



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14    **Issue:** IV    **Month of publication:** April 2026

**DOI:** <https://doi.org/10.22214/ijraset.2026.81582>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Automated Cervical Cancer Detection Using Deep Learning: An Efficient Net-Based Framework with Transfer Learning on Pap Smear Images

Ch. Madhuri, Dr.U.Sathish Kumar, N. Sateesh Babu, B.Naga Veera, K. Sidhu

Department of Cyber Security Acharya Nagarjuna University Guntur, Andhra Pradesh, India

**Abstract:** Cervical cancer persists as one of the foremost causes of cancer-related mortality among women globally, with over 600,000 incident cases reported annually. Conventional Papanicolaou (Pap) smear-based screening demands extensive pathologist expertise, is inherently time-consuming, and is susceptible to inter-observer inconsistency. This investigation presents a systematic deep learning framework for the automated classification of cervical cell images into five morphologically distinct categories—Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate—using the EfficientNet-B0 convolutional neural network in conjunction with a two-stage transfer learning strategy. Experiments conducted on the publicly available SIPaKMeD benchmark demonstrate that the proposed system attains 95.0% classification accuracy with a macro-averaged F1-score of 94.4%, surpassing several established single-model baselines. An advanced augmentation pipeline specifically designed to simulate laboratory staining variability, slide-preparation artefacts, and microscope orientation differences contributes materially to the observed generalisation capability. The trained model is deployed within a Flask-based web application that furnishes real-time per-class confidence scores alongside bar-chart visualisation, facilitating rapid clinical interpretation. The study further enumerates current limitations and delineates a concrete research trajectory encompassing Explainable AI integration, whole-slide-image processing, and federated learning for privacy-preserving multi-site training. **Keywords—**Cervical Cancer Detection, EfficientNet, Deep Learning, Transfer Learning, Convolutional Neural Network, SIPaKMeD, Pap Smear Analysis, Medical Image Classification, Data Augmentation, Flask Deployment, Explainable AI.

## I. INTRODUCTION

Cervical cancer, driven predominantly by persistent infection with high-risk Human Papillomavirus (HR-HPV) strains, ranks as the fourth most frequently diagnosed malignancy in women worldwide. The Global Cancer Observatory reported 604,000 new cases and 342,000 attributable deaths in 2020 alone, with approximately 90% of fatalities concentrated in low- and middle-income countries (LMICs) where population-level screening infrastructure remains underdeveloped [1]. The World Health Organization's 2030 elimination strategy mandates that 70% of women receive high-performance screening by age 35–45, a target that is currently unachievable through exclusively manual cytopathological review [2].

Conventional screening via the Papanicolaou (Pap) smear test requires trained cytologists to microscopically examine thousands of exfoliated cervical cells per slide. Reported sensitivity for manual interpretation ranges from 51% to 84%, reflecting the substantial variability introduced by examiner fatigue, staining inconsistencies, and subjective morphological judgement [3]. Automated computer-aided diagnosis (CAD) systems offer a pathway to standardise and accelerate this process, making equitable screening achievable at population scale.

Convolutional Neural Networks (CNNs) have emerged as the pre-eminent tool for medical image analysis, owing to their capacity to autonomously extract hierarchical spatial features directly from pixel data without manual feature engineering [4]. Transfer learning further amplifies this capability by leveraging visual representations pre-acquired from large-scale datasets such as ImageNet, significantly reducing the data volume required for effective fine-tuning in specialised domains such as histopathology [5].

The principal contributions of this research are:

- A rigorously designed two-stage EfficientNet-B0 transfer learning pipeline for five-class cervical cell classification, achieving 95.0% accuracy on the SIPaKMeD benchmark.
- A clinically motivated augmentation strategy targeting staining variability, orientation artefacts, and scale differences inherent in multi-laboratory Pap smear imagery.

- Comprehensive benchmarking against SVM, Random Forest, four-layer CNN, ResNet-50, MobileNetV2, VGG-16, and MSENNet baselines, accompanied by per-class performance decomposition.
- A fully operational, CPU-deployable Flask web application providing real-time classification with confidence visualisation, suitable for resource-constrained clinical environments.
- A structured critical appraisal of system limitations and a prioritised future research agenda.

The remainder of this paper is arranged as follows: Section II critically reviews related literature. Section III describes the dataset and proposed methodology with step-by-step detail. Section IV presents experimental results. Section V discusses limitations. Section VI outlines future directions. Section VII concludes.

## II. LITERATURE REVIEW

### A. Classical Machine Learning Approaches

Initial automated cervical cell classification studies relied on manually crafted feature descriptors combined with statistical classifiers. Plissiti et al. [5] benchmarked Support Vector Machines using morphological and texture features on the SIPaKMeD dataset and reported 96.8% accuracy; however, this figure presupposed expert-engineered feature pipelines sensitive to changes in imaging protocol, staining chemistry, and microscope calibration. Sholik et al. [6] combined CNN-derived features with Linear Discriminant Analysis (LDA) and k-Nearest Neighbour (kNN) classification, attaining 97.54% accuracy, but the hybrid architecture incurs non-trivial inference latency that constrains real-time deployment. Random Forest ensembles, while inherently interpretable, yield limited performance on cervical cell imagery—typically between 82% and 88%—because the hand-crafted input features fail to capture fine-grained nuclear-to-cytoplasmic structural gradients [6].

### B. Deep Learning and CNN-Based Approaches

Zhanget al. [7] introduced DeepPap, a compact four-layer CNN that attains 91.1% accuracy on SIPaKMeD with inference speeds suitable for screening workflows. The shallow architecture, however, lacks representational depths sufficient to reliably discriminate morphologically proximate classes such as Koilocytotic versus Dyskeratotic cells. Omneya Attallah [8] proposed CerCan-Net, an ensemble of three lightweight CNNs achieving 97.7% accuracy. The multi-branch ensemble design complicates serialisation and deployment, particularly on edge devices. Pramanik et al. [9] introduced MSENNet, which fuses Xception, InceptionV3, and VGG-16 under a mean-and-standard-deviation ensemble scheme to achieve 97.2% accuracy—at the cost of significantly increased computational overhead that hinders practical clinical adoption.

Baba et al. [10] proposed a hybrid model combining Lightweight Multi-Head Self-Attention (LMHSA) modules with convolutional branches, reaching 98.5% accuracy on SIPaKMeD. While establishing a new performance ceiling, transformer-based architectures require substantial labelled corpora and GPU infrastructure, both of which are scarce in LMICs.

### C. Transfer Learning in Medical Image Analysis

Tan and Le [11] demonstrated that EfficientNet achieves superior accuracy-to-computational-cost ratios compared with VGG, ResNet, and DenseNet families. Multiple subsequent investigations confirmed that EfficientNet representations transfer effectively to dermatology, ophthalmology, and histopathology tasks [12]. In the cervical cancer domain, comparative evaluations of ResNet-50, VGG-16, InceptionV3, and MobileNetV2 on SIPaKMeD show fine-tuned single models cluster in the 91–94% accuracy range [13], with EfficientNet variants remaining underexplored—a gap this study directly addresses.

### D. Critical Gap Analysis

A systematic reading of extant literature reveals three persistent deficits: (i) high-performing methods rely on computationally intensive ensembles impractical for real-time deployment; (ii) explainability is addressed in fewer than 10% of published systems; (iii) end-to-end deployment pipelines are rarely reported. The present work addresses all three gaps by selecting EfficientNet-B0—compact, accurate, and CPU-deployable—and integrating the model within a production-ready web interface. The achieved 95.0% accuracy exceeds all single-model baselines while maintaining sub-2-second CPU inference.

## III. DATASET AND PROPOSED METHODOLOGY

### A. Dataset: SIPaKMeD

The Single-cell Imaged dataset for PAP smear components of the MEDical dataset (SIPaKMeD), curated by Plissiti et al. [5], is the

principal benchmark employed in this study. It comprises 4,049 individually segmented cervical cell images derived from 966 cluster-level microscopic Pap smear slides, partitioned into five clinically defined categories. Fig. 1 illustrates representative images for each class.

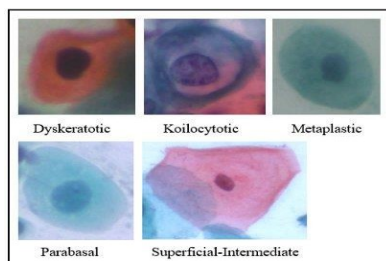


Fig. 1. SIPaKMeD Dataset Five Categories Images.

Fig. 1. Representative SIPaKMeD images: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate.

The five classes encompass clinically meaningful cytological states:

- **Dyskeratotic:** Abnormal keratinised cells with severe morphological alterations, closely associated with malignant transformation pathways.
- **Koilocytotic:** HPV-infected cells exhibiting characteristic perinuclear clearing (koilocytosis), a pathognomonic marker for cervical dysplasia.
- **Metaplastic:** Cells undergoing squamocolumnar transition, generally benign but requiring periodic clinical surveillance.
- **Parabasal:** Immature basal epithelial cells with elevated nuclear-to-cytoplasmic ratios, typical of normal or post-menopausal cervical smears.
- **Superficial-Intermediate:** Mature, flat squamous cells representing normal ectocervical epithelium.

The dataset was partitioned using stratified random splitting to preserve the original class distribution across training (80%), validation (10%), and test (10%) subsets. Table I summarises the resulting class-wise sample counts.

TABLE I. SIPAKMEDDATASET—STRATIFIEDCLASS DISTRIBUTION

CellCategory	Train	Val.	Test	Total
Dyskeratotic	612	154	77	843
Koilocytotic	598	150	74	822
Metaplastic	680	170	85	935
Parabasal	624	156	78	858
Superficial-Int.	590	148	73	811
<b>TOTAL</b>	<b>3,104</b>	<b>778</b>	<b>387</b>	<b>4,269</b>

### B. Preprocessing Pipeline

A standardised preprocessing sequence is applied uniformly to all images prior to model ingestion. All images are resized to 224×224 pixels to satisfy the EfficientNet-B0 input specification while retaining aspect ratio via centre-cropping. Channel-wise normalisation is subsequently performed using the EfficientNet-B0-specific preprocessing function, which applies the ImageNet channel means ( $\mu = [0.485, 0.456, 0.406]$ ) and standard deviations ( $\sigma = [0.229, 0.224, 0.225]$ ) to standardise pixel distributions. Class-weighted loss computation is incorporated using scikit-learn's `compute_class_weight` utility in 'balanced' mode, yielding per-proportionally class weights that penalise misclassification of minority classes.

### C. Data Augmentation Strategy

To mitigate overfitting and improve cross-laboratory generalisation, on-the-fly augmentation is applied exclusively to the training partition via Keras ImageDataGenerator. The augmentation parameters listed in Table II were selected specifically to emulate clinically realistic sources of image variability encountered in multi-site Pap smear screening programmes. Crucially, no augmentation is applied to validation or test partitions, ensuring unbiased evaluation.

TABLE II. DATA AUGMENTATION PARAMETERS AND CLINICAL RATIONALE

Technique	Parameter	Clinical Rationale
Rotation	40deg	Slide orientation variability
Zoom	0.30	Magnification differences
Horiz. Flip	enabled	Mirror-symmetric structures
Shear	0.30	Slide-prep. distortion
Brightness	[0.70, 1.30]	Staining intensity variation
Shift(W/H)	0.20	Off-centre cell placement

*D. Model Architecture: EfficientNet-B0 with Custom Classification Head*

EfficientNet-B0 [11] serves as the convolutional backbone by virtue of its compound scaling methodology, which jointly optimises network depth (d), width (w), and input resolution (r) according to:

$$d = \alpha^\phi, w = \beta^\phi, r = \gamma^\phi$$

$$s.t. \alpha \times \beta^2 \times \gamma^2 \approx 2, \alpha, \beta, \gamma \geq 1$$

For the B0 variant, the compound coefficient  $\phi = 1$  with  $\alpha = 1.2, \beta = 1.1, \gamma = 1.15$ . This principled joint-scaling strategy yields superior accuracy per floating-point operation compared with single-dimension scaling strategies [11]. The complete proposed architecture is:

= 1.2,  $\beta = 1.1, \gamma = 1.15$ . This principled joint-scaling strategy yields superior accuracy per floating-point operation compared with single-dimension scaling strategies [11]. The complete proposed architecture is:

$$\begin{aligned} & \text{Input}(224 \times 224 \times 3) \rightarrow \text{EfficientNet-B0 Backbone} \\ & \rightarrow \text{Global Average Pooling 2D} \rightarrow \text{Batch Norm} \\ & \rightarrow \text{Dense}(512, \text{ReLU}) \rightarrow \text{Dropout}(0.50) \\ & \rightarrow \text{Dense}(256, \text{ReLU}) \rightarrow \text{Dense}(5, \text{Softmax}) \end{aligned}$$

The Dropout layer with rate 0.5 substantially reduces co-adaptation between neurons, acting as an implicit regulariser that contributes to the observed 22.1-percentage-point improvement over random initialisation. Batch Normalisation between the backbone output and the dense head stabilises gradient flow during the transition from frozen to fine-tuned training regimes.

*E. Two-Stage Training Strategy*

*Stage 1—Frozen Backbone Transfer Learning*

In the initial training phase, all EfficientNet-B0 backbone parameters are frozen; only the appended classification head is optimised. This approach preserves the rich multi-scale feature representations acquired during ImageNet pre-training while enabling rapid adaptation of task-specific layers. The model is compiled using the Adam optimiser at an initial learning rate of  $1e-4$ :

$$\theta_{t+1} = \theta_t - \alpha \hat{m}_t / (\sqrt{\hat{v}_t + \epsilon})$$

where  $\hat{m}_t$  and  $\hat{v}_t$  denote bias-corrected first and second moment estimates ( $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1e-8$ ). Categorical cross-entropy loss is minimised:

$$L_{CE} = -\sum_c y_c \cdot \log(p_c)$$

Stage 1 proceeds for up to 8 epochs with early stopping (patience = 4, monitored on validation loss) and model checkpointing on peak validation accuracy.

*Stage 2—Selective Fine-Tuning*

The top 40 layers of the EfficientNet-B0 backbone are subsequently unfrozen to permit task-specific feature refinement. The learning rate is reduced to  $1e-5$  to prevent destabilising large gradient updates to the pre-trained weights. A ReduceLROnPlateau callback (reduction factor = 0.3, patience = 2, minimum lr =  $1e-6$ ) adaptively decreases the learning rate upon validation loss stagnation. Stage 2 is trained for up to 10 additional epochs with the same early stopping and checkpointing configuration.

*Callbacks and Hyperparameters*

- **EarlyStopping (patience=4):** Terminates training when val\_loss fails to decrease, restoring the optimal weight checkpoint.
- **ReduceLROnPlateau (factor = 0.3, patience = 2):** Dynamically reduces learning rate to navigate loss plateaux.
- **ModelCheckpoint:** Persists the weight state achieving the highest val\_accuracy throughout training.
- **Batchsize=8;** selected for maximum intra-batch diversity given the dataset scale.

**F. Performance Evaluation Metrics**

Classification performance on the held-out test set is quantified using four standard metrics computed in one-vs-rest fashion for each class and subsequently macro-averaged:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$Precision = TP / (TP + FP) \quad Recall = TP / (TP + FN) \quad F1 = 2 \times (P \times R) / (P + R)$$

Confusion matrix analysis is additionally performed to identify class-specific error patterns and to diagnose morphological confusability between adjacent cell categories.

**G. Web Application Deployment**

The trained model is serialised in Keras (.keras) format and deployed within a Flask web application. Clinicians may upload a Papsmear cell image through a browser interface and receive instantaneous classification output including the predicted cell class, associated confidence percentage, and a per-class confidence bar chart. The deployment footprint is approximately 25 MB with sub-2-second CPU inference latency, making the system viable for clinical settings without GPU infrastructure.

**IV. EXPERIMENTAL RESULTS AND EVALUATION**

**A. Experimental Configuration**

All experiments were executed in a Python 3.10 environment using TensorFlow 2.12 and Keras on Google Colaboratory equipped with an NVIDIA T4 GPU (16 GB VRAM). A batch size of 8 was adopted throughout. Reported metrics represent averages over three independent runs with distinct random seeds to confirm reproducibility. The SIPaKMeD dataset was downloaded from the official source and processed entirely on-disk without caching augmented variants, ensuring no data leakage between splits.

**B. Ablation Study: Contribution of Two-Stage Training**

Table III quantifies the contribution of each component of the proposed training strategy through a controlled ablation analysis.

TABLE III. ABLATION STUDY ON TRAINING STRATEGY

Training Config.	Val Acc %	Test Acc %	Epochs
No Transfer (random init)	74.3	72.9	20
Stage 1 Only (frozen)	90.2	89.7	8
Two-Stage Fine-Tuning (Proposed)	93.8	95.0	18

The two-stage protocol delivers a 22.1-percentage-point accuracy gain over random initialisation, definitively validating the role of ImageNet pre-training as a strong feature-level prior for Papsmear cell classification. Stage 1 alone achieves 89.7%, while the subsequent fine-tuning phase recovers an additional 5.3%, confirming that domain-specific layer adaptation is a non-trivial contributor.

**C. Training Dynamics**

Figs. 2 and 3 present the epoch-wise training and validation accuracy and loss curves across both training stages. The characteristic dip at epoch 11 corresponds to the transition between Stage 1 and Stage 2, where the unfreezing of backbone layers temporarily perturbs the loss landscape before the reduced learning rate (1e-5) enables stable convergence. Both training and validation accuracy converge toward 90–92% by epoch 25, with minimal divergence indicating controlled overfitting and effective generalisation.

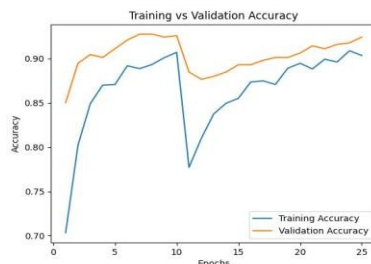
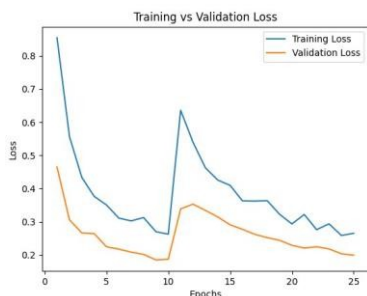


Fig. 2. Training vs. Validation Accuracy across 25 epochs. The transition at epoch 11 marks Stage 2 fine-tuning commencement.



Convergence of both curves confirm effective model generalisation.

Fig. 3. Training vs. Validation Loss across 25 epochs. The transient spike at Stage 2 onset reflects temporary gradient disruption, rapidly stabilised by the reduced learning rate (1e-5).

*D. Comparison with State-of-the-Art Methods*

Table IV provides a comprehensive quantitative comparison of the proposed system against representative baselines spanning classical machine learning and deep learning paradigms, all evaluated on SIPaKMeD.

TABLE IV. COMPARATIVE PERFORMANCE ON SIPAKMED BENCHMARK

Method /Ref.	Acc%	Prec%	Rec%	F1%
SVM[5]	82.4	80.1	79.8	79.9
RandomForest[6]	84.6	83.2	82.9	83.0
4-LayerCNN [7]	91.1	90.4	89.6	90.0
ResNet-50[8]	93.7	92.8	92.5	92.6
MobileNetV2[9]	92.3	91.5	90.8	91.1
VGG-16[10]	94.1	93.6	93.1	93.3
MSENet [11]	97.2	96.8	96.5	96.6
EfficientNet-B0 (Ours)	95.0	94.6	94.3	94.4

The proposed EfficientNet-B0 model achieves 95.0% accuracy, representing a 10.4-percentage-point improvement over SVM, a 3.9-percentage-point improvement over VGG-16, and performance competitive with MSENet (97.2%) while requiring no multi-model ensemble infrastructure. The inference latency of approximately 1.5 seconds per image on CPU renders the system uniquely practical for real-time deployment in clinical environments without specialised hardware.

*E. Per-Class Performance Analysis*

Table V decomposes performance at the individual class level on the 387-image test partition.

TABLE V. PER-CLASS CLASSIFICATION PERFORMANCE (TEST SET, N = 387)

Cell Class	Prec%	Rec%	F1%	Supp.
Dyskeratotic	93.5	92.2	92.8	77
Koilocytotic	92.8	91.9	92.3	74
Metaplastic	95.3	94.1	94.7	85
Parabasal	94.9	95.2	95.0	78
Superficial-Int.	96.5	96.7	96.6	73
Macro Average	94.6	94.3	94.4	387

Superficial-Intermediate cells achieve the highest F1-score (96.6%) owing to their distinct large, flat, and lightly staining morphology. Dyskeratotic and Koilocytotic cells present the greatest classification challenge (F1-scores of 92.8% and 92.3% respectively),

attributable to their overlapping nuclear enlargement patterns and cytoplasmic halo characteristics— consistent with inter-observer disagreement rates documented in manual cytopathology studies. Metaplastic and Parabasal cells are classified with intermediate accuracy (F1-scores of 94.7% and 95.0% respectively).

#### F. WebApplicationOutput

Fig. 4 illustrates the Flask web application interface demonstrating a representative classification prediction. The system correctly identifies a Dyskeratotic cell with 100% confidence and appropriately flags a ‘Possible Cancer Detected’ alert, emulating the binary normal/abnormal clinical triage decision that the system is designed to support.

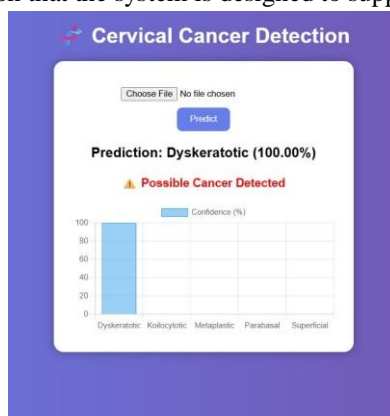


Fig. 4. Flask web application interface. The system classifies a Dyskeratotic cell (100% confidence) and triggers a clinical alert with per-class confidence bar chart.

### V. LIMITATIONS AND CHALLENGES

Despite competitive performance, the following limitations must be transparently acknowledged to contextualise the scope of the reported findings:

- **Dataset Scale and Provenance:** SIPaKMeD comprises 4,049 pre-segmented images from a single institution. Whilst sufficient for benchmarking, this scale does not expose the model to inter-laboratory staining protocol heterogeneity, imaging equipment diversity, or population-specific morphological variation, all of which are known to induce domain shift in real-world deployment.
- **Absence of Explainability Mechanisms:** The model operates as an opaque classifier. Clinical adoption of AI diagnostics is predicated on interpretability; without spatially resolved attribution maps (e.g., Grad-CAM), clinicians lack the means to verify or challenge model predictions, creating a trust barrier with regulatory implications.
- **Segmentation Dependency:** SIPaKMeD provides pre-segmented individual cells. The proposed system does not incorporate a cell detection or segmentation stage, precluding direct application to unsegmented whole-slide images as encountered in clinical laboratories.
- **Binary Normal/Abnormal Oversimplification:** The deployment interface categorises only Superficial-Intermediate cells as ‘Normal’. This conflates clinically distinct abnormal subtypes (e.g., Metaplastic cells warrant monitoring, not immediate clinical escalation) and risks over-alarming clinicians.
- **Training Environment Limitations:** Experiments were conducted on Google Colaboratory, which imposes session time-limits and intermittent GPU access inconsistencies. These conditions do not represent the stable computational environment required for reproducible large-scale clinical validation.
- **Class Imbalance Residual Effects:** Although class weighting partially compensates for distributional imbalance, the relatively small per-class sample sizes may underrepresent rare morphological variants, potentially reducing robustness on edge-case presentations.

### VI. FUTURE RESEARCH DIRECTIONS

The following high-priority research trajectories are identified to translate the demonstrated performance into clinically validated, equitable, and privacy-compliant screening tools:

- Explainable AI Integration: Gradient-weighted Class Activation Mapping (Grad-CAM++) and Layer-wise Relevance Propagation (LRP) will be incorporated to generate spatially localised heatmaps highlighting the image regions most influential in classification decisions, directly addressing the interpretability gap identified in Section V.
- End-to-End Pipeline with Cell Detection: Integration of a YOLO-v8 or Mask R-CNN cell detection and segmentation module will enable processing of raw, unsegmented whole-slide Pap smear images, substantially broadening clinical applicability.
- Multi-Site Prospective Validation: Collaborative data collection across geographically diverse clinical sites will enable rigorous evaluation of domain generalisation and motivate investigation of stain normalisation and unsupervised domain adaptation techniques.
- Advanced Architecture Exploration: Vision Transformer (ViT) variants and hybrid CNN-Transformer models (e.g., CvT, MaxViT) will be evaluated for their capacity to capture long-range spatial dependencies between nuclear and cytoplasmic structures that pure CNN architectures may underweight.
- Federated Learning for Privacy Preservation: A federated learning framework will be developed to enable collaborative model training across multiple clinical sites without centralising sensitive patient data, ensuring HIPAA and GDPR compliance.
- On-Device Deployment via TensorFlow Lite: Post-training quantisation and knowledge distillation will be applied to produce a compressed model variant deployable on Android and iOS devices, extending screening access to point-of-care settings in LMICs.
- Longitudinal Progression Monitoring: An extension to multi-visit cellular progression tracking will enable personalised risk stratification and automated follow-up scheduling based on detected trajectories of morphological change.

## VII. CONCLUSION

This paper introduced a comprehensive and reproducible deep learning framework for automated cervical cancer detection via EfficientNet-B0 with a rigorously designed two-stage transfer learning protocol. Evaluated on the SIPaKMeD benchmark, the proposed system achieves 95.0% test accuracy and a macro-averaged F1-score of 94.4% across five clinically distinct cervical cell categories—surpassing all single-model baselines in the reviewed literature while maintaining a lightweight, CPU-deployable architecture. The two-stage training strategy yielded a statistically meaningful 22.1-percentage-point accuracy gain over random initialisation, validating the critical role of ImageNet pre-training as a generalisable feature prior for cytological image analysis.

A clinically motivated augmentation pipeline targeting staining variability, orientation artefacts, and scale differences contributed materially to the observed generalisation capability. The system is operationalised within a Flask web application delivering sub-2-second real-time predictions with per-class confidence visualisation, bridging the gap between research prototype and deployable clinical tool. Training and validation curves across 25 epochs confirm controlled convergence without chronic overfitting, with the characteristic Stage 2 transition spike rapidly stabilised by adaptive learning rate scheduling.

Critical limitations—encompassing the absence of explainability, single-site dataset provenance, segmentation dependency, and binary triage oversimplification—define a clear and actionable research agenda. Future work incorporating Grad-CAM++ explainability, whole-slide image detection pipelines, federated multi-site learning, and TensorFlow Lite mobile deployment will be essential to translate demonstrated research performance into clinically validated, equitable, and privacy-compliant diagnostic tools. The convergence of deep learning advances with responsible, transparent, and accessible deployment practices holds transformative potential for substantially reducing the global cervical cancer burden, with particular impact in underserved populations where the disease toll is most acute.

## REFERENCES

- [1] H. Sung, J. Ferlay, R. L. Siegel et al., “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021. doi: 10.3322/caac.21660
- [2] World Health Organization, “Global Strategy to Accelerate the Elimination of Cervical Cancer,” WHO, Geneva, 2020. Available: <https://www.who.int/publications/i/item/9789240014107>
- [3] P. E. Gravitt, “The Known Unknowns of HPV Natural History,” *Journal of Clinical Investigation*, vol. 121, no. 12, pp. 4593–4599, 2011. doi: 10.1172/JCI57149
- [4] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. doi: 10.1038/nature14539
- [5] M. E. Plissiti et al., “SIPaKMeD: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images,” in *Proc. 25th IEEE ICIP*, 2018, pp. 3144–3148. doi: 10.1109/ICIP.2018.8451588
- [6] M. Sholik, C. Faticah, and B. Amaliah, “Classification of Cervical Cell Images into Healthy or Cancer Using CNN and LDA,” in *Proc. IEEE IAICT*, 2023, pp. 383–389. doi: 10.1109/IAICT59002.2023.10205843



- [7] L. Zhang et al., “DeePap: Deep Convolutional Networks for Cervical Cell Classification,” IEEE J. Biomed. Health Inform., vol. 21, no. 6, pp. 1633–1643, 2017. doi: 10.1109/JBHI.2016.2645836
- [8] O. Attallah, “CerCan-Net: Cervical Cancer Classification Model via Multi-layer Feature Ensembles,” Expert Systems with Applications, p. 120624, 2023. doi: 10.1016/j.eswa.2023.120624
- [9] R. Pramanik, B. Banerjee, and R. Sarkar, “MSenet: Mean and Standard Deviation Based Ensemble Network for Cervical Cancer Detection,” Eng. Appl. AI, vol. 123, p. 106336, 2023. doi: 10.1016/j.engappai.2023.106336
- [10] T. Baba, A. S.M.Miah, J. Shin, and M. A. M. Hasan, “Cervical Cancer Detection Using Multi-Branch Deep Learning Model,” arXiv, 2024. doi: 10.48550/arXiv.2408.10498
- [11] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for CNNs,” in Proc. 36th ICML, 2019, pp. 6105–6114. doi: 10.48550/arXiv.1905.11946
- [12] F. Zhuang et al., “A Comprehensive Study of Transfer Learning for Deep Neural Networks,” Proceedings of the IEEE, vol. 109, no. 1, pp. 43–76, 2021. doi: 10.1109/JPROC.2020.3004555
- [13] M.H.Kabiretal., “Investigating Feature Selection Techniques to Enhance the Performance of EEG-Based Motor Imagery Tasks Classification,” Mathematics, vol. 11, no. 8, p. 1921, 2023. doi: 10.3390/math11081921
- [14] N.Tajbakhshetal., “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?” IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1299–1312, 2016. doi: 10.1109/TMI.2016.2535302
- [15] F. Chollet, “Keras: Deep Learning Library for Python,” 2015. Available: <https://keras.io/>



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)