D.; Tanaka, L.F.; Lee, K.K.; Khalil, A.I.; Baussano, I.; Shah, A.S.; McAllister, D.A.; Gottlieb, S.L.; Klug, S.J.; Winkler, A.S.; et al. Estimates of the global burden of cervical cancer associated with HIV. *Lancet Glob. Health* **2021**, *9*, e161–e169. [CrossRef] [PubMed]
4. 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]
5. Ginsburg, O.; Bray, F.; Coleman, M.P.; Vanderpuye, V.; Eniu, A.; Kotha, S.R.; Sarker, M.; Huong, T.T.; Allemani, C.; Dvaladze, A.; et al. The global burden of women's cancers: A grand challenge in global health. *Lancet* **2017**, *389*, 847–860. [CrossRef]
6. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
7. Jemal, A.; Bray, F.; Center, M.M.; Ferlay, J.; Ward, E.; Forman, D. Global cancer statistics. *CA A Cancer J. Clin.* **2011**, *61*, 69–90. [CrossRef] [PubMed]
8. Mehmood, M.; Rizwan, M.; Gregus ml, M.; Abbas, S. Machine learning assisted cervical cancer detection. *Front. Public Health* **2021**, *9*, 788376. [CrossRef] [PubMed]
9. Goldie, S.J.; Gaffikin, L.; Goldhaber-Fiebert, J.D.; Gordillo-Tobar, A.; Levin, C.; Mahé, C.; Wright, T.C. Cost-effectiveness of cervical-cancer screening in five developing countries. *N. Engl. J. Med.* **2005**, *353*, 2158–2168. [CrossRef] [PubMed]
10. Hajimirzaei, B.; Navimipour, N.J. Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm. *Ict Express* **2019**, *5*, 56–59. [CrossRef]
11. Jeronimo, J.; Schiffman, M. Colposcopy at a crossroads. *Am. J. Obstet. Gynecol.* **2006**, *195*, 349–353. [CrossRef] [PubMed]
12. Apgar, B.S.; Brotzman, G.L.; Spitzer, M. *Colposcopy E-Book: Principles and Practice*; Elsevier Health Sciences: Amsterdam, The Netherlands, 2008.

13. Cantor, S.B.; Cárdenas-Turanzas, M.; Cox, D.D.; Atkinson, E.N.; Nogueras-Gonzalez, G.M.; Beck, J.R.; Follen, M.; Benedet, J.L. Accuracy of colposcopy in the diagnostic setting compared with the screening setting. *Obstet. Gynecol.* **2008**, *111*, 7–14. [CrossRef] [PubMed]

14. Carpenter, A.B.; Davey, D.D. ThinPrep® Pap Test™: Performance and biopsy follow-up in a university hospital. *Cancer Cytopathol. Interdiscip. Int. J. Am. Cancer Soc.* **1999**, *87*, 105–112. [CrossRef]

15. Nanda, K.; McCrory, D.C.; Myers, E.R.; Bastian, L.A.; Hasselblad, V.; Hickey, J.D.; Matchar, D.B. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: A systematic review. *Ann. Intern. Med.* **2000**, *132*, 810–819. [CrossRef] [PubMed]

16. Mayrand, M.H.; Duarte-Franco, E.; Rodrigues, I.; Walter, S.D.; Hanley, J.; Ferenczy, A.; Ratnam, S.; Coutlée, F.; Franco, E.L. Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *N. Engl. J. Med.* **2007**, *357*, 1579–1588. [CrossRef]

17. Lytwyn, A.; Sellors, J.W.; Mahony, J.B.; Daya, D.; Chapman, W.; Ellis, N.; Roth, P.; Lorincz, A.T.; Gafni, A.; The HPV Effectiveness in Lowgrade Paps (HELP) Study No. 1 Group. Comparison of human papillomavirus DNA testing and repeat Papanicolaou test in women with low-grade cervical cytologic abnormalities: A randomized trial. *Cmaj* **2000**, *163*, 701–707. [PubMed]

18. Shen-Gunther, J.; Wang, Y.; Lai, Z.; Poage, G.M.; Perez, L.; Huang, T.H. Deep sequencing of HPV E6/E7 genes reveals loss of genotypic diversity and gain of clonal dominance in high-grade intraepithelial lesions of the cervix. *BMC Genom.* **2017**, *18*, 231. [CrossRef] [PubMed]

19. Grønhøj, C.; Jensen, D.H.; Agander, T.; Kiss, K.; Høgdall, E.; Specht, L.; Bagger, F.O.; Nielsen, F.C.; von Buchwald, C. Deep sequencing of human papillomavirus positive loco-regionally advanced oropharyngeal squamous cell carcinomas reveals novel mutational signature. *BMC Cancer* **2018**, *18*, 640. [CrossRef] [PubMed]

20. Liu, P.; Iden, M.; Fye, S.; Huang, Y.W.; Hopp, E.; Chu, C.; Lu, Y.; Rader, J.S. Targeted, deep sequencing reveals full methylation profiles of multiple HPV types and potential biomarkers for cervical cancer progression. *Cancer Epidemiol. Biomarkers Prev.* **2017**, *26*, 642–650. [CrossRef] [PubMed]

21. Arroyo Mühr, L.S.; Lagheden, C.; Lei, J.; Eklund, C.; Nordqvist Kleppe, S.; Sparén, P.; Sundström, K.; Dillner, J. Deep sequencing detects human papillomavirus (HPV) in cervical cancers negative for HPV by PCR. *Br. J. Cancer* **2020**, *123*, 1790–1795. [CrossRef] [PubMed]

22. Ai, W.; Wu, C.; Jia, L.; Xiao, X.; Xu, X.; Ren, M.; Xue, T.; Zhou, X.; Wang, Y.; Gao, C. Deep sequencing of HPV16 E6 region reveals unique mutation pattern of HPV16 and predicts cervical cancer. *Microbiol. Spectr.* **2022**, *10*, e01401-22. [CrossRef] [PubMed]

23. Long, N.P.; Jung, K.H.; Yoon, S.J.; Anh, N.H.; Nghi, T.D.; Kang, Y.P.; Yan, H.H.; Min, J.E.; Hong, S.S.; Kwon, S.W. Systematic assessment of cervical cancer initiation and progression uncovers genetic panels for deep learning-based early diagnosis and proposes novel diagnostic and prognostic biomarkers. *Oncotarget* **2017**, *8*, 109436. [CrossRef] [PubMed]

24. Wang, Y.; Liu, L.; Chen, Z. Transcriptome profiling of cervical cancer cells acquired resistance to cisplatin by deep sequencing. *Artif. Cells Nanomed. Biotechnol.* **2019**, *47*, 2820–2829. [CrossRef] [PubMed]

25. Juan, L.; Tong, H.l.; Zhang, P.; Guo, G.; Wang, Z.; Wen, X.; Dong, Z.; Tian, Y. Identification and characterization of novel serum microRNA candidates from deep sequencing in cervical cancer patients. *Sci. Rep.* **2014**, *4*, 6277. [CrossRef] [PubMed]

26. Hong, Z.; Barton, J.P. popDMS infers mutation effects from deep mutational scanning data. *bioRxiv* **2024**. [CrossRef] [PubMed]

27. Sohail, M.S.; Louie, R.H.; Hong, Z.; Barton, J.P.; McKay, M.R. Inferring epistasis from genetic time-series data. *Mol. Biol. Evol.* **2022**, *39*, msac199. [CrossRef] [PubMed]

28. Burmeister, C.A.; Khan, S.F.; Schäfer, G.; Mbatani, N.; Adams, T.; Moodley, J.; Prince, S. Cervical cancer therapies: Current challenges and future perspectives. *Tumour Virus Res.* **2022**, *13*, 200238. [CrossRef] [PubMed]

29. Goldie, S.J.; Kuhn, L.; Denny, L.; Pollack, A.; Wright, T.C. Policy analysis of cervical cancer screening strategies in low-resource settings: Clinical benefits and cost-effectiveness. *JAMA* **2001**, *285*, 3107–3115. [CrossRef] [PubMed]

30. Chauhan, A.S.; Prinja, S.; Srinivasan, R.; Rai, B.; Malliga, J.; Jyani, G.; Gupta, N.; Ghoshal, S. Cost effectiveness of strategies for cervical cancer prevention in India. *PLoS ONE* **2020**, *15*, e0238291. [CrossRef] [PubMed]

31. Coppola, L.; Cianflone, A.; Grimaldi, A.M.; Incoronato, M.; Bevilacqua, P.; Messina, F.; Baselice, S.; Soricelli, A.; Mirabelli, P.; Salvatore, M. Biobanking in health care: Evolution and future directions. *J. Transl. Med.* **2019**, *17*, 1–18. [CrossRef] [PubMed]

32. Damodaran, S.; Berger, M.F.; Roychowdhury, S. Clinical tumor sequencing: Opportunities and challenges for precision cancer medicine. *Am. Soc. Clin. Oncol. Educ. Book* **2015**, *35*, e175–e182. [CrossRef] [PubMed]

33. Safaeian, M.; Solomon, D.; Castle, P.E. Cervical cancer prevention—Cervical screening: Science in evolution. *Obstet. Gynecol. Clin. N. Am.* **2007**, *34*, 739–760. [CrossRef] [PubMed]

34. Bedell, S.L.; Goldstein, L.S.; Goldstein, A.R.; Goldstein, A.T. Cervical cancer screening: Past, present, and future. *Sex. Med. Rev.* **2020**, *8*, 28–37. [CrossRef] [PubMed]

35. Sun, M.; Gao, L.; Liu, Y.; Zhao, Y.; Wang, X.; Pan, Y.; Ning, T.; Cai, H.; Yang, H.; Zhai, W.; et al. Whole genome sequencing and evolutionary analysis of human papillomavirus type 16 in central China. *PLoS ONE* **2012**, *7*, e36577. [CrossRef]

36. Hou, X.; Shen, G.; Zhou, L.; Li, Y.; Wang, T.; Ma, X. Artificial intelligence in cervical cancer screening and diagnosis. *Front. Oncol.* **2022**, *12*, 851367. [CrossRef]

37. Gallay, C.; Girardet, A.; Viviano, M.; Catarino, R.; Benski, A.C.; Tran, P.L.; Ecabert, C.; Thiran, J.P.; Vassilakos, P.; Petignat, P. Cervical cancer screening in low-resource settings: A smartphone image application as an alternative to colposcopy. *Int. J. Women's Health* **2017**, *9*, 455–461. [CrossRef] [PubMed]

38. Xue, P.; Ng, M.T.A.; Qiao, Y. The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence. *BMC Med.* **2020**, *18*, 169. [CrossRef] [PubMed]

39. Khan, M.J.; Werner, C.L.; Darragh, T.M.; Guido, R.S.; Mathews, C.; Moscicki, A.B.; Mitchell, M.M.; Schiffman, M.; Wentzensen, N.; Massad, L.S.; et al. ASCCP colposcopy standards: Role of colposcopy, benefits, potential harms, and terminology for colposcopic practice. *J. Low. Genit. Tract Dis.* **2017**, *21*, 223–229. [CrossRef] [PubMed]

40. Baasland, I.; Hagen, B.; Vogt, C.; Valla, M.; Romundstad, P.R. Colposcopy and additive diagnostic value of biopsies from colposcopy-negative areas to detect cervical dysplasia. *Acta Obstet. Gynecol. Scand.* **2016**, *95*, 1258–1263. [CrossRef] [PubMed]

41. Sambyal, D.; Sarwar, A. Recent developments in cervical cancer diagnosis using deep learning on whole slide images: An Overview of models, techniques, challenges and future directions. *Micron* **2023**, *173*, 103520. [CrossRef] [PubMed]

42. Orfanoudaki, I.M.; Kappou, D.; Sifakis, S. Recent advances in optical imaging for cervical cancer detection. *Arch. Gynecol. Obstet.* **2011**, *284*, 1197–1208. [CrossRef] [PubMed]

43. Kundrod, K.A.; Smith, C.A.; Hunt, B.; Schwarz, R.A.; Schmeler, K.; Richards-Kortum, R. Advances in technologies for cervical cancer detection in low-resource settings. *Expert Rev. Mol. Diagn.* **2019**, *19*, 695–714. [CrossRef] [PubMed]

44. Richards-Kortum, R.; Lorenzoni, C.; Bagnato, V.S.; Schmeler, K. Optical imaging for screening and early cancer diagnosis in low-resource settings. *Nat. Rev. Bioeng.* **2024**, *2*, 25–43. [CrossRef]

45. Drezek, R.A.; Richards-Kortum, R.; Brewer, M.A.; Feld, M.S.; Pitris, C.; Ferenczy, A.; Faupel, M.L.; Follen, M. Optical imaging of the cervix. *Cancer Interdiscip. Int. J. Am. Cancer Soc.* **2003**, *98*, 2015–2027. [CrossRef] [PubMed]

46. Lee, J.G.; Jun, S.; Cho, Y.W.; Lee, H.; Kim, G.B.; Seo, J.B.; Kim, N. Deep learning in medical imaging: General overview. *Korean J. Radiol.* **2017**, *18*, 570. [CrossRef] [PubMed]

47. Razzak, M.I.; Naz, S.; Zaib, A. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps: Automation of Decision Making*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 323–350.

48. Alakwaa, W.; Nassef, M.; Badr, A. Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 409–417. [CrossRef]

49. Jia, A.D.; Li, B.Z.; Zhang, C.C. Detection of cervical cancer cells based on strong feature CNN-SVM network. *Neurocomputing* **2020**, *411*, 112–127.

50. Melekoodappattu, J.G.; Dhas, A.S.; Kandathil, B.K.; Adarsh, K. Breast cancer detection in mammogram: Combining modified CNN and texture feature based approach. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 11397–11406. [CrossRef]

51. Yu, H.; Xiong, J.; Ye, A.Y.; Cranfill, S.L.; Cannonier, T.; Gautam, M.; Zhang, M.; Bilal, R.; Park, J.E.; Xue, Y.; et al. Scratch-AID, a deep learning-based system for automatic detection of mouse scratching behavior with high accuracy. *eLife* **2022**, *11*, e84042. [CrossRef] [PubMed]

52. Sengupta, A.; Ye, Y.; Wang, R.; Roy, K. Going deeper in spiking neural networks: VGG and residual architectures. *Front. Neurosci.* **2019**, *13*, 425055. [CrossRef] [PubMed]

53. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 13733–13742.

54. Tammina, S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *Int. J. Sci. Res. Publ. (IJSRP)* **2019**, *9*, 143–150. [CrossRef]

55. Lin, H.; Hu, Y.; Chen, S.; Yao, J.; Zhang, L. Fine-grained classification of cervical cells using morphological and appearance based convolutional neural networks. *IEEE Access* **2019**, *7*, 71541–71549. [CrossRef]

56. Alyafeai, Z.; Ghouti, L. A fully-automated deep learning pipeline for cervical cancer classification. *Expert Syst. Appl.* **2020**, *141*, 112951. [CrossRef]

57. Allehaibi, K.H.S.; Nugroho, L.E.; Lazuardi, L.; Prabuwono, A.S.; Mantoro, T. Segmentation and classification of cervical cells using deep learning. *IEEE Access* **2019**, *7*, 116925–116941.

58. Targ, S.; Almeida, D.; Lyman, K. Resnet in resnet: Generalizing residual architectures. *arXiv* **2016**, arXiv:1603.08029.

59. Hussain, E.; Mahanta, L.B.; Das, C.R.; Talukdar, R.K. A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. *Tissue Cell* **2020**, *65*, 101347. [CrossRef] [PubMed]

60. Cao, L.; Yang, J.; Rong, Z.; Li, L.; Xia, B.; You, C.; Lou, G.; Jiang, L.; Du, C.; Meng, H.; et al. A novel attention-guided convolutional network for the detection of abnormal cervical cells in cervical cancer screening. *Med Image Anal.* **2021**, *73*, 102197. [CrossRef] [PubMed]

61. de Lima, C.R.; Khan, S.G.; Shah, S.H.; Ferri, L. Mask region-based CNNs for cervical cancer progression diagnosis on pap smear examinations. *Heliyon* **2023**, *9*, e21388. [CrossRef] [PubMed]

62. Zhang, X.; Cao, M.; Wang, S.; Sun, J.; Fan, X.; Wang, Q.; Zhang, L. Whole slide cervical cancer screening using graph attention network and supervised contrastive learning. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Vancouver, BC, Canada, 8–12 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 202–211.

63. Mustafa, S.; Dauda, M. Evaluating convolution neural network optimization algorithms for classification of cervical cancer macro images. In Proceedings of the 2019 15th International Conference on Electronics, Computer and Computation (ICECCO), Abuja, Nigeria, 10–12 December 2019; pp. 1–5.

64. Liu, L.; Wang, Y.; Liu, X.; Han, S.; Jia, L.; Meng, L.; Yang, Z.; Chen, W.; Zhang, Y.; Qiao, X. Computer-aided diagnostic system based on deep learning for classifying colposcopy images. *Ann. Transl. Med.* **2021**, *9*, 1045. [CrossRef] [PubMed]

65. Peng, G.; Dong, H.; Liang, T.; Li, L.; Liu, J. Diagnosis of cervical precancerous lesions based on multimodal feature changes. *Comput. Biol. Med.* **2021**, *130*, 104209. [CrossRef] [PubMed]

66. Asiedu, M.N.; Simhal, A.; Chaudhary, U.; Mueller, J.L.; Lam, C.T.; Schmitt, J.W.; Venegas, G.; Sapiro, G.; Ramanujam, N. Development of algorithms for automated detection of cervical pre-cancers with a low-cost, point-of-care, pocket colposcope. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 2306–2318. [CrossRef] [PubMed]

67. Hu, L.; Bell, D.; Antani, S.; Xue, Z.; Yu, K.; Horning, M.P.; Gachuhi, N.; Wilson, B.; Jaiswal, M.S.; Befano, B.; et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *JNCI J. Natl. Cancer Inst.* **2019**, *111*, 923–932. [CrossRef] [PubMed]

68. Chandran, V.; Sumithra, M.; 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**, *2021*, 5584004. [CrossRef] [PubMed]

69. Attallah, O. CerCan· Net: Cervical cancer classification model via multi-layer feature ensembles of lightweight CNNs and transfer learning. *Expert Syst. Appl.* **2023**, *229*, 120624. [CrossRef]

70. Tomko, M.; Pavliuchenko, M.; Pavliuchenko, I.; Gordienko, Y.; Stirenko, S. Multi-label classification of cervix types with image size optimization for cervical cancer prescreening by deep learning. In Proceedings of the ICICIT—Inventive Computation and Information Technologies, Coimbatore, India, 25–26 August 2022; Springer: Berlin/Heidelberg, Germany, 2023; pp. 885–902.

71. MobileODT. Intel & Mobileodt Cervical Cancer Screening, Kaggle. 2017. Available online: https://kaggle.com/competitions/intel-mobileodt-cervical-cancer-screening (accessed on 21 February 2024).

72. Rebolj, M.; van Ballegooijen, M.; Lynge, E.; Looman, C.; Essink-Bot, M.L.; Boer, R.; Habbema, D. Incidence of cervical cancer after several negative smear results by age 50: Prospective observational study. *BMJ* **2009**, *338*, b1354. [CrossRef] [PubMed]

73. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

74. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

75. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

76. 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, 26 June–1 July 2016; pp. 770–778.

77. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.

78. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.

79. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.

80. Dehghani, M.; Djolonga, J.; Mustafa, B.; Padlewski, P.; Heek, J.; Gilmer, J.; Steiner, A.P.; Caron, M.; Geirhos, R.; Alabdulmohsin, I.; et al. Scaling vision transformers to 22 billion parameters. In Proceedings of the International Conference on Machine Learning. PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 7480–7512.

81. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.

82. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. Lora: Low-rank adaptation of large language models. *arXiv* **2021**, arXiv:2106.09685.

83. Zhu, Y.; Shen, Z.; Zhao, Z.; Wang, S.; Wang, X.; Zhao, X.; Shen, D.; Wang, Q. Melo: Low-rank adaptation is better than fine-tuning for medical image diagnosis. *arXiv* **2023**, arXiv:2311.08236.

84. Darwish, M.; Altabel, M.Z.; Abiyev, R.H. Enhancing Cervical Pre-Cancerous Classification Using Advanced Vision Transformer. *Diagnostics* **2023**, *13*, 2884. [CrossRef] [PubMed]

85. Zhang, S.; Chen, C.; Chen, C.; Chen, F.; Li, M.; Yang, B.; Yan, Z.; Lv, X. Research on application of classification model based on stack generalization in staging of cervical tissue pathological images. *IEEE Access* **2021**, *9*, 48980–48991. [CrossRef]

86. Tanimu, J.J.; Hamada, M.; Hassan, M.; Kakudi, H.; Abiodun, J.O. A machine learning method for classification of cervical cancer. *Electronics* **2022**, *11*, 463. [CrossRef]

87. Waly, M.I.; Sikkandar, M.Y.; Aboamer, M.A.; Kadry, S.; Thinnukool, O. Optimal Deep Convolution Neural Network for Cervical Cancer Diagnosis Model. *Comput. Mater. Contin.* **2022**, *70*, 3295–3309.

88. Gorantla, R.; Singh, R.K.; Pandey, R.; Jain, M. Cervical cancer diagnosis using cervixnet-a deep learning approach. In Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, 28–30 October 2019; pp. 397–404.

89. Shao, R.; Shi, Z.; Yi, J.; Chen, P.Y.; Hsieh, C.J. On the adversarial robustness of vision transformers. *arXiv* **2021**, arXiv:2103.15670.

90. Chen, S.; Ge, C.; Tong, Z.; Wang, J.; Song, Y.; Wang, J.; Luo, P. Adaptformer: Adapting vision transformers for scalable visual recognition. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 16664–16678.

91. Chen, Z.; Duan, Y.; Wang, W.; He, J.; Lu, T.; Dai, J.; Qiao, Y. Vision transformer adapter for dense predictions. *arXiv* **2022**, arXiv:2205.08534.

92. Li, X.L.; Liang, P. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv* **2021**, arXiv:2101.00190.

93. Wang, X.; Wang, G.; Chai, W.; Zhou, J.; Wang, G. User-aware prefix-tuning is a good learner for personalized image captioning. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Shenzhen, China, 14–17 October 2022; Springer: Berlin/Heidelberg, Germany, 2023; pp. 384–395.

94. Jia, M.; Tang, L.; Chen, B.C.; Cardie, C.; Belongie, S.; Hariharan, B.; Lim, S.N. Visual prompt tuning. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 709–727.

95. Sohn, K.; Chang, H.; Lezama, J.; Polania, L.; Zhang, H.; Hao, Y.; Essa, I.; Jiang, L. Visual prompt tuning for generative transfer learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 19840–19851.

96. Hörst, F.; Rempe, M.; Heine, L.; Seibold, C.; Keyl, J.; Baldini, G.; Ugurel, S.; Siveke, J.; Grünwald, B.; Egger, J.; et al. Cellvit: Vision transformers for precise cell segmentation and classification. *Med. Image Anal.* **2024**, *94*, 103143. [CrossRef] [PubMed]

97. Abel, S.M.; Hong, Z.; Williams, D.; Ireri, S.; Brown, M.Q.; Su, T.; Hung, K.Y.; Henke, J.A.; Barton, J.P.; Le Roch, K.G. Small RNA sequencing of field Culex mosquitoes identifies patterns of viral infection and the mosquito immune response. *Sci. Rep.* **2023**, *13*, 10598. [CrossRef] [PubMed]

98. Cai, Z.; Yin, W.; Zeng, A.; Wei, C.; Sun, Q.; Yanjun, W.; Pang, H.E.; Mei, H.; Zhang, M.; Zhang, L.; et al. Smpler-x: Scaling up expressive human pose and shape estimation. *Adv. Neural Inf. Process. Syst.* **2024**, *36*, 1–26. [CrossRef]

99. Pfaendler, R.; Hanimann, J.; Lee, S.; Snijder, B. Self-supervised vision transformers accurately decode cellular state heterogeneity. *bioRxiv* **2023**. [CrossRef]

100. Doron, M.; Moutakanni, T.; Chen, Z.S.; Moshkov, N.; Caron, M.; Touvron, H.; Bojanowski, P.; Pernice, W.M.; Caicedo, J.C. Unbiased single-cell morphology with self-supervised vision transformers. *bioRxiv* **2023**. [CrossRef] [PubMed]

101. Bhattacharjee, D.; Süsstrunk, S.; Salzmann, M. Vision transformer adapters for generalizable multitask learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 19015–19026.

102. Alijani, S.; Fayyad, J.; Najjaran, H. Vision Transformers in Domain Adaptation and Generalization: A Study of Robustness. *arXiv* **2024**, arXiv:2404.04452.

103. Cruz, D.A.; Villar-Patiño, C.; Guevara, E.; Martinez-Alanis, M. Cervix type classification using convolutional neural networks. In Proceedings of the VIII Latin American Conference on Biomedical Engineering and XLII National Conference on Biomedical Engineering: Proceedings of CLAIB-CNIB 2019, Cancún, Mexico, 2–5 October 2019; Springer: Berlin/Heidelberg, Germany, 2020; pp. 377–384.

104. Aina, O.E.; Adeshina, S.A.; Aibinu, A. Classification of cervix types using convolution neural network (cnn). In Proceedings of the 2019 15th International Conference on Electronics, Computer and Computation (ICECCO), Abuja, Nigeria, 10–12 December 2019; pp. 1–4.

105. Payette, J.; Rachleff, J.; de Graaf, C. *Intel and Mobileodt Cervical Cancer Screening Kaggle Competition: Cervix Type Classification Using Deep Learning and Image Classification*; Stanford University: Stanford, CA, USA, 2017.