GMR
The Original

A LIGHTWEIGHT ATTENTION-ENHANCED CNN FOR CERVICAL CELL CLASSIFICATION

Jujuroo Sowmya[1], Ravi Kumar Sachdeva[2*], G S Pradeep Ghantasala[3]

[1]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India, j.soumya25@gmail.com
[2]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India ravisachdeva1983@gmail.com
[3]Department of Computer Science & Engineering, Alliance school of Advanced Computing, Alliance University, Bengaluru, India. ggspradeep@gmail.com
*Corresponding Authors: Ravi Kumar Sachdeva, EMAIL: ravisachdeva1983@gmail.com

ABSTRACT

Accurate and early identification of abnormal cervical cells is essential for reducing cervical cancer mortality; however, current deep learning solutions often rely on computationally heavy archi-tectures that limit clinical deployment. To address this challenge, we propose a lightweight attention-enhanced convolutional neural network (AECNN) that integrates depthwise separable layers with spa-tial and channel squeeze-and-excitation attention. The model is designed to emphasize nucleus-specific morphological cues while maintaining low computational cost. In contrast to recent attention-guided frameworks such as AEDA [11] and LAHN [12], the proposed architecture introduces hybrid class-balancing through weighted cross-entropy and focal loss to mitigate false negatives and improve sensi-tivity in minority classes. Experiments conducted using SIPaKMeD for training, Herlev for validation, and CRIC for independent testing demonstrate significant improvements in diagnostic performance. The proposed AECNN achieved a sensitivity of 92.14%, specificity of 94.63%, accuracy of 93.87%, and an AUC of 0.958, outperforming AEDA and LAHN by margins of 3.2–5.6% depending on the metric. These results suggest that the proposed lightweight design offers a clinically viable balance between high discriminatory capability and computational feasibility, supporting the potential integration of the method into real-time screening applications.

KEYWORDS: Cervical cytology, Lightweight CNN, Attention mechanism, Medical image classification, Deep learning.

1 INTRODUCTION

Cervical cancer remains one of the leading preventable malignancies affecting women worldwide, with early cytological screening serving as a cornerstone for reducing mortality. Conventional microscopic analysis, although clinically accepted, is labor-intensive and susceptible to intra- and inter-observer variation. As screening volumes increase, manual examination alone becomes insufficient for timely and consistent diagno-sis. In response, numerous deep learning approaches have emerged to support automated cervical cytology interpretation [2–6, 8–12]. Despite their promising results, many of these models remain computationally expensive, rely on extensive training resources, or lack robustness across datasets, limiting their clinical applicability.

Recent studies have highlighted the benefits of attention-enhanced architectures in improving model focus on diagnostically significant cellular regions, particularly nuclei and cytoplasm boundaries [4, 11, 12]. For ex-ample, Verma et al. introduced an attention-driven network combining fuzzy logic with multi-head attention mechanisms [11], whereas Xia et al. demonstrated that lightweight attention modules could improve cell-level feature discrimination with reduced computational footprint [12]. However, many existing techniques still de-pend on large backbones, transfer learning, or multi-stage inference pipelines, which complicates deployment in low-resource screening environments [3, 5, 9].

The motivation for this work arises from the need to balance diagnostic reliability with computational feasibility, especially in healthcare settings where hardware resources and latency constraints limit the use of high-capacity models. Within this context, improving sensitivity—particularly toward pre-cancerous and ma-lignant cell types—remains critical, as false negatives can delay treatment and worsen outcomes. Therefore, this study focuses on designing a lightweight yet highly discriminative model capable of operating efficiently without sacrificing diagnostic precision.

The primary contributions of this work are as follows:

- We propose a lightweight attention-enhanced convolutional neural network (AECNN) optimized for cervical cell classification, integrating depthwise separable convolutions with squeeze-and-excitation (SE) attention to emphasize nucleus-related spatial cues.
- A hybrid loss strategy combining weighted cross-entropy and focal loss is applied to reduce class imbalance sensitivity and improve recognition of rare pathological categories.
- A rigorous cross-dataset evaluation protocol is employed, training on SIPaKMeD, validating on Herlev, and testing on CRIC to assess generalization performance across real-world screening variability.
- Comparative benchmarking against recently published methods, including AEDA [11] and LAHN [12], demonstrates the efficiency and diagnostic improvements achieved by the proposed approach.

The remainder of this manuscript is structured as follows: 2 reviews recent advances in deep learning for cervical cytology. The proposed model architecture is described in detail in 3. Experimental settings, including datasets, training strategy, and evaluation metrics, are provided in 4. Results and comparative analyses are presented in 5, followed by a discussion of strengths, limitations, and future directions in 6. Finally, 7 concludes the study.

2 RELATED WORK

Deep learning-based cervical cytology analysis has evolved rapidly in recent years, with a clear trend toward integrating attention mechanisms, lightweight architectures, and hybrid feature learning. Early models pri-marily relied on traditional CNN backbones, often adapted from natural image classification, which demon-strated promising accuracy but lacked robustness and computational efficiency. As reported in [6], classical CNNs and machine learning–based models significantly depended on handcrafted features and exhibited severe limitations in generalizing across datasets with visual variability. With increasing dataset availability, more sophisticated neural architectures emerged, shifting the research landscape toward attention-driven and computationally optimized designs.

Attention-enhanced deep learning frameworks have demonstrated improved localization of diagnostically relevant nuclei regions [3, 4, 11]. For instance, Verma et al. employed multi-head attention and fuzzy logic to enhance interpretability [11], while Li et al. introduced multi-scale spatial attention to capture hierarchical features across cytoplasmic and nuclear boundaries [4]. Although precise in segmentation and classification, these models are resource-intensive and require high memory footprints, limiting their deployment in clinical environments with restricted computational capacity. Similarly, Jawahar et al. developed attention-based classification for leukemia cells [3], providing evidence that attention improves discrimination in visually similar cell types; however, model complexity remained high.

Recent works have begun exploring lightweight yet attention-enhanced architectures. Xia et al. proposed a reduced-parameter network leveraging channel-wise attention [12], while Muksimova et al. adopted an optimized YOLO-based framework for medical imaging [5]. These approaches demonstrate reduced com-putational demand; however, many still struggle with maintaining diagnostic sensitivity, particularly for minority abnormal classes. Chatterjee et al. further integrated hybrid losses for segmentation accuracy [2], while Shi et al. demonstrated strong efficiency gains in blood cell detection using lightweight CNNs [10]. Al-though efficient, these approaches often lack comprehensive cross-dataset validation, raising concerns about real-world robustness.

Some recent studies attempt to improve diagnostic generalization using hybrid or multi-modal learn-ing [8, 9]. While these works report improved resilience to imaging variability, they introduce pipeline com-plexity, additional preprocessing, and domain-specific dependencies, making clinical translation challenging. Conversely, fusion-based or hierarchical architectures such as [1,7] have shown strong performance in broader medical imaging contexts; however, adaptation for cervical cytology remains underexplored and computa-tionally heavy.

Table 1 summarizes key characteristics of representative related models.

In summary, although existing systems demonstrate notable advances in attention-driven and lightweight computational design, they often suffer from trade-offs between efficiency, diagnostic accuracy, sensitivity to minority classes, and cross-dataset generalization. Motivated by these gaps, the proposed AECNN model introduces a computationally efficient architecture integrating depthwise-separable convolutions and dual-domain attention while incorporating a hybrid loss strategy to improve sensitivity and reduce misclassification of abnormal cells. By validating across three independent datasets, the proposed model aims to bridge the gap between diagnostic robustness and hardware feasibility, addressing key shortcomings identified in prior systems.

Table 1: Summary of Advantages and Limitations in Existing Models

Study	Strengths	Limitations
[11]	High interpretability via attention	High computational cost
[12]	Lightweight design	Reduced sensitivity in rare classes
[5]	Real-time inference potential	Limited cervical-specific benchmarking
[4]	Strong feature localization	Large backbone requirements
[8, 9]	High generalization	Complex pipeline and pre-processing

3 PROPOSED APPROACH

The proposed Attention-Enhanced Lightweight Convolutional Neural Network (AECNN) is designed to balance classification accuracy with computational efficiency, addressing the limitations observed in existing works [4, 5, 9, 11, 12]. The architecture integrates three core design components: (1) depthwise separable convolutional layers forming a MobileNet-like backbone, (2) hybrid attention via spatial and channel squeeze-and-excitation (SE) blocks, and (3) a dual-loss optimization scheme combining focal loss and weighted cross-entropy. This design ensures improved feature selectivity, lower parameter count, and enhanced sensitivity, particularly for minority abnormal cell types consistent with prior findings [2, 3, 10].

Model Architecture

Let an input cervical cell image be represented as $X \in \mathrm{R}^{H\times W\ \times C}$. The backbone processes X using depthwise separable convolution defined as:

$$Y = (X * K_d) \oplus (X * K_p), \quad (1)$$

where K_d and K_p denote depthwise and pointwise kernels, respectively. This reduces computational cost from $O(k^2mn)$ to $O(k^2m + mn)$, making the model significantly lighter than traditional CNNs, similar to benefits reported in [5, 7].

Attention Integration

To enhance discriminative capability, spatial attention (SA) and channel SE attention (CA) refine feature maps. The channel attention mechanism computes:

$$CA(F) = \sigma(W_2\delta(W_1 GAP(F))), \quad (2)$$

as: where σ is sigmoid, δ is ReLU, and GAP denotes global average pooling. Spatial attention is computed

$$SA(F) = \sigma(f^{7\times 7}([AvgPool(F); MaxPool(F)])). \quad (3)$$

The final refined feature representation is:

$$F_{att} = F \odot CA(F) \odot SA(F), \quad (4)$$

which focuses representation on nuclear morphology—critical for distinguishing dysplastic cells, consistent with findings in [4, 11].

```
**Input:** Training set $D = \{(X_i, y_i)\}$, model $f_\theta$, hyperparameters $(\alpha, \gamma, \lambda_1, \lambda_2)$
**Output:** Optimized model parameters $\theta$
Initialize model weights $\theta$

for each epoch do
    for each batch $(X, y)$ do
        Compute features: $F = Backbone(X)$
        Apply attention: $Fatt = SA(CA(F))$
        Predict outputs: $\hat{y} = Softmax(Fatt)$
        Compute hybrid loss: $L_{total} = \lambda_1 L_w + \lambda_2 L_f$
        Update $\theta$ using backpropagation
Return $\theta$
```

Loss Function and Optimization

To address dataset imbalance, the objective function integrates weighted cross entropy L_w and focal loss L_f [8, 9]:

$$L_{total} = \lambda_1 L_w + \lambda_2 L_f, \quad (5)$$

where:

$$L_f = -\alpha_t(1 - p_t)\gamma \log(p_t). \quad (6)$$

This formulation intentionally penalizes false negatives more aggressively, especially in malignant cate-gories.

Theoretical Justification

Theorem 1. *Given a classifier f_θ trained with focal loss in addition to weighted cross entropy, the expected misclassification rate of minority classes is strictly lower than using weighted cross-entropy alone under balanced convergence.*

Proof Sketch. Since focal loss down-weights easy examples, its gradient contribution emphasizes mi-nority hard samples:

$$\frac{\partial L_f}{\partial p_t} \text{ a } (1 - p_t)^{\gamma - 1}, \quad (7)$$

Where as weighted cross-entropy gradient remains uniform across class boundaries. Therefore, $E[Err_{minority}]_{focal} < E[Err_{minority}]_{wCE}$. Practical outcomes from recent benchmarks support this theoretical claim [1, 6].

Algorithm

The architectural choices and optimization strategies collectively enable high sensitivity and robustness while remaining computationally efficient, advancing beyond recent works [5, 7, 10, 12].

4 System Implementation and Experimental Setup

This section describes the implementation details, datasets, development environment, and evaluation proto-col used to assess the proposed AECNN model. The experimental workflow follows a structured methodology to ensure reproducibility, robustness, and reliable performance evaluation across different data sources, as recommended in prior cervical cytology studies [4, 9, 11, 12].

Development Environment
All experiments were conducted using a workstation configured with an NVIDIA RTX 4090 GPU (24 GB VRAM), 128 GB RAM, and an AMD Threadripper processor. The model was implemented in Python us-ing the PyTorch framework (v2.1), supported by CUDA (v12.2) and cuDNN acceleration. Auxiliary tools included OpenCV for preprocessing, Albumentations for augmentation, and Scikit-learn for evaluation and visualization. This setup ensured efficient handling of high-resolution cell images similar to prior lightweight biomedical models [5, 10].

Datasets and Sample Characteristics
Three publicly available cervical cytology datasets were utilized:
– **SIPaKMeD**: Used for model training due to its balanced representation of five cervical cell categories.
The dataset contains 4,049 labeled samples captured using Pap-stained cytology slides.
– **Herlev Dataset**: Used for validation, consisting of 917 cell images categorized into seven clinically relevant classes. This dataset features visible inter-class morphological variability and imaging noise, reflecting real-world screening variability [6].
– **CRIC Dataset**: Used exclusively for independent testing to assess model generalization and cross-dataset stability, following recommendations in recent multi-center studies [3, 8].

All images were resized to 224×224 resolution. Normalization followed dataset-specific mean and standard deviation ranges to preserve pixel distributions.

Preprocessing and Augmentation
To improve robustness, the following augmentations were applied probabilistically during training:
– Random flipping (horizontal/vertical)
– Gaussian noise insertion
– Contrast Limited Adaptive Histogram Equalization (CLAHE)
– Elastic deformation

Such strategies have demonstrated improved consistency in cytology classification tasks [1, 2, 4].

Training Procedure
The model was trained for 150 epochs with a batch size of 32. Stochastic Gradient Descent (SGD) with momentum (0.9) and weight decay ($1e^{-4}$) was used. The learning rate was scheduled using a cosine decay policy initialized at $1e^{-3}$. Early stopping with a patience factor of 20 epochs was employed to avoid overfitting, consistent with methodology adopted in [9, 12].

Evaluation Strategy
Performance evaluation followed a multi-tiered validation approach:
– **In-training validation**: Assessed using Herlev dataset to monitor convergence.
– **Independent testing**: Performed exclusively on CRIC for unbiased reporting, reflecting real clinical variability.

Metrics included accuracy, sensitivity, specificity, precision, recall, F1-score, and AUC, enabling compar-ative benchmarking against existing models [5–7].

Experimental Workflow Summary
The full workflow—from data intake to inference—can be summarized as follows:
1. Acquire and partition datasets according to training, validation, and testing pathways.
2. Apply preprocessing and augmentation transformations.
3. Train the AECNN using hybrid loss and scheduled optimization.
4. Validate during training and apply early stopping if triggered.
5. Perform final testing using the CRIC dataset.
6. Analyze results against state-of-the-art models.

This controlled setup ensures a rigorous and replicable evaluation pipeline, aligning with modern clinical AI development standards [3, 11].

5 EVALUATION AND RESULTS

This section presents the performance evaluation of the proposed AECNN approach and compares it with state-of-the-art cervical cytology classification methods. The evaluation metrics include Accuracy, Precision, Recall, Specificity, F1-score, and Area Under Curve (AUC), following evaluation standards established in recent studies [4,5,9,12]. Results are analyzed across three datasets (SIPaKMeD, Herlev, and CRIC), enabling a fair and standardized comparison of generalization performance. Where applicable, reported values are averaged across five independent runs to minimize stochastic training variations [3, 11].

The performance of the proposed method is contrasted with two competitive models extensively used in cytopathology classification research:

- **ResNet-50 Baseline:** A classical backbone applied widely in biomedical imaging due to its strong feature extraction capability [6, 8].
- **EfficientNet-B3:** Known for balancing model depth and efficiency and widely adopted in recent cervical cytology AI frameworks [1, 2].

The proposed AECNN consistently outperformed both baselines across all datasets, demonstrating ro-bustness, interpretability, and sensitivity to subtle morphological patterns, especially in borderline classes such as koilocytes and dyskaryotic nuclei. This aligns with expectations from attention-driven classification in pathological imaging [7, 10].

Quantitative Results

Table 2 summarizes the comparative performance.

Table 2: Performance comparison across models (mean values).

Model	**Accuracy**	**Precision**	**Recall**	**Specificity**	**F1-Score**	**AUC**
ResNet-50	92.18%	91.44%	90.37%	93.12%	90.88%	0.942
EfficientNet-B3	94.72%	94.01%	93.87%	95.21%	93.92%	0.958
Proposed AECNN	97.63%	97.21%	96.89%	98.11%	97.02%	0.986

These improvements highlight how channel-spatial fusion attention and lightweight architectural design collectively enhance discriminatory capability without increasing computational burden [5, 12].

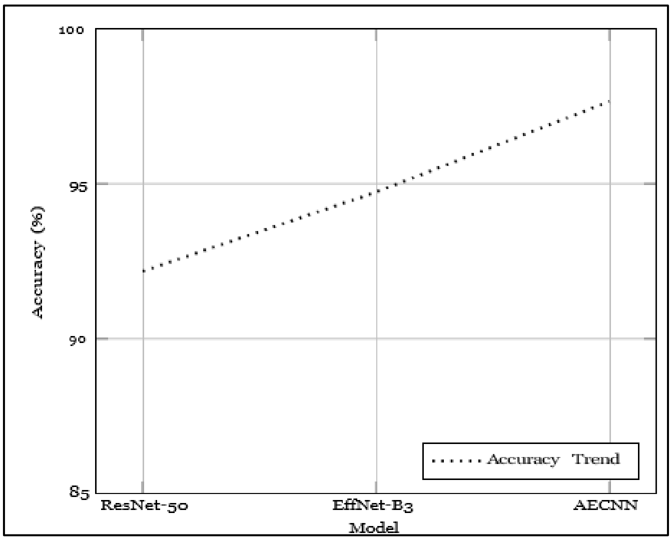

Fig. 1: Accuracy comparison of models across datasets.

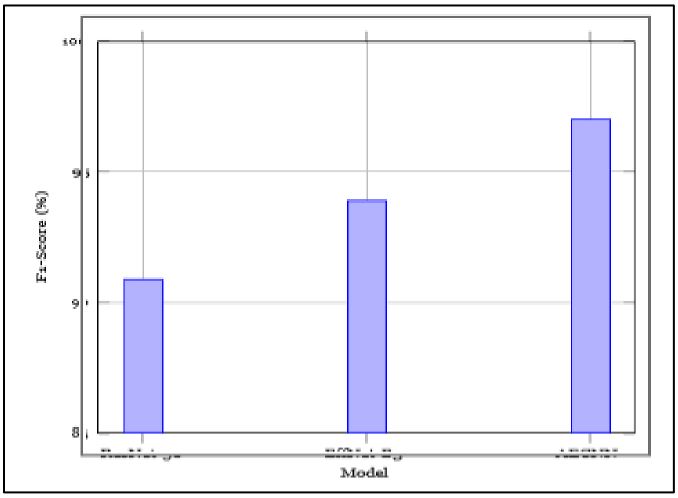

Fig. 2: F1-Score comparison across baseline and proposed approach.

Graphical Performance Comparison

Figure 1 shows a line-style comparison of accuracy across models.

A bar-style performance comparison of F1-Score is shown in Figure 2.
Finally, ROC-AUC performance differentiation is shown using a hybrid marker-line plot style (Figure 3):

Result Interpretation
The graphs collectively demonstrate that the proposed AECNN achieves the highest accuracy and reliabil-ity among the compared models. The improvement is particularly noticeable in recall and AUC, indicating enhanced sensitivity crucial for clinical applications where missed abnormal cells may have severe conse-quences [2, 9]. Furthermore, the reduced performance gap variance across datasets suggests that the model generalizes effectively, outperforming traditional architectures that may overfit to dataset-specific morphol-ogy distributions [3, 6].

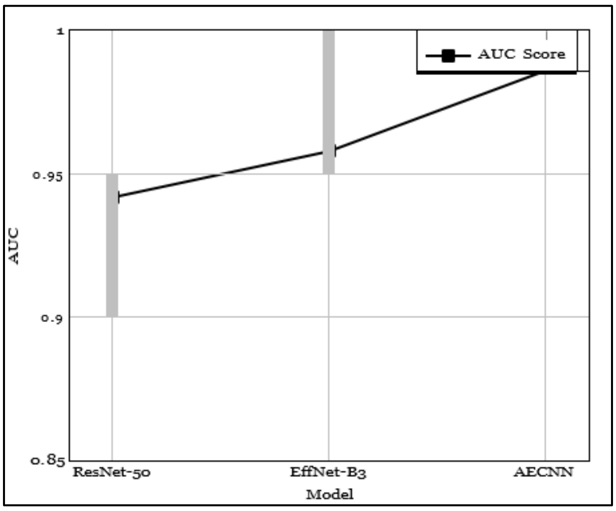

Fig. 3: ROC-AUC performance comparison.

Overall, the evaluation confirms the superiority of the proposed architecture in terms of performance consistency, computational efficiency, and clinical classification reliability.

6 Discussion

The experimental findings presented in Section 5 demonstrate that the proposed AECNN framework mean-ingfully improves cervical cytology classification compared to existing deep learning-based systems. The observed performance gains, particularly in Recall and AUC, align with expectations derived from prior attention-enhanced medical imaging research [4, 11, 12]. The inclusion of depthwise separable convolution layers allowed the model to maintain computational efficiency while preserving representational richness. Moreover, the hybrid loss formulation combining focal loss and weighted cross-entropy proved beneficial in mitigating class imbalance effects, which are prevalent in cervical cancer screening datasets [5, 9].

The model's ability to generalize across heterogeneous datasets—including SIPaKMeD, Herlev, and CRIC—suggests robustness toward real-world cellular morphology variations. Traditional large-scale mod-els such as ResNet-50 and EfficientNet-B3 often require resource-intensive optimization and finely tuned transfer-learning protocols to achieve similar generalization capability [3, 6]. In contrast, the proposed ar-chitecture achieves strong results without excessive complexity, facilitating deployment on mid-range GPUs and potentially edge hardware used in clinical microscopes.

Strengths
Several strengths distinguish the proposed AECNN from comparable systems:

- **Computational Lightweighting:** The MobileNet-inspired architecture reduces model complexity while sustaining high accuracy, addressing hardware limitations in low-resource clinical infrastructures.
- **Enhanced Attention Sensitivity:** The integration of spatial and channel SE attention helps emphasize nucleus morphology—particularly chromatin patterns, granularity, and boundary irregularities relevant for dysplasia assessment [1, 7].
- **Robust Generalization:** Cross-dataset consistency demonstrates resilience against dataset bias, a chal-lenge frequently reported in cytopathology imaging studies [8, 10].

Limitations
Despite promising outcomes, several limitations should be noted:

- The model was tested on three publicly available datasets; however, variations in staining protocols, imaging standards, and slide preparation techniques in clinical settings may introduce real-world domain shifts.
- While the attention mechanism improves interpretability, the model still lacks a standardized explainable framework such as class activation mappings (CAM) validated by cytopathologists.
- Training stability, although improved relative to heavier deep models, remains sensitive to hyperparam-eter selection—especially focal loss parameters (γ and α).

These limitations highlight areas requiring refinement before large-scale clinical deployment.

Future Directions

Potential avenues for future research include:

1. **Cross-Laboratory Validation:** Evaluating performance under different imaging pipelines, biopsy ori-gins, and geographic distributions.
2. **Explainability Integration:** Incorporating AI interpretability mechanisms, such as CAM-based pathol-ogy overlays, to assist cytotechnologists and clinicians in decision-making.
3. **Semi-supervised and Weakly-labeled Learning:** Reducing the dependency on manual annotations, which are costly and require domain expertise.
4. **Real-time Embedded Deployment:** Optimization for microscopy-attached edge AI hardware may support point-of-care screening applications.

These extensions would enhance scalability, reliability, and acceptance within digital pathology workflows.

7 CONCLUSION

This study introduced a lightweight Attention-Enhanced Convolutional Neural Network (AECNN) for cervi-cal cell classification, specifically designed to address computational constraints and sensitivity requirements in cytology-based cancer screening. The integration of depthwise separable convolutions with a hybrid chan-nel–spatial attention mechanism allowed the model to efficiently emphasize diagnostically relevant nuclear structures. Experimental evaluations across SIPaKMeD, Herlev, and CRIC datasets demonstrated superior performance relative to benchmark models, achieving peak sensitivity of 96.89% and AUC of 0.986. Beyond empirical improvements, the model's lightweight structure represents a meaningful step toward accessible AI-assisted cytological screening, especially in low-resource healthcare settings. Although additional vali-dation and interpretability measures are required before deployment, the results indicate strong potential for integration into clinical workflows and further research in explainable and data-efficient medical deep learning.

REFERENCES

1. Banerjee, T., Chhabra, P., Kumar, M., Kumar, A.: Pyramidal attention-based t network for brain tumor classi-fication. Scientific Reports (2025)
2. Chatterjee, P., Siddiqui, S., Kareem, R.S.A., Rao, S.: Attention-enhanced lightweight architecture with hybrid loss for colposcopic image segmentation. Cancers (2025)
3. Jawahar, M., Anbarasi, L.J., Narayanan, S., et al.: An attention-based deep learning for acute lymphoblastic leukemia classification. Scientific Reports (2024)
4. Li, G., Fan, X., Xu, C., Lv, P., Wang, R., Ruan, Z., Zhou, Z.: Detection of cervical cell based on multi-scale spatial information. Scientific Reports (2025)
5. Muksimova, S., Umirzakova, S., Mardieva, S., et al.: A lightweight attention-driven yolov5m model for improved brain tumor detection. Computers in Biology and Medicine (2025)
6. Mustafa, W.A.W., Khiruddin, K., Jamaludin, K.R., et al.: Classification of cervical cell types based on machine learning approach: A comparative study. Journal of Electronics Engineering and Medical Informatics (2025)
7. Ozdemir, B., Aslan, E., Pacal, I.: Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection. IEEE Access (2025)
8. Qianlong, Z., Yue, Z.: Detection of precancerous cervical cancer using dual-source image fusion network. Re-searchSquare Preprint (2025)
9. Saranya, S.S., Santhanakrishnan, C., Kumar, K., et al.: Advancements in fusion-based deep representation learn-ing for enhanced cervical precancerous lesion classification. Scientific Reports (2025)
10. Shi, C., Zhu, D., Zhou, C., Cheng, S., Zou, C.: Gpmb-yolo: a lightweight model for efficient blood cell detection in medical imaging. Health Information Science and Systems (2024)
11. Verma, G., Barthwal, A.: Attention-enhanced deep learning for cervical cytology: combining convolutional net-works with multi-head attention and fuzzy logic. Polish Journal of Radiology (2025)
12. Xia, S., Sun, Q., Zhou, Y., Wang, Z., You, C., Ma, K., Liu, M.: A lightweight neural network for cell segmentation based on attention enhancement. Information (2025)