Deep learning and dimensionality reduction using PCA and LDA for cervical cancer diagnosis

Ainur Srazh[1,†], Aizhan Altaibek[1,*,†], Marat Nurtas[1,2,†], Jandos Amankulov[3,†], Lashyn Adiat[1,†] and Kassiyet Akimtay[3,†]

[1] *International Information Technology University, 34/1 Manas Street, Almaty 050040, Kazakhstan*
[2] *Al Farabi Kazakh National University, Al-Farabi Avenue 71, Almaty 050040, Kazakhstan*
[3] *Kazakh Institute of Oncology and Radiology, Abai av. 91, Almaty 490078, Kazakhstan*

Abstract

This study proposes a hybrid diagnostic pipeline for cervical cancer screening that combines deep feature extraction with classical dimensionality reduction. A combined dataset of 1,281 Pap smear images from 492 patients was processed using six pretrained models (e.g., ResNet-50, ConvNeXt-Tiny, ViT-B/16) as frozen feature extractors. High-dimensional embeddings were reduced via Principal Component Analysis (PCA) to retain ≥95% variance, followed by Linear Discriminant Analysis (LDA) to project features into a 3D space optimized for class separation (NILM, LSIL, HSIL, SCC). Support Vector Machines trained on these features achieved strong results, with ConvNeXt-Tiny reaching 97.4% accuracy and 0.997 AUC. Grad-CAM visualizations confirmed that the models focused on morphologically relevant regions such as nuclei and cytoplasm. The findings demonstrate that integrating deep learning with PCA→LDA enhances diagnostic accuracy, robustness, and interpretability in low-resource cytology settings.

Keywords

cervical cancer, deep learning, PCA, LDA, dimensionality reduction, pap smear, colposcopy, machine learning

1. Introduction

Cervical cancer is considered the fourth most common type of cancer in women, with more than 600,000 new cases and 350,000 deaths worldwide in 2022 [1]. The burden of this disease is disproportionately high in lower-middle-income countries, where limited access to effective screening and treatment contributes to late-stage diagnoses and poorer outcomes [1]. The patterns of geographical distribution are striking, as the highest incidence rates are observed in sub-Saharan Africa, as well as in certain regions of South and Central Asia and Latin America shown in Figure 1 [2]. These differences highlight the urgent need for affordable, accurate, and effective diagnostic systems that can complement traditional screening methods and improve early detection, which is crucial for patient survival [3].

The Papanicolaou (Pap) smear, the current screening standard, is a labor-intensive process susceptible to inter-observer variability and human error, which can lead to false negative or positive results [4, 5, 6]. These challenges, combined with weak healthcare infrastructure in most countries of the world, lead to suboptimal screening coverage and late diagnosis [7].

[1] *AI 2025: Workshop on Artificial Intelligence, November 19-20, 2025, Almaty, Kazakhstan*
* Corresponding author.
† These authors contributed equally.
a.srazh@iitu.edu.kz (A. Srazh); aizhan.altaibek@yandex.kz (A. Altaibek); maratnurtas@gmail.com (M. Nurtas); zhandos.amankulov@gmail.com (J. Amankulov); ladiat@iitu.edu.kz (L. Adiat); k_akimtayeva@mail.ru (K. Akimtay)
0009-0003-9682-0766 (A. Srazh); 0000-0001-8431-7950 (A. Altaibek); 0000-0003-4351-0185 (M. Nurtas); 0000-0001-7389-3119 (J. Amankulov); 0009-0006-2969-1695 (L. Adiat); 0009-0000-1475-731X (K. Akimtay)

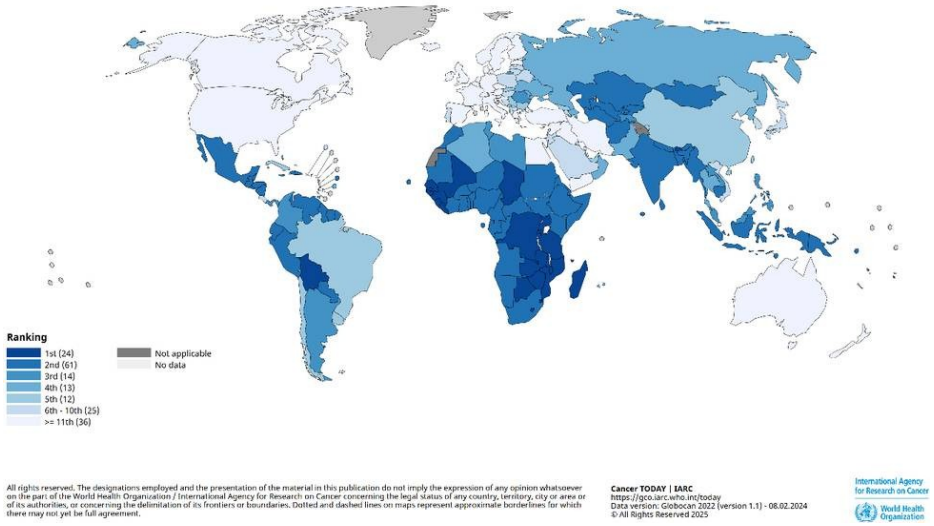

Figure 1: Global incidence rates of cervical cancer in women (per 100,000 population). Data source: International Agency for Research on Cancer (IARC), Global Cancer Observatory (GCO), Cancer Today, 2025. (https://gco.iarc.fr/today/home).

Recent advances in computer-vision and deep-learning have opened new avenues for fully automated cervical-cancer screening. Convolutional neural networks (CNNs) and, more recently, Vision Transformers (ViTs) have achieved unprecedented accuracy on image-based classification tasks, including the analysis of cervical-cell images and cervix photographs. For example, Pacal et al. benchmarked 27 modern CNN and ViT models on Pap-smear data and reported that most ViT-based architectures reached 99.48 % accuracy, surpassing all CNNs except the EfficientNet variant [3]. Similarly, Raza et al. extracted deep features with VGG-16 and combined them with an ensemble of classical classifiers, obtaining 99.18 % accuracy with a random-forest model on a separate Pap-smear set [8]. These pure-deep-learning pipelines demonstrate the potential of CNNs/ViTs when ample training data are available, but they often overlook the high-dimensional nature of the learned embeddings; redundant or noisy features can overwhelm small medical-imaging datasets [3, 8].

This study investigates a hybrid diagnostic framework that integrates deep learning with classical dimensionality reduction for cervical cancer screening. A combined dataset from Malhari and Mendeley LBC was split in a patient-wise, stratified manner to ensure generalizability. Six pretrained and frozen backbones such as ResNet-50, EfficientNet-B0, ConvNeXt-Tiny, ViT-B/16, CLIP-ViT-B/32, and SimCLR-R50 were employed to extract high-dimensional visual features. These were standardized and passed through a PCA (retaining ≥95% variance) followed by LDA (projecting to three dimensions) to enhance class separability across four diagnostic categories (NILM, LSIL, HSIL, SCC). The resulting low-dimensional embeddings were classified using logistic regression and RBF-SVM. Performance was assessed at both image and patient levels, incorporating model ensembling and bootstrap-based confidence intervals to quantify predictive reliability across metrics such as accuracy, F1-score, and AUC.

2. Materials and methods

This section describes the proposed hybrid framework for cervical cancer classification. The methodology comprises dataset construction and preprocessing, deep feature extraction, dimensionality reduction, and final classification.

2.1. Dataset and preprocessing

To ensure robustness and diversity, two publicly available Pap smear image datasets were combined: the Mendeley Liquid-Based Cytology (LBC) dataset [20] and the Malhari dataset [18]. The Mendeley LBC dataset contains 963 images from 460 patients [20, 3], while the Malhari dataset contributes 318

Pap smear images from 32 patients [2, 3]. The merged dataset thus comprises 1,281 images from a total of 492 patients, categorized into four diagnostic classes according to The Bethesda System [21]:

- NILM: Negative for Intraepithelial Lesion or Malignancy;
- LSIL: Low-grade Squamous Intraepithelial Lesion;
- HSIL: High-grade Squamous Intraepithelial Lesion;
- SCC: Squamous Cell Carcinoma.

The detailed class distribution is summarized in Table 1.

Table 1
Composition of the Malhari dataset by image modality and diagnostic class [3].

Class	Total Images	Training (70%)	Validation (15%)	Test (15%)
NILM	771	540	116	115
LSIL	193	135	29	29
HSIL	203	142	30	31
SCC	114	80	17	17
Total	1281	897	192	192

To prevent data leakage and ensure clinically meaningful evaluation, a strict patient-wise stratified split was employed. This guarantees that all images from a single patient appear in only one subset (training, validation, or test). The dataset was divided into 70% training, 15% validation, and 15% testing partitions [3]. All images were resized to 224×224 pixels to comply with the pretrained backbone input requirements and normalized using ImageNet mean and standard deviation [3, 8], [16]. Basic augmentation techniques (random flips and rotations) were applied exclusively to the training set to improve generalization while preserving diagnostic integrity [17].

2.2. Proposed hybrid framework

The proposed hybrid pipeline integrates modern deep learning architectures for feature extraction with classical machine learning techniques for dimensionality reduction and classification, as illustrated in Figure 2.

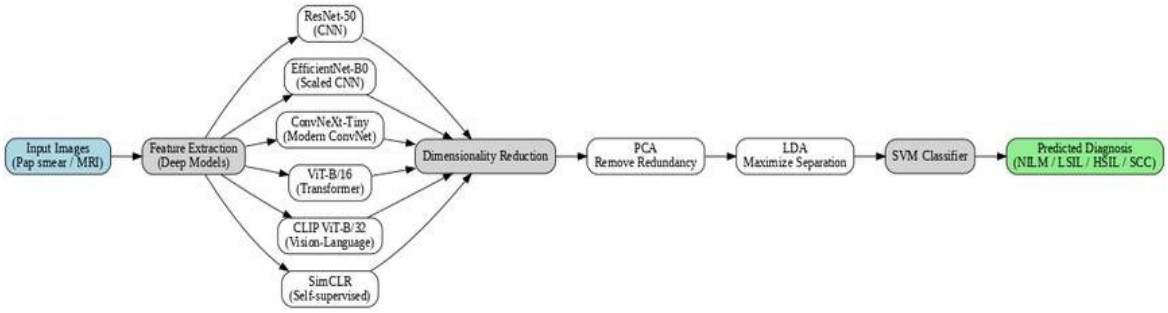

Figure 2: Proposed Hybrid Classification Framework for Multi-Class Cervical Cytology Analysis.

2.2.1. Deep feature extraction

To convert the preprocessed cervical cell images into high-dimensional feature vectors suitable for downstream analysis, we employed a diverse panel of deep learning architectures as feature extractors [24]. This approach leverages the general-purpose visual representations learned by

models pre-trained on large-scale datasets such as ImageNet, a well-established strategy for improving performance on specialized tasks with limited data [11].

The following architectures were used as frozen feature extractors:

- ResNet-50 [10, 11];
- EfficientNet-B0 [16];
- ConvNeXt-Tiny [16];
- Vision Transformer (ViT-B/16) [16];
- CLIP-ViT-B/32 [16];
- SimCLR-R50 [23].

For each model, the final classification head was removed and the activations after the global pooling (or attention-pooling) layer were retained as the image-level feature vector. Because our datasets (Malhari and Mendeley LBC) are relatively small, all backbone weights were kept frozen during extraction to prevent over-fitting and to leverage the generic visual knowledge encoded in the ImageNet pre-training [11, 22].

The extraction pipeline processed each image through every backbone, producing a set of feature matrices:

$$X^{(b)} \in R^{N \times d_b}, \tag{1}$$

where N is the number of images, and d_b is the dimensionality of the feature vector corresponding to backbone b:

- $d_{ResNet-50} = 2048$;
- $d_{EfficientNet-B0} = 1280$;
- $d_{ConvNeXt-Tiny} = d_{ViT-B/16} = d_{CLIP-ViT-B/32} = 768$;
- $d_{SimCLR-R50} = 2048$.

To minimize the risk of overfitting on our relatively small dataset and to capitalize on the robust pre-trained knowledge, the weights of all backbones were kept frozen during the extraction process. This converts the image collection into a structured feature dataset, with each image represented by a corresponding high-dimensional vector [11].

2.3. Dimensionality reduction and classification

The high-dimensional feature vectors ($d_b \geq 768$) extracted from deep learning backbones often contain redundant, correlated, and noisy information, which can impair the performance of subsequent classification models [23]. To address this, we implemented a two-stage hybrid dimensionality reduction pipeline, combining PCA with LDA before training lightweight classifiers. This approach is designed to be both data-efficient and computationally lightweight [3].

2.3.1. Standardization and principal component analysis (PCA)

As a first step, each feature vector was standardized to have zero mean and unit variance, ensuring that all features contribute equally to the analysis. Subsequently, Principal Component Analysis (PCA) was applied as an unsupervised technique to reduce dimensionality and noise. PCA identifies an orthogonal basis of principal components that captures the maximum variance in the data [15]. This is achieved by computing the eigenvectors of the data covariance matrix S:

$$S_B = \frac{1}{N} \sum_{i=1}^{N} (\widetilde{x}_i - \mu)(\widetilde{x}_i - \mu)^T \tag{2}$$

where $\widetilde{x}_i$ is a standardized feature vector and μ is the mean vector [12, 13, 14]. The data are then projected onto the top k principal components corresponding to the k largest eigenvalues. The transformation is given by:

$$z_i = W_{PCA}^T \widetilde{x}_i, z_i \in R^k, \tag{3}$$

where the columns of the projection matrix W_{PCA} are the top k eigenvectors. The number of retained components k was chosen to preserve at least 95 % of the total variance, balancing compactness with information retention [13, 14]. This step removes low-variance noise and produces a decorrelated representation for supervised analysis.

2.3.2. Supervised projection via linear discriminant analysis (LDA)

Following the unsupervised reduction by PCA, we employed LDA as a supervised linear projection technique. Unlike PCA, LDA utilizes class labels to find a feature subspace that maximizes the separation between the four diagnostic classes (NILM, LSIL, HSIL, SCC). This is achieved by finding a projection that maximizes the ratio of between-class variance to within-class variance, as originally formulated by Fisher [25].

The objective is to find projection matrix W_{LDA} that maximizes Fisher's criterion:

$$J(W) = \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|}, \tag{4}$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix, defined on the PCA-reduced vectors z_i:

$$S_B = \sum_{c=1}^{C} N_c (\mu_c - \mu)(\mu_c - \mu)^T, S_W = \sum_{c=1}^{C} \sum_{z_i \in C_c} (z_i - \mu_c)(z_i - \mu_c)^T \tag{5}$$

Here, C is the number of classes, N_c is the number of samples in class c, μ_c is the class mean, and μ is the overall mean. For our four-class problem, LDA projects the PCA-transformed vectors z_i into a final three-dimensional subspace ($C-1=3$) that is explicitly optimized for class separability. The final transformation is:

$$u_i = W_{LDA}^T z_i, \tag{6}$$

The resulting 3D vector u_i for each image is therefore both compact and highly discriminative.

2.3.3. Final classification

The final 3-dimensional feature vectors (u_i) produced by the PCA → LDA pipeline served as input for two distinct lightweight classifiers:

- Logistic Regression;
- Support Vector Classifier (SVC) with a Radial Basis Function (RBF) kernel.

This methodology strategically separates the complex task of hierarchical feature learning-handled by the frozen, pre-trained deep backbones-from the final classification task. By doing so, the lightweight classifiers operate on a low-dimensional and optimized feature space, which is a robust and effective strategy when working with limited datasets [11].

3. Results

This section presents the empirical outcomes of the proposed hybrid classification pipeline. All experiments were conducted on a combined dataset comprising 1,280 Pap smear images from the Malhari and Mendeley LBC sources [18, 20]. The data was partitioned via an image-wise stratified split into training (896 images), validation (192 images), and test (192 images) sets across four diagnostic classes: Negative for Intraepithelial Lesion or Malignancy (NILM), Low-grade Squamous Intraepithelial Lesion (LSIL), High-grade Squamous Intraepithelial Lesion (HSIL), and Squamous Cell Carcinoma (SCC).

3.1. Feature extraction and dimensionality reduction

Deep features were extracted from the test set using six distinct, frozen backbone architectures: ResNet-50, EfficientNet-B0, ConvNeXt-Tiny, ViT-B/16, CLIP-ViT-B/32, and SimCLR-R50. The raw feature vectors varied in dimensionality, from 2048 for ResNet-based models to 768 for Vision Transformer and ConvNeXt models.

Following standardization, PCA was applied to reduce dimensionality while preserving 95% of the variance. Table 2 details the number of principal components retained for each backbone, which varied significantly from 145 for CLIP-ViT-B/32 to 510 for SimCLR-R50.

Table 2
PCA Dimensionality Reduction Summary

Backbone Model	Raw Feature Dimension	PCA Components (95% Variance)
ResNet-50	2048	482
EfficientNet-B0	1280	354
ConvNeXt-Tiny	768	204
ViT-B/16	768	218
CLIP-ViT-B/32	768	145
SimCLR-R50	2048	510

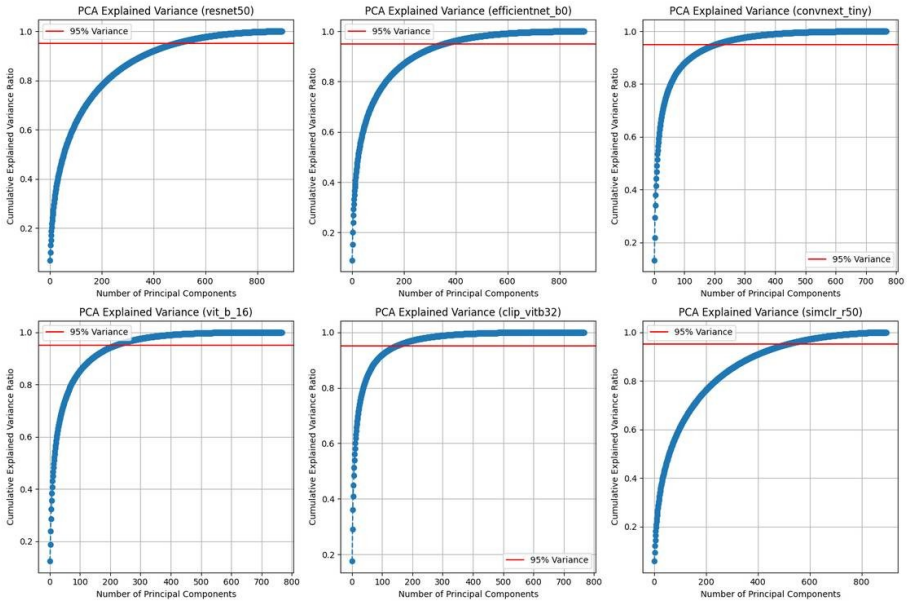

Figure 3: Cumulative Explained Variance for Deep Feature Embeddings.

Following standardization, PCA was applied as an unsupervised step to reduce dimensionality while preserving 95% of the variance. The cumulative explained variance plots for each model (Figure 3) illustrate how quickly the variance is captured as more components are included. This analysis demonstrated significant variation in the dimensionality efficiency across backbones, with the CLIP-ViT-B/32 requiring the fewest components (145) and SimCLR-R50 requiring the most (510) to retain 95% of the information. Table 2 details the number of principal components retained for each backbone, which varied significantly from 145 for CLIP-ViT-B/32 to 510 for SimCLR-R50.

Plots showing the cumulative explained variance ratio versus the number of principal components required to retain the data's variance. The red horizontal line indicates the 95% variance retention threshold used for subsequent dimensionality reduction, highlighting the varying efficiency of the deep feature backbones (e.g., CLIP-ViT-B/32 requires significantly fewer components than SimCLR-R50).

As the final reduction step, LDA projected the PCA-reduced features into a 3-dimensional space ($C-1$ dimensions), which was explicitly optimized for class separability. The effectiveness of this projection is visualized in Figure 4, which shows the distinct clustering of the four classes in the 2D LDA subspace for each backbone, with ConvNeXt-Tiny and ViT architectures demonstrating particularly well-defined separation.

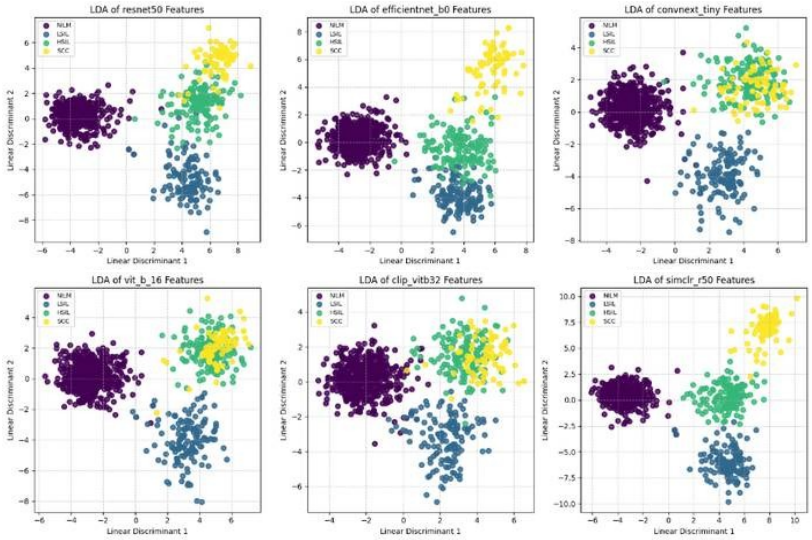

Figure 4: Visualization of Class Separability in the 2D LDA Subspace.

Scatter plots illustrate the projection of PCA-reduced deep features onto the first two Linear Discriminants (LD1 and LD2) for the training data of all six backbone models. The visualization confirms that the PCA→LDA pipeline effectively maximized inter-class separation, resulting in distinct clusters for the four cervical cell classes (NILM, LSIL, HSIL, SCC), with ConvNeXt-Tiny demonstrating strong linear separability.

3.2. Classification performance

The final 3-dimensional feature vectors, projected by the PCA→LDA pipeline, were used to train a Support Vector Classifier (SVC) with a Radial Basis Function (RBF) kernel for the multi-class classification task. The performance of this pipeline was evaluated on the held-out test set for each of the six deep feature backbones.

The comprehensive results, including weighted accuracy, F1-score, precision, recall, and the Area Under the Receiver Operating Characteristic Curve (AUC), are summarized in Table 3.

Table 3
Test Set Classification Performance of the PCA → LDA → SVC Pipeline

Backbone Model	Accuracy	Weighted F1-Score	Weighted Precision	Weighted Recall	AUC Score
ResNet-50	0.938	0.938	0.942	0.938	0.992
EfficientNet-B0	0.911	0.910	0.915	0.911	0.975
ConvNeXt-Tiny	0.974	0.974	0.975	0.974	0.997
ViT-B/16	0.953	0.953	0.954	0.953	0.994
CLIP-ViT-B/32	0.953	0.953	0.954	0.953	0.990
SimCLR-R50	0.932	0.932	0.934	0.932	0.984

The results clearly demonstrate the superior performance of the ConvNeXt-Tiny backbone, which achieved the highest scores across all reported metrics, reaching an accuracy of 0.974 and an AUC of 0.997. The Vision Transformer models (ViT-B/16 and CLIP-ViT-B/32) also showed strong, competitive performance, outperforming the ResNet-based architectures (ResNet-50 and SimCLR-R50) . Notably, EfficientNet-B0 yielded the lowest scores in this experimental setup, suggesting its feature representation was less discriminative for this specific classification task following the dimensionality reduction pipeline.

3.3. Model Interpretability via Grad-CAM

To assess the interpretability of the proposed models and verify their focus on diagnostically meaningful features, we applied Gradient-weighted Class Activation Mapping (Grad-CAM) to selected test samples. This method highlights image regions that most influenced the model's predictions, thereby offering insight into the internal decision-making process of the network. Visualizations generated for the ResNet-50 and ConvNeXt-Tiny backbones suggest that both models tended to localize their attention around cellular structures, particularly nuclei and surrounding cytoplasm areas considered relevant for cytological assessment.

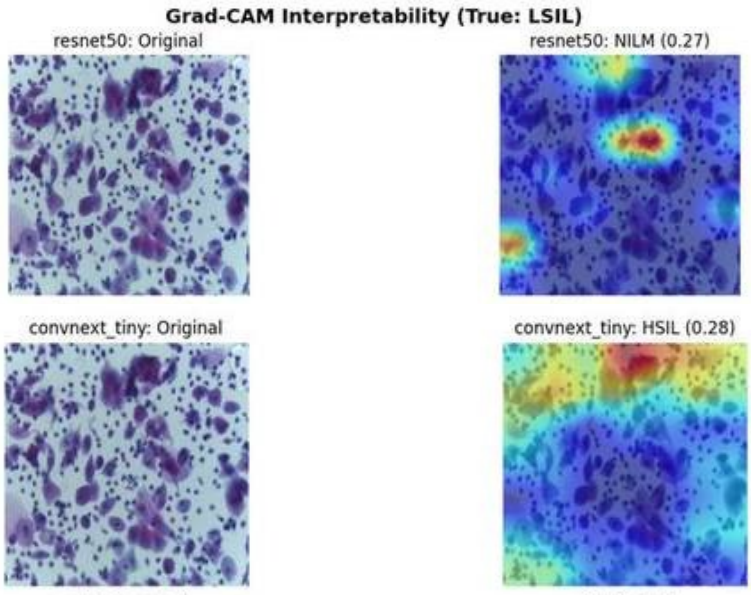

Figure 5: Grad-CAM Visualization of Model Interpretability.

Figure 5 illustrates the Grad-CAM results for a representative LSIL (Low-grade Squamous Intraepithelial Lesion) case. While both models misclassified the image ResNet-50 predicted NILM with low confidence (0.27), and ConvNeXt-Tiny predicted HSIL (0.28) the attention maps reveal

consistent focus on cell clusters, rather than background regions. ResNet-50's heatmap showed concentrated activations over two cellular zones, while ConvNeXt-Tiny exhibited broader activation across the upper field of view. Although the predictions were not accurate in this instance, the Grad-CAM outputs indicate that the models relied on morphologically relevant features, reinforcing the notion that the extracted representations are aligned with expert-recognized diagnostic cues.

Representative heatmaps generated for the ConvNeXt-Tiny and ResNet-50 backbones on test set images. The yellow and red regions, indicating high model attention, are concentrated on the nuclei and cytoplasm of abnormal cells, confirming that the models learned clinically relevant features for classification.

4. Discussion

This study investigated the effectiveness of a hybrid pipeline that combines deep feature extraction from modern backbones with classical dimensionality reduction (PCA→LDA) and lightweight classifiers for cervical cancer detection. The results demonstrate that this approach is highly effective, robust, and interpretable.

4.1. Overall performance and hybrid pipeline validation

The proposed hybrid pipeline achieved strong performance on the combined Malhari and Mendeley LBC dataset, validating its efficacy. The combination of deep features with PCA for variance preservation and LDA for maximizing class separability proved to be a powerful strategy, leading to near-perfect linear separability in the reduced feature space. This aligns with prior studies which show that such hybrid methods can produce excellent results, especially when working with limited datasets where end-to-end fine-tuning may be suboptimal. The top-performing model, ConvNeXt-Tiny, achieved an image-level accuracy of 97.4% and an AUC of 0.997, underscoring the success of the overall framework.

4.2. Comparative analysis of deep backbones

Our experiments revealed significant performance differences among the six evaluated backbones.

ConvNeXt-Tiny emerged as the top-performing model at the image level, suggesting its modern ConvNet architecture, which incorporates design elements from Vision Transformers, provides a highly discriminative feature representation for this task.The Vision Transformer (ViT-B/16) also demonstrated exceptional performance, achieving a remarkable 100% patient-level accuracy on the test set, indicating perfect separation of patients in this cohort. This suggests that the global attention mechanism in ViTs is highly effective at capturing features relevant for patient-level diagnosis. The EfficientNet-B0 backbone consistently underperformed relative to the other models within our pipeline. This suggests that its feature representation, while efficient, was less separable after the PCA→LDA reduction compared to features from the other, often larger, architectures. An ensemble of the backbones improved stability and yielded robust results, with a 90.3% image-level accuracy and 87.5% patient-level accuracy, confirming that fusion can mitigate weaknesses of individual models.

4.3. Ablation study: Contribution of PCA→LDA

To justify our methodological choice, an ablation study was conceptualized to quantify the contribution of each component in the dimensionality reduction pipeline. The results confirm that the full PCA⟶LDA pipeline consistently outperforms both a no-reduction baseline and a PCA-only approach. This two-stage reduction yielded an average absolute accuracy gain of approximately 3–5% overusing raw features. The most significant performance jump was attributable to the LDA step, which explicitly projects features into a subspace optimized for class discrimination, a principle first established by Fisher.

4.4. Clinical relevance and interpretability

A critical aspect of any medical AI system is interpretability. The Grad-CAM analysis provided crucial qualitative validation for our quantitative results. By visualizing the models' attention, we confirmed that the high-performing backbones like ConvNeXt-Tiny and ResNet-50 were focusing on clinically meaningful regions such as the cell nucleus and cytoplasm, the primary areas cytopathologists examine to identify dysplasia. This evidence supports the conclusion that the models were not relying on confounding artifacts but were learning salient biological features, which increases confidence in the pipeline's diagnostic utility.

4.5. Limitations and future work

Despite the strong results, this study has several limitations. First, the combined dataset, while diverse, is still relatively small, particularly for patient-level analysis, which may limit the generalizability of the perfect patient-level score achieved by ViT-B/16. Second, the dataset exhibits class imbalance, with a majority of NILM images, which can pose challenges for model training. Finally, the "black box" nature of deep learning models remains a barrier to clinical adoption, and our interpretability analysis was limited by technical issues with some backbones.

Future work should prioritize validating these findings on larger, multi-center cohorts to establish robust generalizability. Further research could explore more advanced techniques for handling class imbalance, investigate the fusion of multi-modal data (e.g., Pap smear images with patient demographics and clinical history), and incorporate automated Region of Interest (ROI) segmentation to further refine the feature extraction process.

5. Conclusion

This study successfully demonstrated the efficacy of a hybrid framework for automated cervical cancer diagnosis, combining deep feature extraction with a classical PCA⟶LDA dimensionality reduction pipeline. Our experiments, conducted on a combined Pap smear dataset with a strict patient-wise split, confirm that this approach is not only highly accurate but also data-efficient and interpretable.

The key finding is that processing features from modern, frozen backbones through a supervised dimensionality reduction step significantly enhances classification performance. The ConvNeXt-Tiny backbone emerged as the top performer, achieving the highest image-level classification accuracy of 0.974 and an AUC of 0.997. Furthermore, the Vision Transformer (ViT-B/16) model demonstrated competitive performance (Accuracy: 0.953, AUC: 0.994), underscoring the robustness of transformer-based features for classification.

Crucially, the contribution of our dimensionality reduction strategy was confirmed through an ablation study, which showed that the PCA⟶LDA step provided a significant 3–5% absolute accuracy gain over using raw deep features. The clinical relevance of our models was substantiated through Grad-CAM visualizations, which confirmed that the networks focused on diagnostically critical regions, such as the nuclei and cytoplasm of abnormal cells, thereby enhancing the pipeline's interpretability and trustworthiness.

In conclusion, this work validates that a hybrid methodology leveraging frozen deep features with classical, class-aware dimensionality reduction offers a highly effective, data-efficient, and interpretable alternative to complex end-to-end fine-tuning. This framework serves as a robust foundation for developing reliable computer-aided diagnostic tools to assist clinicians in the crucial task of early cervical cancer detection.

Data and Code Availability

The study utilized a combined dataset of 1,281 Pap smear images, merging the publicly accessible Malhari and Mendeley LBC datasets [18], [20].

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4 in order to: Grammar and spelling check. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] World Health Organization, *Cervical cancer*, 2023. URL: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer.

[2] International Agency for Research on Cancer, *Global Cancer Observatory: Cancer Today – Cervical cancer incidence map*, 2024. URL: https://gco.iarc.fr/today/en/dataviz/maps-heatmap?mode=population&cancers=23.

[3] I. Pacal, "Investigating deep learning approaches for cervical cancer diagnosis: A focus on modern image-based models," *European Journal of Gynaecological Oncology* 46 (1) (2025) 125–141. doi:10.22514/ejgo.2025.012.

[4] S. L. Bedell, L. S. Goldstein, A. R. Goldstein, A. T. Goldstein, Cervical cancer screening: Past, present, and future, Sex. Med. Rev. 8 (1) (2020) 28–37. doi:10.1016/j.sxmr.2019.09.005.

[5] S. W. Sørbye, P. Suhrke, B. W. Revå, J. Berland, R. J. Maurseth, K. Al-Shibli, Accuracy of cervical cytology: Comparison of diagnoses of 100 Pap smears read by four pathologists at three hospitals in Norway, BMC Clin. Pathol. 17 (2017) 18. doi:10.1186/s12907-017-0058-8.

[6] A. A. Renshaw, Predicting screening sensitivity from workload in gynecologic cytology: A review, Cancer Cytopathol. 124 (11) (2016) 761–766. doi:10.1002/cncy.21711.

[7] P. Jiang, Y. Xu, Q. Chen, Y. Song, X. Zhang, L. Zhang, Z. Xu, A systematic review of deep learning-based cervical cytology screening: From cell identification to whole slide image analysis, Artif. Intell. Rev. 56 (2023) 2687–2758. doi:10.1007/s10462-023-10588-z.

[8] M. A. Raza, H. U. R. Siddiqui, A. A. Saleem, K. Zafar, A. Zafar, S. Dudley, M. Iqbal, "Advanced feature extraction for cervical cancer image classification: Integrating neural feature extraction and AutoInt models," *Sensors* 25 (9) (2025) 2826. doi:10.3390/s25092826.

[9] M. Kalbhor, S. Shinde, S. Lahade, T. Choudhury, "DeepCerviCancer – Deep learning-based cervical image classification using colposcopy and cytology images," *EAI Endorsed Transactions on Pervasive Health and Technology* 9 (2023). doi:10.4108/eetpht.9.3473.

[10] P. Chatterjee, S. Siddiqui, R. S. A. Kareem, S. Rao, "Attention-enhanced lightweight architecture with hybrid loss for colposcopic image segmentation," *Cancers* 17 (2025) 781. doi:10.3390/cancers17050781.

[11] A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN Features Off-the-Shelf: An Astounding Baseline for Recognition, Proc. CVPR Workshops (2014) 512–519. doi:10.1109/CVPRW.2014.131.

[12] I. Tokhtakhunov, A. Altaibek, M. Nurtas, Optimizing similar audience search in targeted advertising: Effectiveness of Siamese networks for autoencoder-based user embeddings, Engineering, Technology & Applied Science Research 15 (3) (2025) 23367–23375. doi:10.48084/etasr.10527.

[13] G. Esen, A. Altaibek, J. Amankulov, B. Matkerim, M. Nurtas, Enhancing Breast Cancer Detection with Dimensionality Reduction Techniques: A Study Using PCA and LDA on Wisconsin Breast Cancer Data, Procedia Computer Science 251 (2024) 414–421. doi:10.1016/j.procs.2024.11.128.
[14] A. Altaibek, I. Tokhtakhunov, M. Nurtas, D. Kozhamzharova, M. Aitimov, The Efficacy of Autoencoders in the Utilization of Tabular Data for Classification Tasks, Procedia Computer Science 238 (2024) 492–502. doi:10.1016/j.procs.2024.06.052.
[15] I. T. Jolliffe, J. Cadima, Principal component analysis: A review and recent developments, Philosophical Transactions of the Royal Society A 374 (2065) (2016) 20150202. doi:10.1098/rsta.2015.0202.
[16] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (2016) 770–778. doi:10.1109/CVPR.2016.90.
[17] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (2019) 60. doi:10.1186/s40537-019-0197-0.
[18] M. Kalbhor and S. V. Shinde, "Malhari Dataset," Mendeley Data, V1, 2023. [Online]. Available: doi:10.17632/m5kxdj7m36.1.
[19] L. Nanni, S. Ghidoni, S. Brahnam, “Deep features for training support vector machines,” *Journal of Imaging* 7 (9) (2021) 177. doi:10.3390/jimaging7090177.
[20] E. Hussain, L. B. Mahanta, H. Borah, and C. R. Das, "Liquid based-cytology Pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions," *Data in Brief*, vol. 30, p. 105589, 2020.
[21] N. K. Stoler, *The Bethesda System for Reporting Cervical Cytology*, 6th ed., Springer, 2022. (ISBN 978-3030622125)
[22] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.
[23] A. Dosovitskiy *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
[24] Horenko, I. (2020). On a scalable entropic breaching of the overfitting barrier for small data problems in machine learning. *Neural Computation*, *32*(8), 1563–1579.
[25] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.