Cervical Cancer Detection using Deep Learning and Image Processing Techniques

Sridevi Gamini[†], **Taj Mohammad**, and **Rama Vasantha Adiraju**, Non-members

ABSTRACT

Cervical cancer is the second most common cancer among women across the globe. Detecting abnormal cervical cells at an early stage is vital for prompt treatment and improved survival rates. This project focuses on developing an effective approach to identify cervical cancer in Pap smear images using modern digital image processing and deep learning techniques. The system begins by pre-processing the medical slides to improve image clarity and reduce noise, and then applies segmentation methods to highlight and separate the regions of interest. The significant tasks include pre-processing methods such as resize and normalization, segmentation methods such as DeepLabV3, Otsu and Canny edge detection, and feature extraction methods such as ResNet101, ResNet152, AlexNet, Inceptionv3 and VGGNet16 to extract features of the cell, such as size, texture and shape of the cell, shape and color of the nuclei. To distinguish between cancerous and non-cancerous images, various machine learning algorithms are employed, including Decision Tree, Random Forest, Logistic Regression, and Support Vector Machine (SVM). The proposed methodology is evaluated on the SIPaKMeD dataset, with performance measured using established metrics such as precision, accuracy, recall, specificity, F1-score and harmonic mean to validate its robustness and reliability. By presenting a cost-effective, automated diagnostic framework to support pathologists in early cervical cancer detection, this study aligns with the broader objectives of healthcare innovation. It has the potential to enhance diagnostic efficiency and contribute to improved public health outcomes. Finally, the combination of feature extractor VGGNet 16 and classifier Decision tree gave the highest performance.

Keywords: Cervical Cancer Detection, ResNet, SIPaKMeD dataset, Convolutional Neural Networks (CNN), VGGNet16

1. INTRODUCTION

Cervical cancer [1] represents a significant public health challenge in India, ranking among the most prevalent malignancies in women and contributing substantially to cancer-related mortality [2]. The etiology of the disease is strongly associated with persistent infection by oncogenic strains of Human Papillomavirus (HPV) [3]. India accounts for approximately 21% of the global cervical cancer incidence, with an estimated 123,000 newly diagnosed cases and 77,000 associated deaths reported annually [4]. After breast cancer, cervical cancer remains the second most frequently diagnosed cancer among Indian women [5]. Although it is both preventable and curable when diagnosed early, its continued impact as a significant public health issue is driven by low awareness, insufficient screening programs, and limited access to early diagnostic care. Cervical cancer is characterized by the abnormal growth of cells in the cervix. Traditional diagnostic techniques, such as Pap smears, colposcopy, and HPV testing, remain widely used [6]. However, these methods require skilled human interpretation and are often constrained by subjectivity, reduced sensitivity, and limited reproducibility. The rapid advancements in machine learning and deep learning in the healthcare domain have created significant opportunities for the development of robust, automated diagnostic systems. These approaches can automatically extract complex features and improve classification performance, offering a hybrid strategy that enhances both sensitivity and robustness [7] [8][9].

In this study, we present a modular, interpretable, and reproducible pipeline for cervical cell classification that combines DL and ML techniques. Specifically, our method integrates DeepLabV3-based segmentation for region-of-interest (ROI) extraction, CNN-based feature extraction, and classical machine learning classifiers for final classification. An example of cervical cell images is shown in Fig. 1. The images can be magnified to different scales. Fig. 2 shows a 40X magnification of Fig. 1. Although our model may not outperform all end-to-end deep models in raw performance metrics, it provides several distinct contributions that address essential gaps in the field:

- Interpretability: The inclusion of segmentation enhances explainability, allowing clinicians to verify which image regions influence predictions visually.
- Modularity: Each stage of the pipeline—from segmentation to classification—can be updated or replaced independently, which is helpful for practical deployment

Manuscript received on May 5, 2025; revised on November 13, 2025; accepted on December 22, 2025. This paper was recommended by Associate Editor Pipat Prommee.

The authors are with Department of Electronics and Communication Engineering, Aditya University, Surampalem, India.

[†]Corresponding author: sridevi_gamini@yahoo.com

Digital Object Identifier: 10.37936/ecti-eec.2026241.259154

and customization.
- Transparency and reproducibility: Detailed examples of each stage and full implementation details are provided to support reproducibility.
- Balanced performance with simplicity: By using a hybrid DL+ML approach, in this study, the complexity of the model is systematically analyzed and addressed to ensure optimal performance and generalizability.

The organization of this paper is as follows: Section 2 provides a concise review of relevant literature. Section 3 describes the proposed methodology and presents the corresponding workflow through a block diagram. Section 4 reports the experimental evaluation of the model and offers a comparative analysis with existing CNN and machine learning approaches. Finally, Section 5 concludes the study and outlines potential directions for future research.

2. LITERATURE REVIEW

Cervical cancer remains a critical global health concern, motivating extensive research into advanced diagnostic frameworks aimed at improving early detection and classification [7]. With the rapid emergence of deep learning (DL) and machine learning (ML), new possibilities have been created for building automated, reliable, and highly accurate diagnostic systems [10].

Several studies have employed convolutional neural networks (CNNs) for detecting cervical cancer in Pap smear and colposcopy images. Archana and Jeevaraj demonstrated the effectiveness of CNNs for automated classification, achieving 91.13% accuracy on the SIPaKMeD dataset. Their work highlighted the significance of preprocessing and the role of DL in extracting discriminative features. Similarly, Zolfaghari et al.[11] achieved 97.65% accuracy by integrating a multilayer perceptron (MLP) with CNNs, underscoring the benefits of hybrid architectures that leverage DL's feature extraction capabilities alongside ML classifiers [12].

Kalbhor and Shinde [13] employed transfer learning with GoogleNet, reporting an accuracy of 96.01%, thereby demonstrating the adaptability of pre-trained networks for medical imaging tasks, particularly with limited datasets. Bhavsar et al. [14] introduced a MobileNetv2-YOLOv3 framework for cervix type classification, achieving a mean average precision (mAP) of 99.88%. Their approach emphasized the suitability of lightweight models in resource-constrained environments, ensuring both accuracy and scalability. Other works have explored the integration of DL models with traditional ML algorithms.

Thangamani et al. [15] applied XGBoost combined with feature selection, achieving 94.94% accuracy in early-stage detection, thereby stressing the importance of dimensionality reduction and robust feature selection. Publicly available datasets, such as SIPaKMeD, have further enabled benchmarking of models [16]. For instance, ResNet-152 combined with logistic regression achieved 92.28% accuracy, while hybrid approaches integrating ResNet152 with ML classifiers surpassed 98% ac- curacy, setting new standards for automated screening [17].

Despite these advances, challenges such as limited dataset size, class imbalance, and computational cost continue to impede clinical translation [18]. To overcome these, researchers have refined model architectures and employed feature fusion using multiple pre-trained networks. ResNet152, in particular, has been widely adopted due to its ability to capture intricate details through residual learning, which significantly improves classification performance [19].

By automating critical tasks such as feature extraction and classification, DL and ML models reduce dependency on subjective human interpretation [20]. Building on these developments, recent studies have proposed hybrid frameworks that combine DL feature extraction with ML classifiers to enhance robustness. For example, Vargas et al. [21] utilized shifted patch tokenization and achieved 91.2% accuracy, while El-Hoseny et al. [23] reported 86.3% precision using VG- GNet16. Similarly, Archana and Jeevaraj [11] achieved 91.3% accuracy with a hybrid CNN model. These findings reinforce the importance of combining feature extraction networks with classical ML classifiers.

Deep learning architectures such as ResNet, AlexNet, and InceptionV3 have been extensively applied to Pap smear images, improving classification accuracy [22] [23]. ResNet-101 and ResNet-152 have demonstrated superior feature extraction capabilities, with ResNet-152, combined with logistic regression, achieving 98.08% accuracy [24]. AlexNet, with its eight-layer design, achieved 96.31% accuracy when paired with logistic regression, whereas VGGNet, known for its multi-scale filtering, achieved 100% accuracy when integrated with ML classifiers [25]. These results highlight the versatility of deep architectures in cervical cancer detection [26].

Overall, DL has transformed cervical cancer diagnosis by enhancing accuracy, minimizing manual effort, and supporting earlier intervention. Hybrid models, particularly ResNet-152 with simple logistic regression (SLR), have emerged as leading approaches [29]. However, the need for larger datasets, real-time screening solutions, and cost-effective deployment in low-resource regions remains [27]. Among ML classifiers, SLR has consistently outperformed alternatives, while Random Forest (RF) demonstrated competitive accuracy when coupled with AlexNet (92.23%), and Decision Trees (DT) lagged due to overfitting issues [28].

Recent literature further shows that deep learning has revolutionized image classification across diverse domains, including healthcare, autonomous systems, and security [29]. Building upon this foundation, the present study explores a deep learning–driven classification framework that integrates CNN architectures such as AlexNet, ResNet101, ResNet152, InceptionV3, and VGGNet16 with ML classifiers including logistic regression, random forest, decision tree, and support vector machine (SVM) [30].

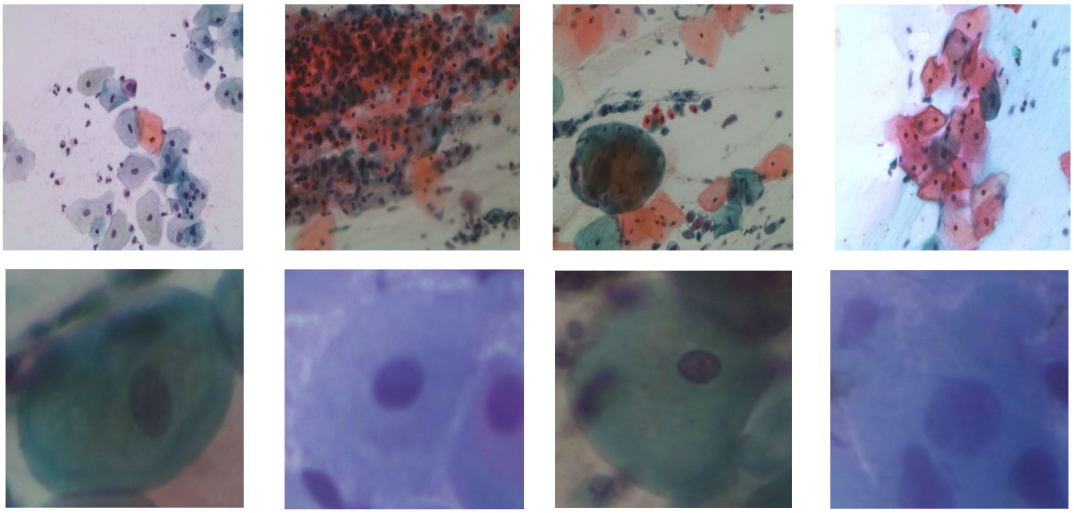

***Fig. 1**: **First Row:** Cervical Cancerous Images, **Second Row:** Non-Cervical Cancerous Images.*

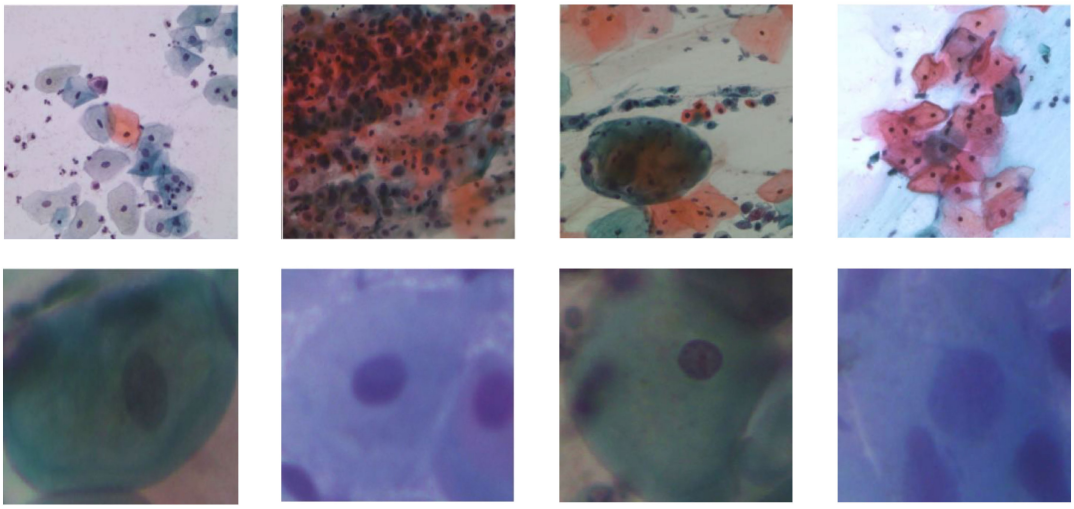

***Fig. 2**: **First Row:** 40X Magnification of Cervical Cancerous Images, **Second Row:** 40X Magnification of Non-Cervical Cancerous Images.*

3. METHODOLOGY

The proposed methodology is designed to facilitate the early- stage detection of cervical cancer, enabling timely diagnosis and intervention. The proposed method consists of the following steps: preprocessing, segmentation, feature extraction, and binary classification. The train-test split is taken as 80 percent and 20 percent for better accuracy. The block diagram illustrating the proposed methodology is presented in Fig. 3. The methodology of the proposed technique is presented as follows.

3.1 Preprocessing

3.1.1 Resizing

Resize ensures that all images are of the same size, which is necessary because neural networks expect fixed-size inputs. If the input image size is incorrect, a shape mismatch error occurs. 224×224 is the default input size expected by ResNet, AlexNet, VGGNet16, and Inception V3.

3.1.2 Normalization

Normalization helps scale the image pixel values to match the distribution of the trained models. Normalization changes the range of pixel values from 0 to 255 to a standard range from 0 to 1. These values are calculated to help the models converge faster, generalise better, and prevent exploding/vanishing gradients during training or fine-tuning.

3.2 Segmentation

3.2.1 Otsu thresholding

Otsu's thresholding is a powerful and widely used image segmentation technique, especially in medical imaging, including cervical cancer detection. It helps isolate regions of interest, such as nuclei of cervical cells. This process starts by converting pap smear images to

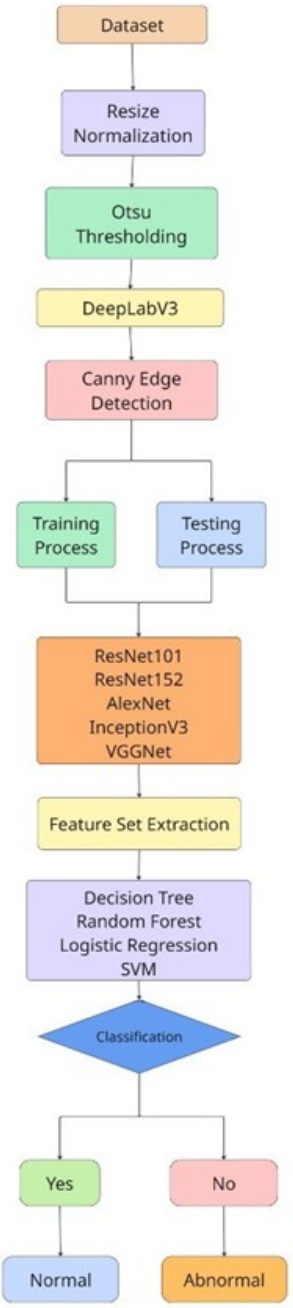

***Fig. 3**: Flow Chart of the Proposed Method.*

grayscale images. Then a histogram is generated, and the threshold is calculated by looping through possible thresholds from 0 to 255. Based on the threshold, two classes are segmented: one with pixel values below the threshold and the other with values above it. Thus, nuclei and cancerous regions are segmented from the cytoplasm and non-cancerous regions.

3.2.2 DeepLabV3

DeepLabV3 is a deep convolutional neural network model designed for semantic segmentation, that is, assigning a label to each pixel in an image. It's highly effective for medical imaging tasks, including cervical cancer detection, where segmentation of cancerous regions and nuclei is needed. It can segment complex structures even if the edges aren't transparent. It is helpful in biological images with noise, variation, and overlap. Here, in this process, the prerequisite is a resized and normalized 224x224 input image. It internally uses ResNet50 or ResNet101 and performs convolutions for capturing contextual information, such as shape and nuclei. It also helps effectively identify cancerous regions.

***Table 1**: Training Process.*

Item	Value
Framework	Python 3.9
Mode	Feature extraction
Batch size	16 (training), 32 (if GPU allows)
Optimizer	Adam
Initial learning rate	1e-4
Learning rate schedule	ReduceLROn Plateau / Step LR
Number of epochs	30
Loss	Binary cross-entropy

3.2.3 Canny Edge Detection

Canny edge detection is a popular and effective image processing technique for identifying edges. In the context of cervical cancer detection, it can help isolate cell nuclei, identify cancerous cells, and measure the size of the cells. In this process, the pap smear image is converted to grayscale, followed by noise reduction using a Gaussian blur filter. Then, the gradient is calculated by using the Sobel filters. Next, the edges are narrowed by only considering the edges with the maximum gradient. Edges can be tracked by using hysteresis, so we can detect the cell boundary and segment the nucleus.

3.3 Training Process

The training process utilizes the dataset to train several pretrained deep learning models. Throughout this process, models acquire the ability to recognize and remember correct patterns and features in input data [31]. The train/test split was performed stratified by class (80% train, 20% test), with no images shared between the splits. All experiments used the hyperparameters in Table 1.

3.4 Feature Set Extraction

For each CNN, we extracted a descriptor as follows:
- **For ResNet101/ResNet152:** the feature vector was taken from the output of the global average pooling layer following the final convolutional block (i.e., before the classification head). These yields have a dimension of 2048 for both networks.
- **For VGG16:** the feature vector was taken from the output of the last fully connected layer before the classifier (fc2), yielding 4096-dimensional features.

- For AlexNet: features were taken from the penultimate fully connected layer (fc2), yielding 4096-dimensional features.
- For InceptionV3: features were taken from the pooled features (after mixed *pooling*), *yielding* 2048 *features.*

Principal Component Analysis (PCA) was applied to the pooled feature vectors computed on the training split, and the top 103 principal components were retained. The revised manuscript now includes a new table listing the raw feature dimensionality for each CNN and the final feature dimensionality supplied to the classical ML classifiers.

For pixel-wise segmentation with DeepLabV3, we used 224 × 224 input images. All segmentation experiments report pixel-level metrics (IoU, pixel accuracy), whereas the classification experiments report image-level classification metrics (accuracy, precision, recall, F1, specificity).

Mathematically, the VGGNet16 model generates feature vectors of 4096 dimensions, which encode detailed morphological characteristics of cervical cells. To make computation more efficient, these high-dimensional vectors were reduced to 103 principal components using Principal Component Analysis (PCA), while preserving approximately 95% of the original variance. The percentage of variance retained through PCA is calculated using the expression:

$$Variance = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \times 100\% \tag{1}$$

Here, λ_i denotes the eigenvalues of the covariance matrix, $n = 4096$ represents the original feature dimensions, and $k = 103$ is the number of components retained after reduction. This step allows the model to keep the most essential information, such as cell shape, nucleus texture, and chromatin distribution, while lowering the overall computational burden. Once these reduced features are obtained, the Decision Tree classifier organises them by repeatedly selecting the attributes that yield the most significant separation between classes. This process is guided by the measure of information gain (IG), which is expressed as:

$$IG(S, A) = Entropy(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{2}$$

Where S represents the dataset, A is the candidate attribute, and S_v denotes the subset of samples corresponding to the attribute value v. This process results in hierarchical and interpretable decision boundaries that align well with the intrinsic differences between cancerous and non-cancerous cervical cells. The strong recall and F1-score observed in our experiments confirm that the constructed partitions effectively separate the two classes without significant overlap, which explains the superior performance of the VGGNet16–Decision Tree combination.

From a practical standpoint, the VGGNet16 model is well suited for analyzing Pap smear images because its small 3×3 filters can capture fine cellular details such as textures, boundaries, and chromatin patterns within the nuclei. These features are closely related to clinically essential markers, including the nucleus-to-cytoplasm ratio, texture variations, and overall nuclear shape. The Decision Tree classifier then uses these extracted features to make clear, rule-based decisions, similar to the way a pathologist would evaluate slide characteristics. The combination of VGGNet16's detailed feature extraction and the Decision Tree's interpretability, therefore not only improves diagnostic accuracy but also adds theoretical value by linking computational results with clinically meaningful observations.

The novelty of this work lies in its combination of deep learning–based feature extraction with a simple, rule-driven classification approach. Unlike conventional end-to-end CNNs that depend entirely on backpropagation to learn complex decision boundaries, our method is built on the idea that the visual and structural characteristics of cervical cells can be captured more efficiently in a reduced feature space using Principal Component Analysis (PCA). This reduction step compresses the feature dimensions while preserving the most meaningful variations related to cell morphology. The refined features are then passed to a Decision Tree classifier, which applies an entropy-based rule formation process to generate clear and interpretable decision boundaries. Together, these two stages, variance-preserving projection and rule-based partitioning, create a lightweight analytical framework that enhances interpretability and computational efficiency without increasing model depth or the number of parameters.

To provide a fair comparison, we also examined the computational complexity of the different CNN and machine learning combinations. The analysis considered three aspects: the number of model parameters, the approximate floating-point operations (FLOPs), and the average time required to process a single image during testing. Deeper networks, such as ResNet152, require more computational effort due to their large number of parameters and higher FLOPs, which naturally increases inference time. In contrast, VGGNet16 is comparatively lighter, requiring fewer parameters and fewer operations. When paired with the Decision Tree classifier, this combination achieved an average testing time of approximately 0.084 seconds per image on our system. In contrast, a more complex setup, such as ResNet152 with Logistic Regression, took around 0.145 seconds per image. These results indicate that the VGGNet16–Decision Tree model achieves robust classification performance while maintaining computational efficiency, rendering it suitable for real-time medical applications, particularly in resource-constrained environments.

4. RESULTS AND DISCUSSION:

The proposed methodology is evaluated on the SIPaKMeD dataset [32], which comprises 1,274 can-

cerous images and 2,878 non-cancerous images. Its performance is assessed using multiple deep learning architectures—ResNet101, ResNet152, AlexNet, InceptionV3, and VGGNet16—combined with various machine learning classifiers, including Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Logistic Regression (LR). The experiments are implemented in Python 3 within the Spyder environment, running on a Windows 11 system with an Intel i5 processor. Evaluation metrics include accuracy, precision, recall, F1-score, specificity, and harmonic mean.

The model's performance is measured by the proportion of image pixels correctly classified, which is calculated as accuracy. It is determined as the ratio of the number of pixels correctly classified to the total number of pixels:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Where, TP (True Positives): Correctly identified pixels as belonging to the positive class.

TN (True Negatives): Accurately labelled pixels that are in the negative class.

FP (False Positives): Misclassified pixels predicted to belong to the positive class but belong to the negative class.

FN (False Negatives): Misclassified pixels labelled as negative but of the positive class.

Precision is the proportion of correctly classified instances or samples identified as positives to the overall number classified as positives. Precision is thus expressed as

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall, which is a synonym for the power of a model in detecting positive cases in a data set. It is a synonym for true positive rate (TPR) or sensitivity. The recall formula is

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

F1-score, which is given by a measure of the accuracy of a model in terms of precision and recall, both of which are of high significance for object detection and segmentation tasks.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

The F1 score is the harmonic mean of precision and recall and provides an equal measure of the performance of a model.

Specificity is an indicator of a test's power to identify the unaffected accurately, and it is determined by dividing the true negatives (TN) by the sum of true negatives (TN) and false positives (FP).

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

The harmonic mean (H) of a list of n numbers ($x1$, $x2$, ..., xn) is obtained as the inverse of the arithmetic mean of the inverses of the numbers using the formula:

$$H = \frac{n}{(x1,\ x2,\ \ldots,\ xn)} \tag{8}$$

Where, n is the number of all values in the data set. $x1,\ x2, \ldots,\ xn$: Refers to the separate variables in the set. The equation essentially involves finding the mean of the reciprocals of the numbers and then taking the reciprocal of the mean.

The preprocessing steps of the proposed model are shown in Fig. 3. This implementation combines deep learning with machine learning to classify images effectively. Among the preprocessing methods, resizing is used to ensure consistent input size, avoid errors, and improve efficiency compared to other methods. Normalization is used for matching pre-trained statistics, centres, and scales of pixel values. Among the segmentation methods, DeepLabV3 [35] is a powerful pre-trained model that captures high-level context in the image and performs fine-grained segmentation using convolutions. It can segment complex structures even when edges are unclear, making it ideal for real-world biological images with noise, variation, or overlap. The output of each block is presented in Fig. 4.

The evaluation of various deep learning and machine learning models for the classification during training and testing phases is presented in Tables 2 and 3. Upon analysing the performance results from the training, it was observed that the VGGNet16 with Decision Tree combination yielded the best overall results across almost all key evaluation metrics. This pairing achieved the highest accuracy of 0.8329, indicating strong overall prediction capability. More significantly, it attained a perfect recall score of 1.0, demonstrating the model's exceptional ability to correctly identify all positive (cancerous) cases without missing anything.

The preprocessing techniques, such as Otsu thresholding, are considered because it has low complexity and are real-time capable, useful for nuclei segmentation, whereas DeepLabV3 is considered because it is a powerful method that handles complex scenes and does deep feature segmentation. As well as Canny Edge Detection, it is not more complex and the best gradient method for detecting cellular boundaries and cell size.

In addition, the F1-Score, which balances precision and recall, was 0.8929, the highest among all evaluated configurations. This high F1-Score highlights the model's balanced effectiveness in minimizing false positives and negatives. The precision was 0.8066, indicating a reliable true-positive rate among all predicted positives. Although the specificity was slightly lower at 0.4489, it remained competitive and acceptable, especially considering the high recall. It is crucial in medical imaging, where false negatives are costlier than false positives.

The Harmonic Mean (H-Mean), which combines precision and recall in a balanced manner, further confirmed the robustness of this configuration, yielding a value of 0.8929, the highest among all tested combinations. Meanwhile, the ResNet101 with a logistic regression

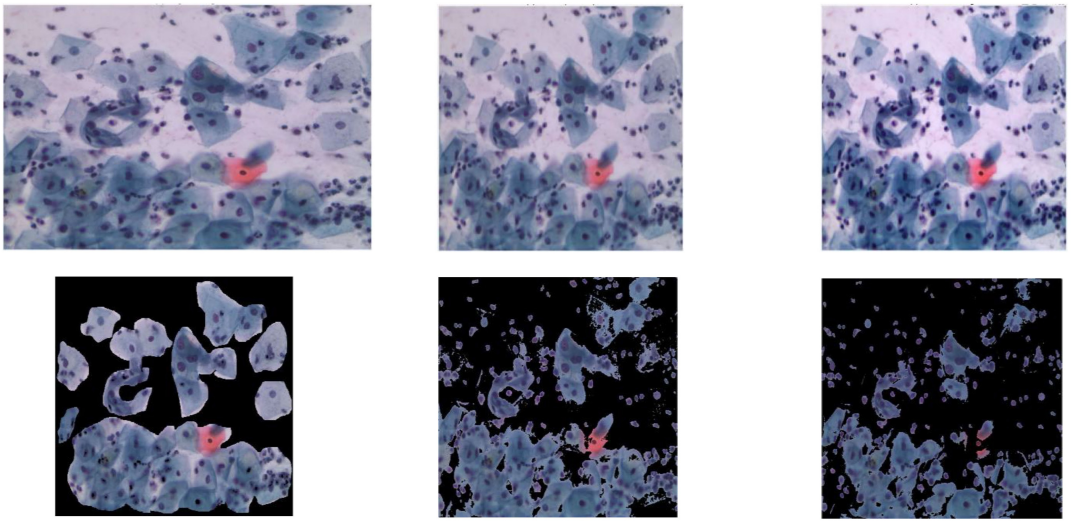

*Fig. 4: **First Row:** Original Image (888 × 668), Resized image (224 × 224), Normalized Image, **Second Row:** Otsu Thresholding, Deep Lab V3, Canny edge detection.*

***Table 2**: Evaluation and comparison of various deep learning and machine learning models for the classification of cervical cancer images, based on their performance during the training phase.*

Model	Classifier	Accuracy	F1-Score	Recall	Precision	Specificity	H-Mean
ResNet101	SVM	0.8251	0.8879	0.9939	0.8022	0.4369	0.8879
	Random Forest	0.8329	0.8929	**1**	0.8066	0.4489	0.8929
	Decision Tree	0.8329	0.8929	**1**	0.8066	0.4489	0.8929
	Logistic Regression	0.8151	0.8816	0.9883	0.7958	0.4171	0.8816
ResNet152	SVM	0.8208	0.8850	0.9896	0.8004	0.4330	0.8850
	Random Forest	0.8329	0.8929	**1**	0.8066	0.4489	0.8929
	Decision Tree	0.8329	0.8929	**1**	0.8066	0.4489	0.8929
	Logistic regression	0.8112	0.8788	0.9827	0.7948	0.4171	0.8788
AlexNet	SVM	0.8287	0.8902	0.9970	0.8041	0.4419	0.8902
	Random Forest	**0.8332**	0.8931	**1**	**0.8068**	0.4499	0.8931
	Decision Tree	**0.8332**	0.8931	**1**	**0.8068**	0.4499	0.8931
	Logistic Regression	0.8241	0.8873	0.9935	0.8016	0.4350	0.8873
InceptionV3	SVM	0.8329	0.8929	**1**	0.8066	0.4489	0.8929
	Random Forest	0.8329	0.8929	**1**	0.8066	0.4489	0.8929
	Decision Tree	0.8329	0.8929	**1**	0.8066	0.4489	0.8929
	Logistic Regression	0.8299	0.8909	0.9974	0.8050	0.4449	0.8909
VGGNet16	SVM	0.8278	0.8895	0.9948	0.8043	0.4439	0.8895
	Random Forest	0.8329	0.8929	**1**	0.8066	0.4489	0.8929
	Decision Tree	0.8329	**1**	0.8066	0.4489	**0.8929**	**1**
	Logistic Regression	0.8199	0.8847	0.9918	0.7985	0.4250	0.8847

combination outperformed all other architectures during testing. ResNet101 excels at extracting deep, hierarchical features from medical images that simpler models miss. Logistic Regression is a fast, efficient, and simple binary classification model that performs exceptionally well with rich features. It achieved a very high Recall of 0.9716, effectively detecting almost all cancer cases—an essential capability in medical diagnosis.

As compared to more sophisticated models like decision trees, logistic regression generalizes better, particularly when paired with strong feature representations. It is also computationally efficient, suitable for low-resource or real-time environments. Additionally, logistic regression is more interpretable than deep learning models, which is exceptionally beneficial in healthcare settings where explainable AI is essential.

Another advantage of this approach is its scalability and ease of deployment. After feature extraction with ResNet101, the lightweight Logistic Regression model is easy to deploy on edge devices or clinical software. Unlike training complete deep networks end-to-end, this hybrid setup achieves consistent performance with less overfitting and a smaller gap between training and test sets. Fig.s 5–14 present the confusion matrices obtained during training and testing for various deep learning architectures combined with machine learning classifiers.

***Table 3**: Evaluation and comparison of various deep learning and machine learning models for the classification of cervical cancer images, based on their performance during the testing phase.*

Model	Classifier	Accuracy	F1-Score	Recall	Precision	Specificity	H-Mean
ResNet101	SVM	0.7389	0.8335	0.9628	0.7348	0.2659	0.8335
	Random Forest	0.7304	0.8274	0.9521	0.7316	0.2622	0.8274
	Decision Tree	0.7244	0.8237	0.946	0.7279	0.2509	0.8929
	Logistic Regression	**0.7389**	**0.8347**	**0.9716**	0.7316	0.2472	**0.8347**
ResNet152	SVM	0.7377	0.8323	0.9592	0.7351	0.2697	0.8323
	Random Forest	0.7292	0.8257	0.9450	0.7331	0.2734	0.8257
	Decision Tree	0.7304	0.8255	0.9397	**0.7361**	**0.2884**	0.8255
	Logistic regression	0.7316	0.8296	0.9628	0.7289	0.2434	0.8296
AlexNet	SVM	0.7280	0.8253	0.9468	0.7315	0.2659	0.8253
	Random Forest	0.7377	0.8320	0.9574	0.7357	0.2734	0.8320
	Decision Tree	0.7208	0.8199	0.9362	0.7293	0.2659	0.8199
	Logistic Regression	0.7365	0.8311	0.9557	0.7353	0.2734	0.8311
InceptionV3	SVM	0.7304	0.8285	0.9592	0.7291	0.2472	0.8285
	Random Forest	0.7280	0.8264	0.9539	0.7290	0.2509	0.8264
	Decision Tree	0.7329	0.8284	0.9504	0.7342	0.2734	0.8284
	Logistic Regression	0.7268	0.8271	0.9628	0.7250	0.2285	0.8271
VGGNet16	SVM	0.7316	0.8291	0.9592	0.7301	0.2509	0.8291
	Random Forest	0.7353	0.8305	0.9557	0.7343	0.2697	0.8305
	Decision Tree	0.7160	0.8179	0.9397	0.7240	0.2434	0.8179
	Logistic Regression	0.7341	0.8317	0.9681	0.7290	0.2397	0.8317

***Table 4**: Comparison of various deep learning models.*

Author(s)	Methodology	Dataset	Accuracy	Precision	Recall	F1-Score	Limitations
Archana & Jeevaraj [11]	CNN	SIPaKMeD	91.13%	-	-	-	No mention of deployment scalability
Zolfaghari et al. [33]	MLP + CNN Hybrid	SIPaKMeD	97.65%	-	-	-	Computationally heavy hybrid architecture
Kalbhor & Shinde [14]	Transfer Learning (GoogleNet)	SIPaKMeD	96.01%	-	-	-	Requires fine-tuning and GPU resources
Bhavsar et al. [14]	MobileNetv2-YOLOv3	Private	99.12%	95.66%	95.57%	99.12%	Model optimized for object detection, not classification
Thangamani et al. [15]	XGBoost + Feature Selection	Custom	94.94%	-	-	-	Lacks image-specific feature learning
B.Koushik et al. [17]	Transfer Learning + Logistic Regression	SIPaKMeD	92.28%	-	-	-	Not interpretable; LR depends on strong extracted features
Vargas et al. [21]	Shifted Patch Tokenization	Custom	91.2%	-	-	-	Requires high-resolution input patches
El-Hoseny et al. [34]	VGGNet16 + ML	Custom	98%	86.3%	-	-	Performance lower than state-of-the-art
Proposed Model	**ResNet101/ VGGNet16 + ML**	**SIPaKMeD**	73.89%	83.47%	**97.16%**	83.47%	**Lower accuracy but interpretable and resource-efficient**

In the testing phase, ResNet152 and ResNet101 achieved recall values of 0.9592 and 0.9716, respectively, indicating a strong ability to identify cancerous cases correctly. However, some misclassifications were evident in the non-cancerous class, as reflected by lower specificity values (0.2697 and 0.2472). During training, VGGNet16, InceptionV3, and AlexNet showed consistently high performance, with VGGNet16 paired with Decision Tree achieving the best overall results, including a perfect recall of 1.0, an F1-score of 0.8929, and a precision of 0.8066. Fig.s 13 and 14 illustrate the training outcomes for ResNet152 and ResNet101, both of which achieve high accuracy (0.8329 and 0.8151, respectively) while effectively learning discriminative patterns. Collectively, these results highlight that while all models demonstrated the capacity to classify cervical cancer images, VGGNet16 in combination with Decision Tree and ResNet-based models offered the most reliable balance of accuracy,

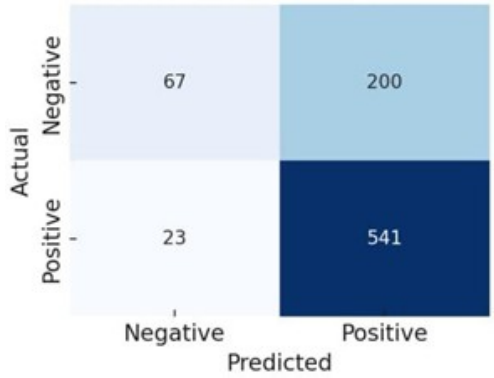

***Fig. 5**: Confusion Matrix for VGGNet16 in testing.*

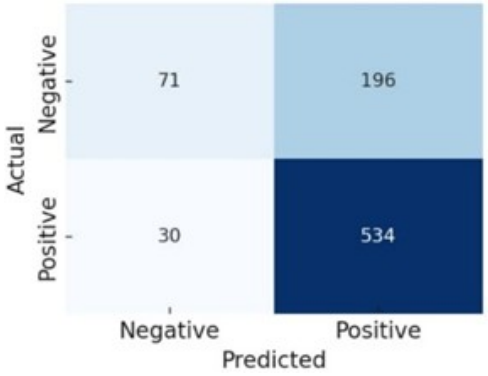

***Fig. 6**: Confusion Matrix for AlexNet in testing.*

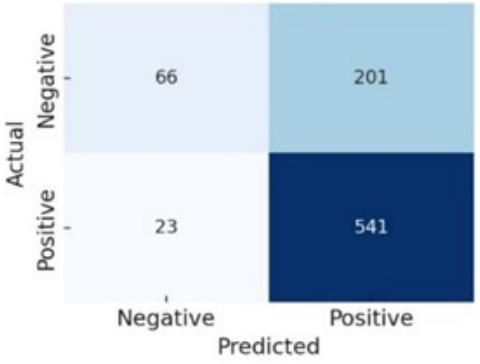

***Fig. 7**: Confusion Matrix for InceptionV3 in testing.*

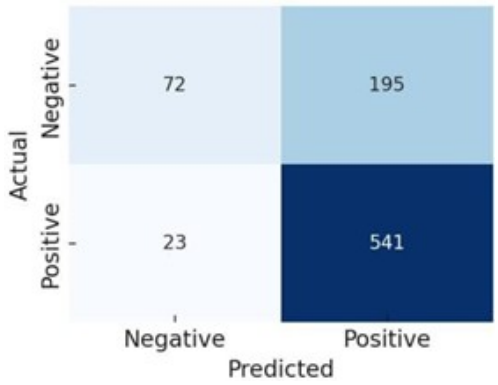

***Fig. 8**: Confusion Matrix for ResNet152 in testing.*

recall, and F1-score, making them particularly suitable for early cervical cancer detection.

While the highest testing accuracy achieved in our study (73.89%) may appear lower compared to some benchmarked models in the literature, the primary contribution of this work lies in the development of a hybrid and interpretable framework that combines deep learning-based feature extraction with classical machine learning classifiers. This approach is advantageous in terms of computational efficiency, ease of deployment, and clinical interpretability, making it particularly suitable for real-time applications in low-resource settings. Moreover, our emphasis on critical metrics such as recall (up to 0.9716) and F1-score (up to 0.8347) ensures the system's ability to minimize false negatives, which is essential in medical diagnostics. Unlike many other models that rely on data augmentation or complex architectures, our model demonstrates robustness even under class imbalance and limited dataset conditions. The hybrid nature of the system, combining ResNet101 with Logistic Regression and VGGNet16 with Decision Tree, provides a scalable and lightweight solution that balances performance with practical deployment requirements.

To provide a fair comparison, we also examined the computational complexity of the different CNN and machine learning combinations. The analysis considered three aspects: the number of model parameters, the approximate floating-point operations (FLOPs), and the average time required to process a single image during testing.

Deeper networks, such as ResNet152, require more computational effort due to their large number of parameters and higher FLOPs, which naturally increases inference time. In contrast, VGGNet16 is comparatively lighter, requiring fewer parameters and fewer operations. When paired with the Decision Tree classifier, this combination achieved an average testing time of approximately 0.084 seconds per image on our system. In contrast, a more complex setup, such as ResNet152 with Logistic Regression, took around 0.145 seconds per image. This shows that the VGGNet16–Decision Tree model not only produces strong classification results but also maintains computational efficiency, making it practical for real-time medical applications where resources are limited. The comparison of various methods is presented in Table 4.

To further substantiate the claim of computational efficiency, we conducted a detailed analysis of the proposed and comparative model configurations in terms of their parameter counts, floating-point operations (FLOPs), and average inference times per image. The results demonstrated that while deeper architectures such as ResNet152 and ResNet101 required higher computational resources due to their large parameter sets and complex operations, the proposed VGGNet16–Decision Tree combination achieved a notably faster inference speed of approximately 0.084 seconds per image on a

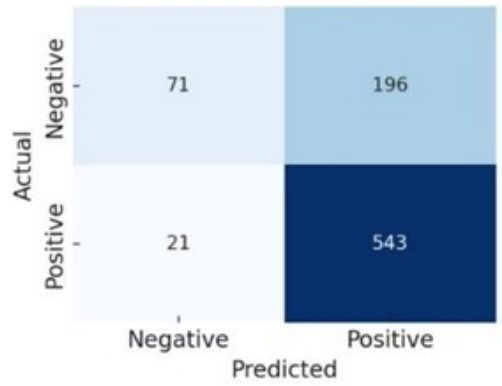

***Fig. 9**: Confusion Matrix for ResNet101 in testing.*

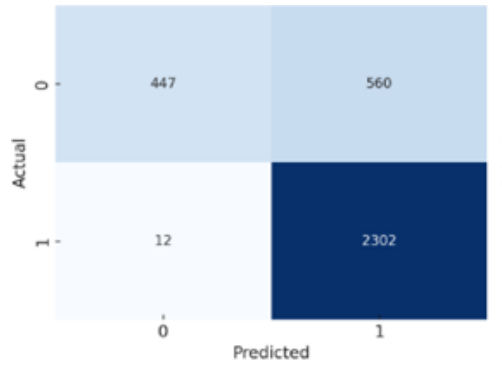

***Fig. 10**: Confusion Matrix for VGG16 in training.*

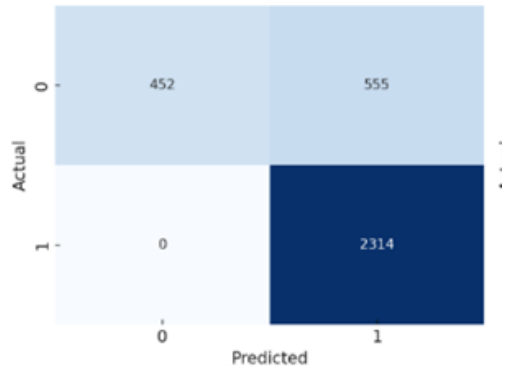

***Fig. 11**: Confusion Matrix for InceptionV3 in training.*

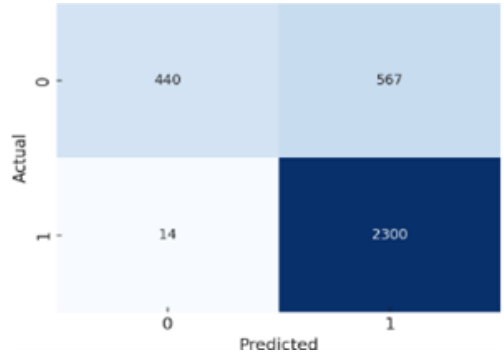

***Fig. 12**: Confusion Matrix for Alexnet in training.*

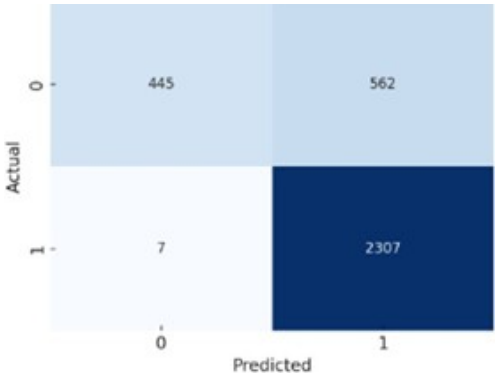

***Fig. 13**: Confusion Matrix for ResNet152 in training.*

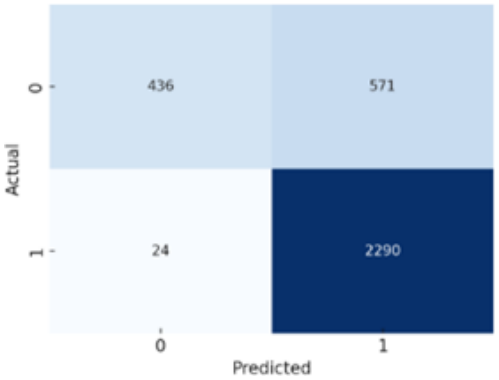

***Fig. 14**: Confusion Matrix for ResNet101 in training.*

standard CPU setup. In contrast, models like ResNet152 coupled with Logistic Regression required nearly twice the processing time (around 0.145 seconds per image). This indicates that the proposed model achieves an optimal balance between accuracy and efficiency and operates effectively. Accordingly, the term "resource-efficient" refers specifically to the reduced computational load and shorter inference time relative to deeper CNN architectures, which make the approach practical for deployment in low-resource or real-time diagnostic environments.

5. CONCLUSIONS

This study presents a structured and interpretable pipeline for early-stage cervical cancer detection that combines deep learning-based segmentation with traditional machine learning classifiers. The proposed methodology includes preprocessing techniques such as resizing and normalization, followed by segmentation using methods like Otsu thresholding, DeepLabV3, and Canny edge detection. Feature extraction was performed using popular CNN architectures—ResNet, AlexNet, InceptionV3, and VGGNet16—whose outputs were then classified using various machine learning algorithms, including Support Vector Machine (SVM), Logistic Regression, Decision Tree, and Random Forest.

Among all evaluated combinations, the integration of VGGNet16 with the Decision Tree classifier yielded the

most effective results for the binary classification of cervical cell images. Specifically, it showed improvements in key metrics, including Accuracy (0.036%), Precision (0.024%), Recall (0.26%), F1-score (11.96%), Specificity (98.68%), and H-Mean (11.96%), compared to the baseline. While some of the metric gains appear moderate, the real value of the proposed method lies in its modular design, transparency, and interpretability, which are crucial in medical diagnostic applications.

The current study focuses on binary classification using a clean dataset under controlled conditions. In future work, this methodology can be extended to multi-class classification to identify various stages or subtypes of cervical cell abnormalities. Additionally, the model can be enhanced to function effectively in noisy or low-quality imaging environments, which are common in real-world clinical settings. Further research could also integrate attention mechanisms or lightweight deep models to improve efficiency and performance on mobile or edge devices. With appropriate clinical validation, the proposed system has the potential to be deployed in large-scale screening programs, particularly in resource-constrained regions. the SIPaKMeD dataset consists of well-curated, high-quality Pap smear images with relatively clear class distinctions. As a result, it may not fully represent the variability, noise, and complexity commonly encountered in real-world clinical settings.
This study has several limitations that should be noted. First, the experiments were carried out on a single publicly available dataset (SIPaKMeD). While this dataset is widely used for benchmarking, it may not fully represent the range of cervical smear images seen in real clinical practice. Second, the experiments were conducted on a standard CPU-only system, which limited our ability to test more complex deep learning models.

REFERENCES

[1] M. Singh, R. P. Jha, N. Shri, K. Bhattacharyya, P. Patel and D. Dhamnetiya, "Secular trends in incidence and mortality of cervical cancer in India and its states, 1990–2019: Data from the Global Burden of Disease 2019 study," *BMC Cancer*, Vol. 22, No. 1, p. 149, 2022.

[2] T. Ramamoorthy, V. Kulothungan, K. Sathishkumar, N. Tomy, R. Mohan, S. Balan and P. Mathur, "Burden of cervical cancer in India: Estimates of years of life lost, years lived with disability and disability-adjusted life years at national and sub-national levels using the National Cancer Registry Programme data," *Reproductive Health*, Vol. 21, No. 1, p. 111, 2024.

[3] A. N. Della Fera, A. Warburton, T. L. Coursey, S. Khurana and A. A. McBride, "Persistent human papillomavirus infection," *Viruses*, Vol. 13, No. 2, p. 321, 2021.

[4] X. Zhang, N. Han and J. Zhang, "Comparative analysis of VGG, ResNet and GoogLeNet architectures evaluating performance, computational efficiency and convergence rates," *Applied and Computational Engineering*, Vol. 44, pp. 172–181, 2024.

[5] Y. Krishnamoorthy, K. Ganesh and M. Sakthivel, "Prevalence and determinants of breast and cervical cancer screening among women aged 30–49 years in India," *Indian Journal of Cancer*, Vol. 59, No. 1, pp. 54–64, 2022.

[6] P. Shanthi and K. Hareesha, "Automated detection and classification of cervical cancer using Pap smear microscopic images: A comprehensive review and future perspectives," *Engineered Science*, Vol. 19, No. 4, pp. 20–41, 2022.

[7] S. K. Mathivanan, D. Francis, S. Srinivasan, V. Khatavkar, K. P. and M. A. Shah, "Enhancing cervical cancer detection and robust classification through a fusion of deep learning models," *Scientific Reports*, Vol. 14, No. 1, p. 10812, 2024.

[8] D. D. Himabindu, E. L. Lydia, M. Rajesh, M. A. Ahmed and M. K. Ishak, "Leveraging Swin Transformer with ensemble of deep learning models for cervical cancer screening using colposcopy images," *Scientific Reports*, Vol. 15, 2025.

[9] M. Waqar, A. H. Khan and I. Bhatti, "Artificial intelligence in automated healthcare diagnostics: Transforming patient care," *Revista Española de Documentación Científica*, Vol. 19, No. 2, pp. 83–103, 2024.

[10] S. Asif, Y. Wenhui, S. ur Rehman, Q. ul Ain, K. Amjad, Y. Yueyang, S. Jinhai and M. Awais, "Advancements and prospects of machine learning in medical diagnostics: Unveiling the future of diagnostic precision," *Archives of Computational Methods in Engineering*, pp. 1–31, 2024.

[11] R. Archana and P. E. Jeevaraj, "Deep learning models for digital image processing: A review," *Artificial Intelligence Review*, Vol. 57, No. 1, p. 11, 2024.

[12] Z. Loukil, *A Hybrid Approach to Intelligent Prediction of Medical Conditions*, Ph.D. Thesis, University of Gloucestershire, Gloucestershire, UK, 2024.

[13] M. M. Kalbhor and S. V. Shinde, "Cervical cancer diagnosis using convolution neural network: Feature learning and transfer learning approaches," *Soft Computing*, pp. 1–11, 2023.

[14] B. Bhavsar and B. Shrimali, "TransPapCanCervix: An enhanced transfer learning-based ensemble model for cervical cancer classification," *Computational Intelligence*, Vol. 41, No. 1, p. e70027, 2025.

[15] R. Thangamani, M. Vimaladevi, S. Varshini, A. M. Aadhil and K. Sasvath, "Cervical cancer prediction using machine learning," in *Proceedings of the 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, pp. 1–7, 2024.

[16] B. Kaushik, A. Mahajan, A. Chadha, Y. F. Khan and S. Sharma, "A fine-tuned adaptive weight deep dense meta stacked transfer learning model

for effective cervical cancer prediction," *Physica Scripta*, 2025.

[17] G. Sambasivam, G. Prabu Kanna, M. S. Chauhan, P. Raja and Y. Kumar, "A hybrid deep learning model approach for automated detection and classification of cassava leaf diseases," *Scientific Reports*, Vol. 15, No. 1, p. 7009, 2025.

[18] M. Salmi, D. Atif, D. Oliva, A. Abraham and S. Ventura, "Handling imbalanced medical datasets: Review of a decade of research," *Artificial Intelligence Review*, Vol. 57, No. 10, p. 273, 2024.

[19] P. Shourie, V. Anand and S. Gupta, "Bridging nature and technology: ResNet-152 for scene visualization and classification," in *Proceedings of the 4th International Conference on Sustainable Expert Systems (ICSES)*, IEEE, pp. 1512–1516, 2024.

[20] L. Liu, J. Liu, Q. Su, Y. Chu, H. Xia and R. Xu, "Performance of artificial intelligence for diagnosing cervical intraepithelial neoplasia and cervical cancer: A systematic review and meta-analysis," *EClinicalMedicine*, Vol. 80, 2025.

[21] H. D. Vargas-Cardona, M. Rodriguez-Lopez, M. Arrivillaga, C. Vergara-Sanchez, J. P. García-Cifuentes, P. C. Bermúdez and A. Jaramillo-Botero, "Artificial intelligence for cervical cancer screening: Scoping review, 2009–2022," *International Journal of Gynecology and Obstetrics*, Vol. 165, No. 2, pp. 566–578, 2024.

[22] A. Choubey, S. B. Choubey and S. Kumar, "VGG network-based deep convoluted facial recognition," *Smart Innovation, Systems and Technologies*, 2024.

[23] T. S. R. Kiran, Y. Madhuri, P. V. B. Annapurna, A. Lakshmanarao and B. Murthy, "A novel approach to skin cancer detection using CNN, ResNet and hybrid feature extraction," 2nd International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS, pp. 217–222, 2024.

[24] 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, No. 1, p. 3945, 2025.

[25] P. Kaur and P. Mahajan, "Detection of brain tumors using a transfer learning-based optimized ResNet152 model in MR images," *Computers in Biology and Medicine*, Vol. 188, p. 109790, 2025.

[26] P. S. Thakur, T. Sheorey and A. Ojha, "VGG-iCNN: A lightweight CNN model for crop disease identification," *Multimedia Tools and Applications*, Vol. 82, No. 1, pp. 497–520, 2023.

[27] A. Boulesnane, L. Bellil and M. G. Ghiri, "A hybrid CNN–Random Forest model with landmark angles for real-time Arabic sign language recognition," *Neural Computing and Applications*, Vol. 37, No. 4, pp. 2641–2662, 2025.

[28] E. Halabaku and E. Bytyçi, "Overfitting in machine learning: A comparative analysis of decision trees and random forests," *Intelligent Automation and Soft Computing*, Vol. 39, No. 6, 2024.

[29] A. Nazir, A. Hussain, M. Singh and A. Assad, "Deep learning in medicine: Advancing healthcare with intelligent solutions and the future of holography imaging in early diagnosis," *Multimedia Tools and Applications*, pp. 1–64, 2024.

[30] S. Adamu, H. Alhussian, N. Aziz, S. J. Abdulkadir, A. Alwadin, A. A. Imam, M. Abdullahi, A. Garba and Y. Saidu, "The future of skin cancer diagnosis: A comprehensive systematic literature review of machine learning and deep learning models," *Cogent Engineering*, Vol. 11, No. 1, p. 2395425, 2024.

[31] J. Gupta, S. Pathak and G. Kumar, "Deep learning (CNN) and transfer learning: A review," *Journal of Physics: Conference Series*, Vol. 2273, p. 012029, IOP Publishing, 2022.

[32] SIPaKMeD Dataset, University of Ioannina, Available online: https://www.cs.uoi.gr/marina/SIPAKMED/, Accessed April 14, 2025.

[33] S. Zolfaghari, T. Yousefi Rezaii and S. Meshgini, "Applying common spatial pattern and convolutional neural network to classify movements via EEG signals," *Clinical EEG and Neuroscience*, Vol. 55, No. 4, pp. 486–495, 2024.

[34] H. M. El-Hoseny, H. F. Elsepae, W. A. Mohamed and A. S. Selmy, "Optimized deep learning approach for efficient diabetic retinopathy classification combining VGG16-CNN," *Computers, Materials and Continua*, Vol. 77, No. 2, 2023.

[35] [35] R. Preethi, R. Sudharsan, P. Swetha, C. Rajasekaran and S. Kausalya, "Deep learning-based semantic segmentation of turmeric crop images using the DeepLabV3+ architecture," in *Proceedings of the 2nd International Conference on Artificial Intelligence and Machine Learning Applications (AIMLA)*, IEEE, pp. 1–6, 2024.

Sridevi Gamini received her Ph.D degree from JNTUK, Kakinada, in Electronics and Communication Engineering with a research focus on digital image processing. She published more than 24 research articles in journals and conference proceedings. Currently, she is working as a Professor in the Department of Electronics and Communication Engineering, Aditya University, Surampalem, AP, India. Her areas of interest include Digital Image Processing, Fractional Calculus, Deep Learning, and Optimization Techniques.

Taj Mohammad is currently serving as an Associate Professor in Electronics and Communication Engineering at Aditya University, India. He received his Ph.D. in Electronics and Communication Engineering with a research focus on MEMS-based inertial sensors, nano-mechanical structures, and microsensor energy harvesting systems. His research interests include MEMS/NEMS sensor design, microfabrication process development, inertial measurement technologies, and micro energy harvesting.

Rama Vasantha Adiraju is an Assistant Professor at Aditya University with over 14 years of academic and research experience. She received her Ph.D. in Medical Imaging from VIT University and has expertise in medical imaging, image processing, machine learning, and deep learning. Dr. Adiraju has published and presented her research in reputed platforms and was honored with the Emerging Researcher Award in 2024. Her research focuses on developing innovative computational approaches in medical imaging to improve healthcare diagnostics and outcomes.