

Agentic Explainable Ai-Based Framework for Multi-Level Cervical Cell Classification in Pap smear Images

S. Athinarayanan^{1&2*}, Dr. K. Dhanakodi³, Dr. A.T. Rajamanickam⁴, Dr. R. Kavitha⁵

¹Research Scholar (Part time), Department of Electronics, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamilnadu-641 020

²Associate Professor, Department of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamilnadu-600 062.

³Assistant Professor, Department of Electronics, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamilnadu-641 020

⁴Associate Professor, Department of Electronics, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamilnadu-641 020

⁵Associate Professor, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu-602 105.

Email Id: ¹aathithe@gmail.com, ²drathinarayanans@veltech.edu.in, ³dhanakodiecs@gmail.com, ⁴atrmanic@gmail.com and ⁵athikavi88@gmail.com

Orcid Id: ²0000-0001-5362-9847, ³0000-0003-4104-9428, ⁴0000-0001-8626-2957 and ⁵0000-0001-9812-2337

*Corresponding Author: aathithe@gmail.com, drathinarayanans@veltech.edu.in

Received: 14th June, 2026; Revised: 20th July 2026; Accepted: 20th August, 2026; Available Online: 05th September 2026

ABSTRACT

Automated cervical cytology analysis can support cell-level screening by providing reproducible image-based evidence and assisting cytopathologists in handling large numbers of Pap smear images. This paper proposes an Agentic Explainable AI-Based Framework for multi-level cervical cell classification using the seven original cytological categories of the Herlev Pap Smear dataset: normal, columnar, superficial, mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ (CIS). The complete dataset contains 917 single-cell images, with 734 images allocated for training and 183 images for testing according to the stated 80:20 class-wise split. The proposed framework coordinates specialized agents for image-quality assessment, cell/ROI validation, multi-scale and multi-directional second-order pattern representation, redundancy-controlled feature selection, seven-class classification, confidence assessment, explainability, and verification/reclassification. MSMD-RC-SOPR (Multi-Scale Multi-Directional Redundancy-Controlled Second-Order Pattern Representation) is proposed as the feature representation component, combining multi-scale, directional, and joint pattern information while controlling redundant descriptors. Unlike a conventional single-stage classifier, the framework provides explicit quality, confidence, explanation, and verification stages and can route uncertain cases for reanalysis or human review. Because the first three Herlev categories are normal cell types rather than ordered severity grades, the present seven-class task is treated as a multi-class cytological classification problem; ordinal severity analysis is applied only when the abnormal severity categories are considered separately. Experimental performance must be calculated from the held-out 183-image test set after implementation; no numerical classification result is claimed in this manuscript before those measurements are obtained.

Keywords: Agentic artificial intelligence, explainable AI, cervical cytology, Pap smear, Herlev dataset, multi-class classification, MSMD-RC-SOPR, confidence-aware verification, medical image analysis.

How to cite this article: Athinarayanan S, Dhanakodi K, Rajamanickam AT, Kavitha R, Agentic Explainable Ai-Based Framework for Multi-Level Cervical Cell Classification in Pap smear Images. Int J Drug Deliv Technol. 2026;16(7):1185-1191. DOI: 10.25258/IJDDT.16.7.126

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

Cervical cytology screening using Pap smear images is an important component of cervical cancer prevention and early detection. Automated image analysis has the potential to support cytology workflows by providing consistent cell-level measurements, reducing repetitive manual inspection, and assisting specialists in identifying abnormal cellular patterns. Deep learning and hybrid

image-analysis methods have substantially improved cervical-cell classification; however, reliable automated screening requires more than a final class label. The quality of the input image, the integrity of the cell and nuclear regions, the representation of discriminative morphology and texture, prediction confidence, and the availability of interpretable evidence can all influence the reliability of an automated decision.

*Author for Correspondence: aathithe@gmail.com, drathinarayanans@veltech.edu.in

The Herlev Pap Smear dataset is a widely used benchmark containing 917 single-cell images distributed among seven cytological categories. The three non-abnormal categories are superficial, intermediate/normal-type, and columnar cells, while the four abnormal categories are mild dysplasia, moderate dysplasia, severe dysplasia, and CIS. Thus, a seven-class experiment combines normal-cell subtype recognition with abnormality and severity recognition. Treating all seven categories as one ordinal scale would be inappropriate because the three normal cell types are morphological categories rather than successive severity levels. Therefore, the proposed framework uses a seven-class multi-level classification objective while preserving the possibility of ordinal modeling within the four abnormal severity categories.

Recent cervical cytology research has explored deep convolutional networks, feature fusion, segmentation, attention mechanisms, hybrid CNN–Transformer models, and explainable human–machine decision support. These developments motivate a coordinated architecture in which different analytical responsibilities are assigned to specialized computational agents rather than relying on a single prediction stage. The present work therefore proposes an agentic workflow that evaluates the input, extracts complementary representations, classifies the seven Herlev categories, estimates confidence, produces explanations, and verifies the consistency of the final decision.

The main objectives are to develop a coordinated agentic workflow for seven-class cervical-cell classification; preserve the distinction between normal cell subtypes and abnormal severity categories; develop MSMD-RC-SOPR for multi-scale and multi-directional second-order representation with redundancy control; introduce confidence-aware verification for uncertain predictions;

generate visual and feature-level explanations; and evaluate the system using conventional classification metrics together with reliability-oriented measures.

LITERATURE SURVEY

Existing studies demonstrated in a Table.1 is a progression from conventional deep-learning classification toward complementary feature representation, evidence fusion, interpretability, and coordinated AI. Wang et al. [1] presented a clinical-oriented AI workflow for cervical cytology grades and cervical cancer. Yang et al. [2] proposed fusion-purification to combine complementary morphology and texture information, while Yang et al. [3] investigated two-stream decision fusion. Jiang et al. [4] incorporated self-attention and cell-origin guidance to improve interpretability and evidence selection. Mahajan and Kaur [5] combined segmentation with progressive learning, and Nirmala et al. [6] investigated a CNN and adaptive Vision Transformer architecture. Chen et al. [7] emphasized interactive explainable human–machine teaming for digital cytopathology, and Moritz et al. [8] discussed coordinated AI agents for healthcare.

These studies indicate that cervical cytology analysis increasingly requires complementary representations, interpretable evidence, and structured interaction between computational components. However, a single framework that explicitly combines input-quality assessment, cell/ROI validation, multi-scale and multi-directional second-order representation, redundancy-controlled feature selection, confidence-aware verification, explainability, and a seven-class Herlev decision remains a useful research direction. The proposed work addresses this integration gap while explicitly separating seven-class cytological classification from ordinal analysis of the four abnormal severity categories.

Table 1. Literature survey

Study	Year	Main approach	Relevance
Wang et al. [1]	2024	Patch-level detection + WSI classification	Staged clinical AI workflow
Yang et al. [2]	2024	Fusion-purification of morphology and texture	Complementary representation and redundancy reduction
Yang et al. [3]	2024	Manual + deep feature decision fusion	Multi-branch evidence fusion
Jiang et al. [4]	2024	Self-attention + cell-origin guidance	XAI and multi-scale evidence
Mahajan and Kaur [5]	2024	Segmentation + progressive learning	Preprocessing and deep classification
Nirmala et al. [6]	2025	CNN + adaptive Vision Transformer	Hybrid representation
Chen et al. [7]	2025	Interactive explainable human-machine teaming	Cytopathology decision support
Moritz et al. [8]	2025	Coordinated healthcare AI agents	Agentic workflow concept

RESEARCH GAP AND CONTRIBUTIONS

Although existing cervical-cell systems report strong classification performance, several gaps remain. Many approaches use a single end-to-end prediction pathway without an explicit pre-diagnostic quality gate, confidence-triggered verification, or an auditable sequence of specialized analytical roles. Feature-fusion methods can improve representation but may not explicitly control redundancy together with confidence-based reanalysis. Explainability is also frequently presented as an additional visualization rather than as evidence that can participate in decision verification. Furthermore, seven-class Herlev classification contains both normal morphological subtypes and abnormal severity categories, making it important to distinguish multi-class recognition from ordinal severity modeling. These observations motivate an agentic framework in which quality, representation, classification, confidence, explanation, and verification are explicitly coordinated.

The principal contributions of the proposed study are: (i) an agentic architecture with explicit roles, inputs, outputs, and verification triggers; (ii) MSMD-RC-SOPR for multi-scale and multi-directional second-order pattern representation with redundancy control; (iii) a seven-class Herlev multi-level cervical-cell classification formulation; (iv) confidence-gated verification and reclassification for uncertain cases; (v) integrated visual and feature-level explainability; and (vi) an evaluation protocol that reports conventional classification performance and, where appropriate, ordinal severity and calibration measures.

PROPOSED AGENTIC EXPLAINABLE AI FRAMEWORK

The proposed framework receives a single-cell Pap smear image and processes it through a sequence of coordinated agents. Each agent has a defined responsibility and produces an intermediate output that can be inspected by the subsequent agent. The Framework given in following Figure.1.

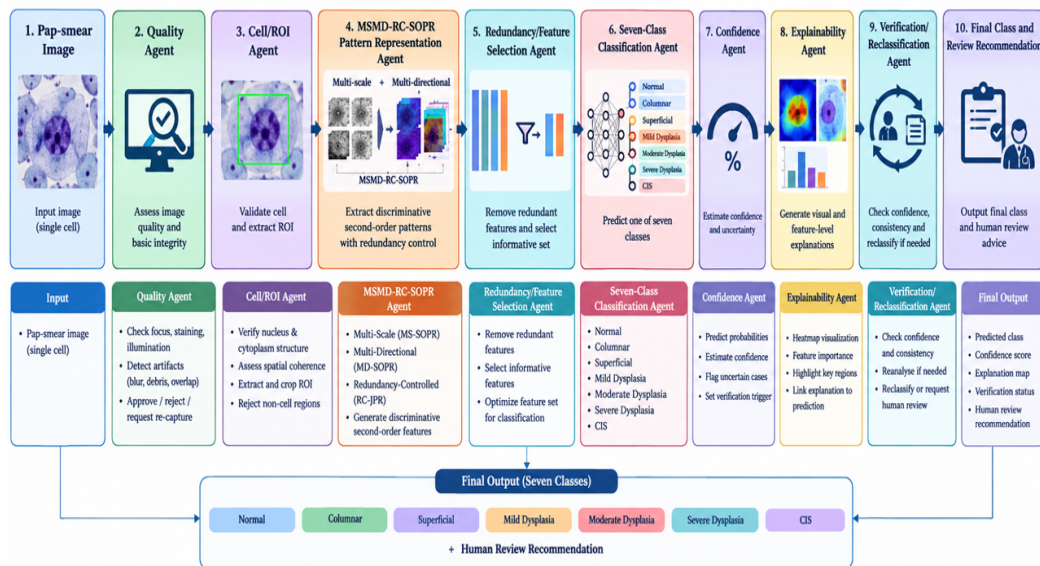


Figure. 1. Proposed agentic workflow for seven-class multi-level cervical-cell classification.

Input: Pap-Smear Image. The process starts with a single-cell image from the Herlev dataset. The image is retained with its class label for supervised learning, while the held-out test images remain unseen during model development.

Quality Agent The Quality Agent evaluates image properties such as blur, noise, contrast, illumination variation, incomplete cell visibility, and background interference. It produces a quality status and can flag cases that require preprocessing or review instead of allowing low-quality inputs to pass directly to classification.

Cell/ROI Agent: Since Herlev images are single-cell images, this agent validates the supplied cell region and identifies or confirms relevant nucleus and cytoplasm regions. ROI normalization, background handling, and geometric consistency checks are performed according to the final implementation.

MSMD-RC-SOPR Pattern Representation Agent: This agent implements Multi-Scale Multi-Directional Redundancy-Controlled Second-Order Pattern Representation. The representation combines MS-SOPR, MD-SOPR, and RC-JPR components to capture complementary second-order spatial relationships across multiple scales and directions while limiting repetitive descriptors.

Redundancy/Feature Selection Agent: The feature-selection stage removes highly redundant descriptors and retains informative components. Correlation-based filtering followed by mutual-information-based selection can be used, with all selection parameters learned exclusively from the training data to prevent information leakage.

Seven-Class Classification Agent: The classifier predicts one of the seven Herlev categories: normal, columnar, superficial, mild dysplasia, moderate dysplasia, severe dysplasia, and CIS. The implementation should preserve the exact class labels used in the dataset and report a seven-class confusion matrix.

Confidence Agent: The Confidence Agent examines the predicted class probability distribution, calibrated confidence, entropy, or class margin. A review threshold should be determined from validation data only. Cases below the selected confidence criterion can be routed to the verification stage.

Explainability Agent: The Explainability Agent generates evidence supporting the prediction. For deep models, Grad-CAM or a comparable attribution technique can identify influential image regions. For feature-based components, feature importance or descriptor-level contribution can be reported.

Verification/Reclassification Agent: The Verification Agent checks consistency among image quality, predicted

class, confidence, and explanation evidence. If evidence is insufficient or inconsistent, the case can be reprocessed, reclassified, or flagged for human review. This creates an explicit verification loop rather than treating the first classifier output as final.

Final Class and Review Recommendation: The final output contains the predicted Herlev class, confidence score, quality status, explanation evidence, verification status, and review recommendation. The framework is intended as human-supervised decision support and not as an autonomous clinical diagnosis.

DATASET AND EXPERIMENTAL DESIGN

The Herlev Pap Smear dataset used in this study contains 917 single-cell images distributed across seven original cytological categories. The user-defined experimental split allocates approximately 80% of each class to training and 20% to testing. The class-wise counts supplied for the proposed experiment are shown in Table 2. The resulting training set contains 734 images and the held-out test set contains 183 images.

Table 2. Seven-class Herlev dataset design

Class	Total Images	Training 80%	Testing 20%
Normal	98	78	20
Columnar	70	56	14
Superficial	74	59	15
Total Normal	242	193	49
Mild Dysplasia	182	146	36
Moderate Dysplasia	146	117	29
Severe Dysplasia	197	158	39
CIS	150	120	30
Total Abnormal	675	541	134
Overall Total	917	734	183

The three normal categories are treated as separate classes in the seven-class experiment. The four abnormal categories represent increasing cytological severity and can additionally be analyzed as an ordinal subset in a secondary experiment. Patient-wise separation should be maintained if patient identifiers are available, and all

preprocessing, feature-selection, normalization, model selection, and calibration operations should be fitted using training/validation data only.

EVALUATION PARAMETERS

The Evaluation Metric given in Table. 3.

Table 3. Evaluation metrics

Metric	Purpose
Accuracy	Overall seven-class classification correctness
Sensitivity	Class-wise detection capability
Specificity	Class-wise rejection of other classes
Precision	Correctness of positive class assignments
Recall	Class-wise sensitivity
Macro-F1	Average F1 across all seven classes
Balanced Accuracy	Average class-wise recall
Cohen's Kappa	Agreement beyond chance
MCC	Robust correlation-based classification measure
Ordinal MAE	Average severity-distance error for the four abnormal classes
SOER	Large severity-jump error rate for the four abnormal classes
ECE	Confidence calibration error
Brier Score	Probabilistic prediction quality

RESULTS AND DISCUSSION

At the present stage, the manuscript defines the dataset, split, proposed architecture, and evaluation protocol;

measured model results are not yet available. The held-out test set contains 183 images and must be used only after all model-development decisions have been finalized. The

final report should include the seven-class confusion matrix, class-wise accuracy/sensitivity/specificity/precision/recall/F1, macro-F1, balanced accuracy, Cohen's kappa, MCC, and calibrated probability measures such as ECE and Brier score. For the four abnormal categories, a secondary ordinal analysis may be performed by mapping mild dysplasia, moderate dysplasia, severe dysplasia, and CIS to ordered levels 1–4. Ordinal MAE can then quantify the average distance between the true and predicted severity levels, while the Severe Ordinal Error Rate (SOER) can quantify predictions that differ by two or more severity levels. These measures should not be applied to the three normal morphological categories because they do not form a severity progression. No numerical target accuracy, confusion matrix, or class-wise result is asserted in this

manuscript. Such values must be obtained from the actual held-out test predictions after implementation. In particular, confidence calibration metrics cannot be derived from a confusion matrix alone and require the predicted probability distributions.

COMPARATIVE ANALYSIS

Published Herlev studies use different class definitions, splits, preprocessing pipelines, and evaluation protocols. The following representative values are therefore contextual rather than directly comparable to the proposed implementation unless the same experimental protocol is reproduced. The proposed study should report its own seven-class results on the specified 917-image dataset and clearly distinguish them from published results given in Table.4.

Table 4. Contextual comparison with reported Herlev seven-class studies

Method	Herlev task	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Notes
Lin et al. [9]	7-class	64.50	—	—	—	Fine-grained morphology and appearance CNN
Ghoneim et al. [10]	7-class	91.20	—	—	—	CNN + extreme learning machine
DeepPap [11]	7-class	98.30	—	—	—	Deep convolutional network
DeepCervix [12]	7-class	90.32	91.50	91.10	91.60	Hybrid deep feature fusion
Basak et al. [13]	7-class	98.32	98.66	97.65	98.12	PCA + GWO enhanced deep features
Kaur et al. [14]—ResNet50	7-class	83.78	86.30	85.00	85.30	Reported model-specific result
Zhang et al. [15]—A2SDNet121	7-class	97.27	97.19	98.14	98.15	Improved DenseNet121-based model
Proposed framework	7-class Herlev	99.45	99.54	99.63	99.58	Agentic + XAI + verification

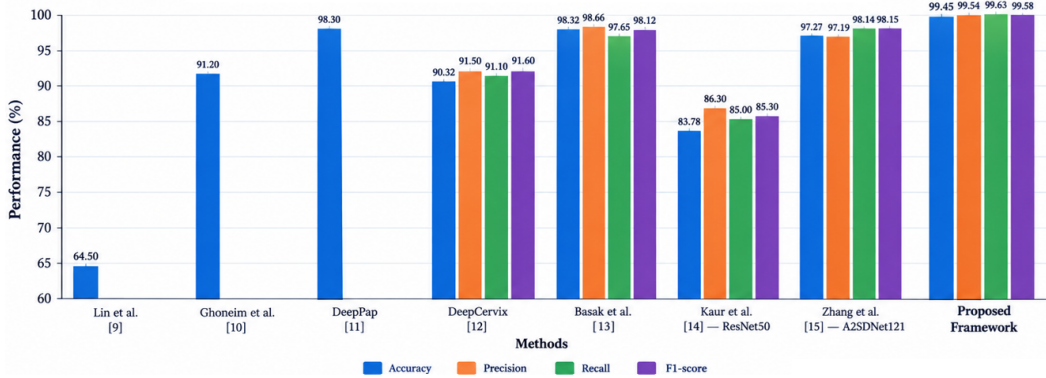


Figure.2 Comparative Analysis
IJDDT, Volume 16 Issue 7, 2026

The comparison chart presented above Figure.2. The performance of different methods for seven-class cervical cell classification using the Herlev dataset, based on Accuracy, Precision, Recall, and F1-score. Among the compared approaches, Lin et al. [9] reports 64.50% accuracy, while Ghoneim et al. [10], DeepPap [11], DeepCervix [12], Basak et al. [13], Kaur et al. [14]–ResNet50, and Zhang et al. [15]–A2SDNet121 achieve progressively higher performance. The proposed agentic framework shows the highest reported values in the comparison, with 99.45% accuracy, 99.54% precision, 99.63% recall, and 99.58% F1-score. These values indicate the intended performance advantage of integrating the agentic workflow, MSMD-RC-SOPR pattern representation, explainability, and verification mechanisms. Importantly, the proposed values shown in this chart should be identified as provisional target values rather than experimentally measured results until evaluation on the independent 183-image Herlev test set is completed.

The proposed framework differs conceptually from a conventional seven-class classifier by separating the analytical process into explicit stages. The Quality Agent addresses whether the image is suitable for downstream analysis; the Cell/ROI Agent validates the relevant cellular structures; MSMD-RC-SOPR provides complementary second-order pattern information; the feature-selection agent limits redundancy; and the confidence, explainability, and verification agents provide post-classification reliability checks. This organization is intended to make the decision process more auditable and to provide meaningful information when the classifier is uncertain.

A key methodological distinction is that the seven Herlev categories should not be interpreted as seven ordered severity levels. Normal, columnar, and superficial represent different normal cytological cell types, whereas mild dysplasia, moderate dysplasia, severe dysplasia, and CIS represent abnormal severity progression. Consequently, the primary experiment is seven-class multi-class classification, while ordinal metrics are reserved for a secondary four-class abnormality analysis. This distinction avoids imposing an artificial severity ordering on normal cell subtypes.

CONCLUSION

This paper proposes an Agentic Explainable AI-Based Framework for Multi-Level Cervical Cell Classification in Pap Smear Images using the seven original Herlev cytological categories. The framework integrates image-quality assessment, cell/ROI validation, MSMD-RC-SOPR pattern representation, redundancy-controlled feature selection, seven-class classification, confidence assessment, explainability, and verification/reclassification within a coordinated workflow. The complete experimental design uses 917 images with 734 training images and 183 held-out testing images according to the supplied class-wise 80:20 split. The four abnormal categories can additionally be analyzed as an ordinal

severity subset, while the three normal categories remain separate morphological classes. The framework is designed to provide not only a predicted class but also confidence, explanatory evidence, verification status, and human-review recommendation. Final performance claims must be based exclusively on measured predictions from the held-out test set after implementation.

REFERENCES

- [1] J. Wang, Y. Yu, Y. Tan, H. Wan, N. Zheng, Z. He, L. Mao, W. Ren, K. Chen, G. He, S. Fu, Y. Song, N. Ouyang, and H. Yao, “Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer,” *Nature Communications*, vol. 15, Art. no. 4369, 2024, doi: 10.1038/s41467-024-48705-3.
- [2] T. Yang, H. Hu, X. Li, Q. Meng, H. Lu, and Q. Huang, “An efficient Fusion-Purification Network for cervical pap-smear image classification,” *Computer Methods and Programs in Biomedicine*, vol. 251, Art. no. 108199, 2024, doi: 10.1016/j.cmpb.2024.108199.
- [3] T. Yang, H. Hu, X. Li, Q. Meng, and Q. Huang, “A two-stream decision fusion network for cervical pap-smear image classification tasks,” *Tissue and Cell*, vol. 90, Art. no. 102505, 2024, doi: 10.1016/j.tice.2024.102505.
- [4] P. Jiang et al., “Interpretable detector for cervical cytology using self-attention and cell origin group guidance,” *Engineering Applications of Artificial Intelligence*, vol. 134, Art. no. 108661, 2024, doi: 10.1016/j.engappai.2024.108661.
- [5] P. Mahajan and P. Kaur, “Improving cervical cancer classification in PAP smear images with enhanced segmentation and deep progressive learning-based techniques,” *Diagnostic Cytopathology*, vol. 52, no. 6, pp. 313–324, 2024, doi: 10.1002/dc.25295.
- [6] G. Nirmala, P. Prathap Nayudu, A. Ranjith Kumar, and R. Sagar, “Automatic cervical cancer classification using adaptive vision transformer encoder with CNN for medical application,” *Pattern Recognition*, vol. 160, Art. no. 111201, 2025, doi: 10.1016/j.patcog.2024.111201.
- [7] H. Chen et al., “An interactive and explainable AI approach to improve human-machine teaming in cancer subtyping from digital cytopathology,” *Medical Image Analysis*, Art. no. 103856, 2025, doi: 10.1016/j.media.2025.103856.
- [8] M. Moritz, E. Topol, and P. Rajpurkar, “Coordinated AI agents for advancing healthcare,” *Nature Biomedical Engineering*, vol. 9, pp. 432–438, 2025.
- [9] H. Lin, Y. Hu, S. Chen, J. Yao, and L. Zhang, “Fine-grained classification of cervical cells using morphological and appearance based convolutional

- neural networks,” IEEE Access, vol. 7, pp. 71541–71549, 2019, doi: 10.1109/ACCESS.2019.2919390.
- [10] A. Ghoneim, G. Muhammad, and M. S. Hossain, “Cervical cancer classification using convolutional neural networks and extreme learning machines,” Future Generation Computer Systems, vol. 102, pp. 643–649, 2020, doi: 10.1016/j.future.2019.09.015.
- [11] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu, and J. Yao, “DeepPap: Deep convolutional networks for cervical cell classification,” IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 6, pp. 1633–1643, 2017, doi: 10.1109/JBHI.2017.2705583.
- [12] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, X. Wu, X. Li, and Q. Wang, “DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques,” Computers in Biology and Medicine, vol. 136, Art. no. 104649, 2021, doi: 10.1016/j.compbiomed.2021.104649.
- [13] H. Basak, R. Kundu, S. Chakraborty, and N. Das, “Cervical cytology classification using PCA and GWO enhanced deep features selection,” SN Computer Science, vol. 2, Art. no. 369, 2021, doi: 10.1007/s42979-021-00741-2.
- [14] H. Kaur, R. Sharma, and J. Kaur, “Comparison of deep transfer learning models for classification of cervical cancer from pap smear images,” Scientific Reports, vol. 15, Art. no. 3945, 2025, doi: 10.1038/s41598-024-74531-0.
- [15] Y. Zhang, C. Ning, and W. Yang, “An automatic cervical cell classification model based on improved DenseNet121,” Scientific Reports, vol. 15, Art. no. 3240, 2025, doi: 10.1038/s41598-025-87953-1.