# A Hybrid Deep Learning Framework for Cervical Cancer Classification Using Multi-Scale Feature Fusion and Autoencoder Compression

V. Ulagamuthalvi[1], Angelin Keziah S[2*], Aswin S[3]

[1,2,3]Department of Computer Science and Engineering, Sathyabama University, India.

**Email:** sakeziah05@gmail.com[2]

## Abstract

Amidst the rapid growth of research and technology in cervical cancer detection, ensuring accuracy, especially with varied sample quality in traditional Pap smears, is still one of the major challenges. Automated classification of cervical cell images facilitates early detection of abnormal cellular changes critical for cervical cancer screening. This paper proposes a hybrid, multi-class cervical cell classification that integrates segmentation, deep feature extraction, feature compression, and ensemble classification. Initially, segmentation was done by training a Swin-UNETR using pseudo masks. These segmented images served as an input for the two main CNNs used for deep feature extractions at multiple resolutions. The fused features were subsequently refined and compressed using an MLP autoencoder. Classification was then performed by a weighted ensemble model incorporating both a Support Vector Machine and a Random Forest. Experiments were carried out on the SIPaKMeD dataset and achieved a test accuracy of 95.55%, confirming robust performance across all five cervical cell types. Therefore, based on the results, we can infer that combining segmentation, deep feature extraction at multiple resolutions, autoencoder's latent features and ensemble learning can be a practical and effective approach for automated cervical cell analysis.

## Keywords

Cervical cytology, Segmentation, Deep features, Autoencoder, Ensemble learning

## Introduction

Cervical malignancy remains a major health challenge for women worldwide, particularly in developing countries where routine screening is limited. The Pap smear remains the standard method for detecting cervical abnormalities through microscopic examination of cervical cell specimens (Yousafzade et al., 2023). Although it has significantly reduced mortality, it relies on manual assessment by cytologists and is susceptible to errors caused by fatigue, subjective interpretation, and irregular slide staining.

**Submission**: 13 May 2026; **Acceptance:** 11 June 2026; **Available online:** June 2026

Modern breakthroughs in computational intelligence have demonstrated the effectiveness of deep learning techniques for cervical cancer diagnosis (Asadi et al., 2020). Transfer learning and CNN-based models have achieved promising classification performance (Allogmani et al., 2024; Alsubai et al., 2023; Göker, 2024**;** Yousafzade et al., 2023). Hybrid CNN-Transformer and transformer-based architectures have further enhanced feature representation and contextual learning in medical image analysis (Brahmareddy & Selvan, 2025; Cao et al., 2022; Cheng et al., 2025; Zhu et al., 2024). Previous studies have emphasized the need for segmentation-guided region selection and image standardization to improve cervical image classification performance (Pal et al., 2021). Autoencoder-based feature learning and ensemble classification approaches have also shown potential for improving diagnostic performance through more discriminative feature representations (Akhila et al., 2024).

Despite these advancements, challenges remain, including low screening rates leading to late diagnosis, the high global burden of cervical cancer, limited access to healthcare in resource-constrained settings, and the prevalence of high-risk (HPV) infections (Sung et al., 2021; Yousafzade et al., 2023). Additionally, technical challenges such as variability in cellular morphology, image artifacts, segmentation difficulties, overlapping cells, and dependence on manual screening emphasize the necessity of robust automated systems for reliable cervical cancer detection (Alsubai et al., 2023). These challenges persist because many approaches focus on isolated stages such as segmentation or classification rather than a unified diagnostic framework (Wubineh et al., 2024). Furthermore, reliance on a single feature extraction strategy may limit the representation of important complementary features which may lead to reduced robustness in challenging cases. Cervical cell images consist of complex non-linear patterns therefore existing PCA-based approach being a linear dimensionality reduction technique, may not fully capture the complex nonlinear feature relationships present in cervical cytology images (Chauhan et al., 2025), whereas autoencoders can learn more discriminative nonlinear latent representations. Although AI-based screening methods such as Automated Visual Evaluation (AVE) have been introduced, studies have reported concerns regarding model portability, overfitting, and the reliability of predictions for borderline abnormalities, highlighting the need for rigorous validation before clinical deployment (Schiffman, 2026). These limitations underscore the need for more effective automated frameworks capable of improving the reliability and accuracy of cervical cancer diagnosis.

Recent studies have highlighted the potential of hybrid feature fusion, ensemble learning, and transformer-based architectures for improving cervical cell classification performance (Wubineh et al., 2024). Motivated by these observations, the main objective of this research is to improve automated cervical cell classification through an integrated framework that combines transformer-based segmentation, multi-resolution CNN feature extraction, autoencoder-based feature representation learning, and ensemble classification within a unified pipeline. By combining these complementary components, the proposed framework aims to improve feature representation, enhance classification robustness, and achieve more reliable cervical cell classification. The developed system is intended to support computer-aided analysis of cervical cytology images and provide a foundation for future automated cervical cancer screening applications.

## Methodology

### 1. Experimental environment:

This study employs the SIPaKMeD dataset, which is a collection of Pap-smear cytology images (Plissiti et al., 2018). The dataset consists of 4,049 cervical cell images distributed across five classes. The dataset was partitioned into training (70%, 2,834 images), validation (20%, 810 images), and testing (10%, 405 images) subsets using a stratified sampling strategy to preserve class distributions across all categories. Each image features a single, isolated cell. These cells are split evenly into five distinct categories, following established cytological and morphological standards.

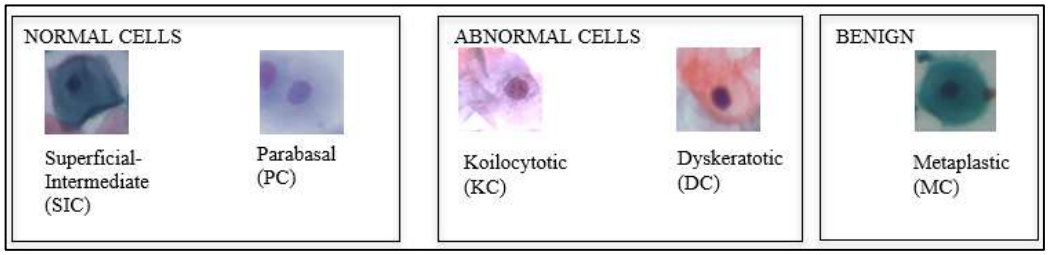


Figure 1. SIPaKMeD dataset images

The proposed framework was implemented using Python 3.8.20 and PyTorch 2.4.1. Model training and evaluation were performed using NVIDIA GPUs with CUDA 11.8 support.

### 2. System Architecture

The overall architecture of the proposed framework is illustrated in Figure 2 and consists of image preprocessing, cervical cell segmentation, deep feature extraction, feature compression, and ensemble-based classification.

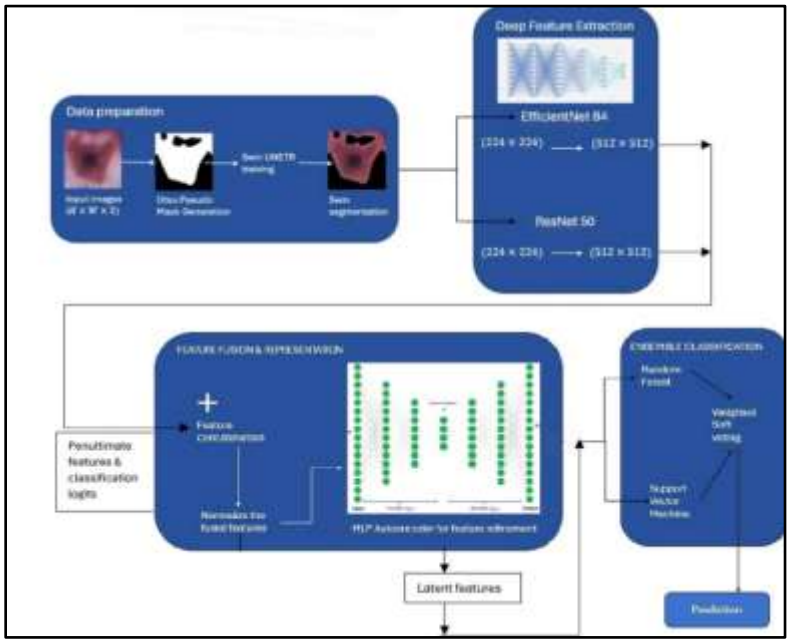


Figure 2. System Architecture

**3. Preprocessing:**

All images were converted to RGB, resized, transformed into PyTorch tensors, and normalized using standard ImageNet statistics to ensure dataset uniformity and compatibility with pre-trained models. A resolution of $224 \times 224$ was used for segmentation and mask generation, while both $224 \times 224$ and $512 \times 512$ resolutions were evaluated for feature extraction using EfficientNet-B4 and ResNet-50. To ensure reproducibility, a fixed random seed of 42 was used for Python, NumPy, and PyTorch operations throughout the experiments.

**4. Pseudo-Mask Generation and Segmentation**

To obtain approximate region-of-interest (ROI) labels without manual annotation, pseudo-masks were generated using Otsu's method. Each $(224 \times 224)$ image was smoothed with a $(5 \times 5)$ Gaussian filter and binarized at the optimal threshold $t^*$, separating cellular foreground from background. Morphological opening and closing with a $(3 \times 3)$ element then healed fragmented regions. These pseudo-masks served as weak labels.

The pseudo-masks were refined using a weakly supervised Swin-UNETR network, taking $224 \times 224 \times 3$ RGB images as inputs and Otsu-generated masks as labels. Swin-UNETR leverages shifted-window self-attention to capture local and long-range contextual information while preserving spatial coherence (Cao et al., 2022). The network outputted a $224 \times 224 \times 1$ probability map $S(x, y) \in [0, 1]$, which was thresholded to create a binary mask that isolated diagnostically relevant regions. Training lasted 20 epochs with a batch size of 4, using the AdamW optimizer (learning rate=$1 \times 10^{-4}$, weight decay=$1 \times 10^{-5}$), and Binary Cross-Entropy with Logits Loss. Random resized cropping, horizontal/vertical flipping, and colour jittering were applied to improve generalization.

**5. Feature Extraction using CNNs**

After lesion-focused masking, two complementary convolutional neural network backbones, Efficientnet-B4 and Resnet-50, were used for supervised classification. Both models initially were dependent on ImageNet pre-trained weights and were tailored to the specific number of classes for this task.

- EfficientNet-B4: It was selected as a high-capacity yet parameter-efficient backbone with compound scaling of depth, width, and resolution. The original classification head was replaced with a task-specific fully connected layer having an output dimension equal to the number of target classes. The network was first fine-tuned on lesion-focused masked images at $224 \times 224$ resolution, with all layers unfrozen to adapt the pre-trained features to cervical cytology images.
- ResNet-50: In parallel, a ResNet-50 backbone was used as a complementary architecture based on residual learning. The ImageNet pre-trained ResNet-50 was modified by replacing its final fully connected layer with a classifier matching the number of classes. Similar to EfficientNet-B4, the model was first fine-tuned on $224 \times 224$ masked images.

For both architectures, the best-performing $224 \times 224$ checkpoint initialized training at $512 \times 512$ resolution. This progressive multi-resolution strategy leveraged global context and fine-grained morphological details via partial weight loading of matching parameters. Both EfficientNet-B4 and ResNet-50 were optimized using AdamW (learning rate = $1 \times 10^{-4}$ and weight decay = $1 \times 10^{-5}$). A dropout rate of 0.4 was applied before the final classification layer, and learning rate adaptation was performed using a ReduceLROnPlateau scheduler.

**6. Penultimate and Logit Feature Extraction and Fusion:**

To build a rich, backbone-agnostic representation suitable for downstream ensemble learning, features were extracted from both trained backbones at $512 \times 512$ resolution. For each model, we captured:

- Penultimate-layer features ($h_{eff}$, $h_{res}$) : It represents high-level semantic embeddings before the final classification layer.
- Logits ($z_{eff}, z_{res}$) : It represents the pre-softmax class scores produced by the classifier head.

For EfficientNet-B4, a forward hook was registered on the average pooling layer and for ResNet-50 a hook was attached to the global pooling layer to obtain the penultimate feature. In parallel, the corresponding logits were collected from each network's final linear layer. For each sample, the hook activated features were flattened and concatenated with the logits to form a single backbone feature vector. Finally, the EfficientNet-B4 and ResNet-50 feature vectors were fused into a unified representation and used as input to the autoencoder.

**7. Feature Compression and Enhancement:**

The fused deep feature representation obtained from EfficientNet-B4 and ResNet-50 is high-dimensional and exhibits scale variability and redundancy due to heterogeneous backbone characteristics. Prior to dimensionality reduction, the fused feature vectors were normalized ($x_{norm}$) to ensure consistent feature scaling and stable optimization. Then, a fully connected MLP autoencoder learns the compact non-linear latent representation of the normalized fused features.

- Encoder Architecture and Function: The encoder projects $x_{norm}$ into a lower-dimensional latent space z, using a sequence of fully connected layers with decreasing dimensionality: $D \rightarrow 1024 \rightarrow 512 \rightarrow L$. Each intermediate layer applies layer normalization followed by ReLU activation and dropout. No normalization or dropout is applied at the latent layer to preserve a compact and discriminative embedding.
- Decoder Architecture and Function: The decoder reconstructs the normalized fused feature vector from the latent representation z, enforcing the retention of informative content in the compressed space. It mirrors the encoder with the dimensional flow: $L \rightarrow 512 \rightarrow 1024 \rightarrow D$. Layer normalization, ReLU activation, and dropout are applied after each intermediate linear layer, while the final reconstruction layer is linear and activation-free. This design enables accurate reconstruction of the normalized continuous feature values and constrains the latent space to capture the principal informative factors of the fused representation.

The autoencoder employed a latent dimension of 256 and was trained for 100 epochs using the AdamW optimizer with a learning rate of $1 \times 10^{-4}$ . Mean Squared Error (MSE) was used as the reconstruction loss function between the normalized input features ($x_{norm}$) and their reconstruction ($\hat{x}_{norm}$). After training, only the encoder was retained, and the resulting latent features were used as inputs to the Random Forest and Support Vector Machine classifiers.

**8. Final classification using ensemble ML models:**

The terminal stage of the proposed framework employs an ensemble learning strategy that combines Random Forest and Support Vector Machine classifiers. The motivation for adopting an ensemble approach lies in the inherent differences in the learning mechanisms of RF and SVM. The input consists of the compact latent features generated by the autoencoder. Random Forest is a tree-based ensemble model capable of capturing complex non-linear relationships through hierarchical partitioning of the feature space and SVM is a margin-based classifier that constructs optimal decision boundaries in high-dimensional spaces.

Hyperparameter optimization was performed using stratified 3-fold cross-validation, where class distributions were preserved across all folds. The average validation performance was used to select the optimal classifier configuration and improve model generalization.

- Random Forest: The optimal Random Forest configuration consisted of 300 decision trees with a maximum depth of 10.
- Support Vector Machine (SVM): Support Vector used a Radial Basis Function (RBF) mechanism to manage data distributions that lack linear divisibility. The RBF kernel allows the model to capture complex, non-linear relationships in the latent feature space generated by the autoencoder. The optimal SVM configuration employed an RBF kernel with a regularization parameter C = 1.0 and $\gamma = scale$ .

Weighted soft voting is an ensemble learning strategy that aggregates the probabilistic predictions of multiple models to achieve a more reliable final result. Unlike standard voting, which only considers a model's final choice, soft voting incorporates specific weights to these models, typically based on their individual accuracy. The final prediction for a feature vector came from blending the class-probability outputs of the RF and SVM models.

## Results and Discussion

This section presents the quantitative and qualitative results obtained from the proposed hybrid classification framework on the SIPaKMED cervical dataset.

1. **Segmentation Performance:** The Swin-UNETR model achieved a Dice score of 98.08% and an IoU of 96.2%, indicating highly accurate delineation of cervical cell regions. Compared with the SPP-SegNet baseline (Wubineh et al., 2025), the proposed segmentation approach produced higher precision, recall, and pixel accuracy. This improvement can be attributed to the transformer-based architecture of Swin-UNETR, which captures both local cellular structures and long-range contextual information through self-attention mechanisms [Table 1]. Furthermore, the accurate localization of diagnostically relevant regions addresses the need for segmentation-guided region selection highlighted in previous studies (Pal, A et al., 2021).

Table 1. Comparison of Swin-UNETR performance against the SPP-SegNet baseline.

| Model | Dice | F1-score | Precision | Recall | IOU | MCC | Pixel accuracy |
|---|---|---|---|---|---|---|---|
| SPP-SegNet (Wubineh et al., 2025) | - | - | 93.87% | 94.94% | 95.08% | - | 94.15% |
| **Swin-UNETR** | **98.08%** | **98.08%** | **98.05%** | **98.1%** | **96.2%** | **96%** | **98%** |

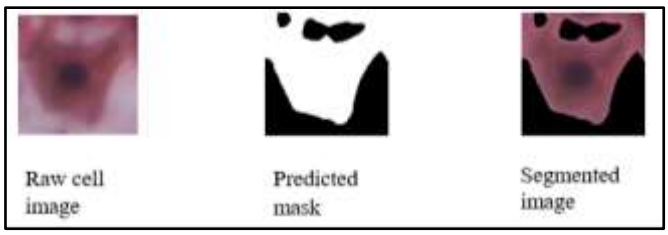


Figure 3. Output of the segmentation process

**Feature Extraction Performance:** Performance scores for EfficientNet-B4 and ResNet-50 are shown in Table 2. ResNet-50 achieved the highest accuracy of 95.31% at 512 x 512 resolution, benefiting from the availability of finer cellular details. However, EfficientNet-B4 performed slightly better at $224 \times 224$ than at $512 \times 512$. This suggests that increasing the image resolution did not provide additional useful information for EfficientNet-B4 and may have introduced variations that affected its classification performance.

Table 2. Performance metrics of CNN's at different resolutions

| **Model Architecture** | Input Resolution | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|---|
| EfficientNet-B4 | 224 × 224 | 94.81 | 94.61 | 94.55 | 94.55 |
| EfficientNet-B4 | 512 × 512 | 94.57 | 94.45 | 94.25 | 94.29 |
| ResNet-50 | 224 × 224 | 95.06 | 94.99 | 94.86 | 94.87 |
| ResNet-50 | 512 × 512 | 95.31 | 95.18 | 95.12 | 95.09 |

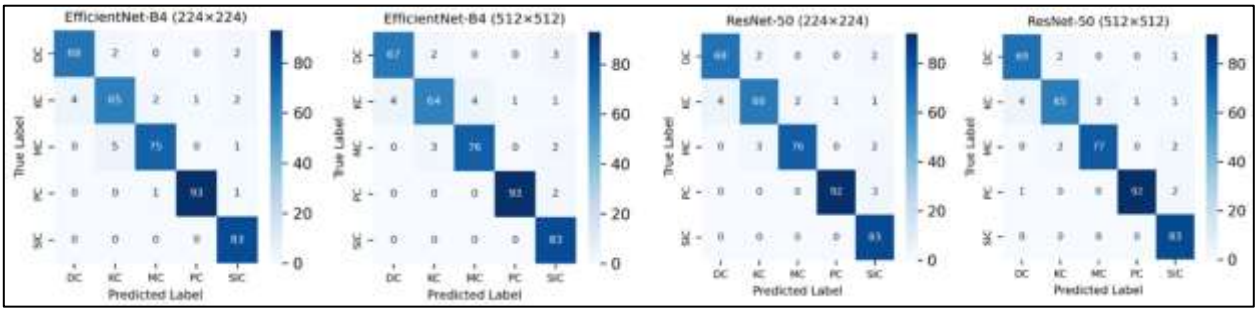


Figure 4. Confusion matrix for ResNet-50

**MLP Autoencoder performance:** The MLP-Autoencoder achieved an accuracy of 93.3% with an MSE of 0.066. The low reconstruction error indicates that the autoencoder successfully compressed the fused deep feature space while preserving most of the discriminative information.

**Performance of the Ensemble Model:** The final ensemble model reached an overall accuracy of 95.55%. The macro average AUC is 99.2% [Figure 6]. Confusion matrices for ensemble model is provided in Figure 5. The ensemble model achieved the same overall accuracy as the standalone SVM classifier. This suggests that the SVM generally produced stronger predictions than the Random Forest classifier therefore it contributed more heavily to the voting process, leading to performance similar to that of the SVM. However, the effect of weighted voting is data-dependent, and different datasets or prediction distributions may produce different ensemble outcomes.

Table 3: Performance of Ensemble models

| Classifier | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Random Forest | 94.57 | 94.43 | 94.30 | 94.28 |
| SVM | 95.56 | 95.40 | 95.30 | 95.30 |
| **Ensemble (Soft Voting)** | **95.56** | **95.40** | **95.30** | **95.30** |

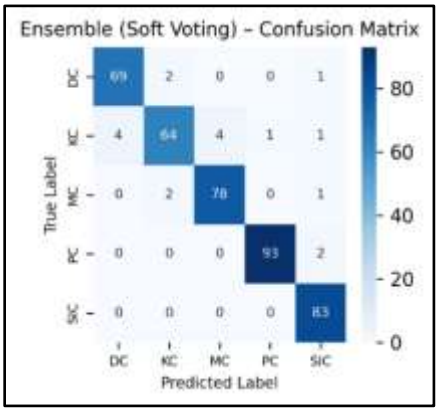


Figure 5. Confusion Matrix for the Ensemble Model

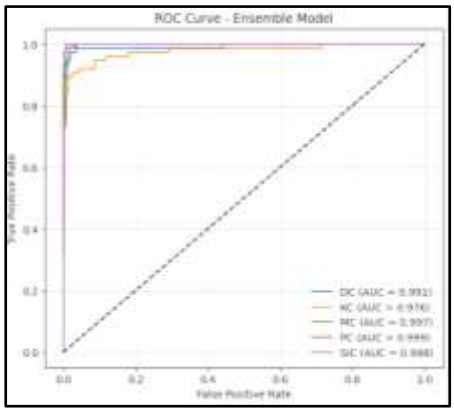


Figure 6. AUC-ROC Curve for the Ensemble Model

When comparing with previous studies, this approach showed some improvements by focusing more on improving feature quality using deeper models and combining their outputs. This helps in creating a stronger representation of the data.

Table 4. Comparison of the proposed system with existing research

| METHOD | DATASET | ACCURACY |
|---|---|---|
| CerviFormer (Transformer with cross-attention) (Deo et al., 2024) | SIPaKMeD | 93.70% |
| SPP-SegNet + SE-DenseNet (Wubineh et al., 2025) | SIPaKMeD | 95% |
| Privacy-Preserved CNN (Simple CNN (4 layers) + segmentation) (Alsubai et al., 2023) | SIPaKMeD | 91.13% |
| VGG16-based Neural Feature Extractor + AutoInt + LGBM (Raza et al., 2025) | Multi-class dataset | 91.11% |
| **Proposed Method (Hybrid - Swin-UNETR + CNN + AE + Ensemble)** | **SIPaKMeD** | **95.55%** |

Although overall accuracy remained unchanged, the ensemble provided a more robust prediction framework by combining complementary decision mechanisms from RF and SVM. The results demonstrate the effectiveness of the proposed framework for cervical cell classification; however, several limitations remain. First, the model was evaluated on cropped cervical cell images, whereas real clinical settings typically involve whole-slide images containing multiple overlapping cells, which may affect practical applicability. Second, the study was conducted using only the SIPaKMeD dataset, limiting the assessment of cross-dataset generalization. Finally, training multiple deep learning models at high resolution $512 \times 512$ required considerable

computational resources and training time, which may constrain scalability and rapid experimentation.

## Conclusion

This work presented a hybrid cervical cell analysis framework that integrates Swin-UNETR-based segmentation, deep feature extraction using EfficientNet-B4 and ResNet-50, latent-space compression through a multi-scale autoencoder, and a weighted soft-voting ensemble classifier. Unlike conventional feature engineering approaches, the autoencoder in this framework serves as a nonlinear representation learning module that compresses fused deep features into compact latent embeddings, which are subsequently used for ensemble classification. The experimental results on the SIPaKMeD dataset demonstrated high segmentation accuracy and strong classification performance across all five cervical cell categories. The combination of clean segmented inputs complementary deep features, and structured latent embeddings significantly enhanced the model. Overall, the proposed architecture provides an effective and reliable approach for automated cervical cytology analysis. Future work will therefore focus on extending the model to WSIs by integrating automatic cell detection and patch extraction for end-to-end analysis. Additionally, the use of a limited number of datasets may restrict generalization. To overcome this, future research will investigate triple-stage transfer learning across multiple cervical cytology datasets from diverse sources, which is expected to reduce dataset-specific bias.

## Acknowledgements

The authors would like to express their sincere gratitude to their institute, teaching staff, and guide for their continuous support and guidance. We also thank the anonymous reviewers, and the journal editors for their valuable and constructive feedback.

## References

Akhila, K. R., Muthukumaran, N., & Ahilan, A. (2024). Classification of cervical cancer using an autoencoder and cascaded multilayer perceptron. IETE Journal of Research, 70(1), 26-36. https://doi.org/10.1080/03772063.2022.2142859

Allogmani, A. S., Mohamed, R. M., Al-Shibly, N. M., & Ragab, M. (2024). Enhanced cervical precancerous lesions detection and classification using Archimedes Optimization Algorithm with transfer learning. Scientific Reports, 14(1), 12076. https://doi.org/10.1038/s41598-024-62773-x

Alsubai, S., Alqahtani, A., Sha, M., Almadhor, A., Abbas, S., Mughal, H., & Gregus, M. (2023). Privacy preserved cervical cancer detection using convolutional neural networks applied to pap smear images. *Computational and Mathematical Methods in Medicine*, *2023*(1), 9676206. https://doi.org/10.1155/2023/9676206

Asadi, F., Salehnasab, C., & Ajori, L. (2020). Supervised algorithms of machine learning for the prediction of cervical cancer. Journal of biomedical physics & engineering, 10(4), 513. https://doi.org/10.31661/jbpe.v0i0.1912-1027

Brahmareddy, A., & Selvan, M. P. (2025). TransBreastNet a CNN transformer hybrid deep learning framework for breast cancer subtype classification and temporal lesion progression analysis. *Scientific Reports*, *15*(1), 35106. https://doi.org/10.1038/s41598-025-19173-6

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2022, October). Swin-unet: Unet-like pure transformer for medical image segmentation. In European Conference on computer vision (pp. 205-218). Cham: Springer Nature Switzerland. https://doi.org/10.48550/arXiv.2105.05537

Chauhan, N. K., Singh, K., Kumar, A., Mishra, A., Gupta, S. K., Mahajan, S., ... & Kim, J. (2025). A hybrid learning network with progressive resizing and PCA for diagnosis of cervical cancer on WSI slides. Scientific Reports, 15(1), 12801. https://doi.org/10.1038/s41598-025-97719-4

Cheng, Y., Wu, H., Wu, F., Wang, Y., Jiang, W., Xiong, M., & Liu, L. (2025). Fine-grained pathomorphology recognition of cervical lesions with a dropped multibranch Swin Transformer. *Quantitative Imaging in Medicine and Surgery*, *15*(4), 3551-3564. https://doi.org/10.21037/qims-24-1590

Deo, B. S., Pal, M., Panigrahi, P. K., & Pradhan, A. (2024). CerviFormer: A pap smear-based cervical cancer classification method using cross-attention and latent transformer. *International Journal of Imaging Systems and Technology*, *34*(2), e23043. https://doi.org/10.48550/arXiv.2303.10222

Göker, H. (2024). Detection of cervical cancer from uterine cervix images using transfer learning architectures. Eskişehir Technical University Journal of Science and Technology A-Applied Sciences and Engineering, 25(2), 222-239. https://doi.org/10.18038/estubtda.1384489

Pal, A., Xue, Z., Desai, K., Banjo, A. A. F., Adepiti, C. A., Long, L. R., ... & Antani, S. (2021). Deep multiple-instance learning for abnormal cell detection in cervical histopathology images. Computers in Biology and Medicine, 138, 104890. https://doi.org/10.1016/j.compbiomed.2021.104890

Raza, M. A., Siddiqui, H. U. R., Saleem, A. A., Zafar, K., Zafar, A., Dudley, S., & Iqbal, M. (2025). Advanced feature extraction for cervical cancer image classification: integrating neural feature extraction and autoint models. *Sensors*, *25*(9), 2826. https://doi.org/10.3390/s25092826

Schiffman, M. (2026). RE: An observational study of deep learning and automated evaluation of cervical images for cancer screening. *JNCI: Journal of the National Cancer Institute*, *118*(2), 364-365. https://doi.org/10.1093/jnci/djaf348

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *71*(3), 209-249. https://doi.org/10.3322/caac.21660

Wu, Y., Hatipoglu, S., Alonso-Álvarez, D., Gatehouse, P., Li, B., Gao, Y., ... & Yang, G. (2021). Fast and automated segmentation for the three-directional multi-slice cine myocardial velocity mapping. Diagnostics, 11(2), 346. https://doi.org/10.3390/diagnostics11020346

Wubineh, B. Z., Rusiecki, A., & Halawa, K. (2024). Classification of cervical cells from the Pap smear image using the RES_DCGAN data augmentation and ResNet50V2 with self-attention architecture. Neural Computing and Applications, 36(34), 21801-21815. https://doi.org/10.1007/s00521-024-10404-x

Wubineh, B. Z., Rusiecki, A., & Halawa, K. (2025). Spp-segnet and se-densenet201: A dual-model approach for cervical cell segmentation and classification. Cancers, 17(13), 2177. https://doi.org/10.3390/cancers17132177

Youneszade, N., Marjani, M., & Pei, C. P. (2023). Deep learning in cervical cancer diagnosis: architecture, opportunities, and open research challenges. IEEe Access, 11, 6133-6149. https://doi.org/10.1109/ACCESS.2023.3235833

Zhu, S., Lin, L., Liu, Q., Liu, J., Song, Y., & Xu, Q. (2024). Integrating a deep neural network and Transformer architecture for the automatic segmentation and survival prediction in cervical cancer. Quantitative Imaging in Medicine and Surgery, 14(8), 5408. https://doi.org/10.21037/qims-24-560