A Deep Learning-Driven Multimodal Healthcare System for the Early Detection of Cervical Cancer

Pratik Oak
Department of Electronics & Telecommunication, Dr. Babasaheb Ambedkar Technological University, Lonere, India
pratik24hours@gmail.com

P. S. Deshpande
Dr. Babasaheb Ambedkar Technological University, Lonere, India
deprachi3@gmail.com

Brijesh Iyer
Department of Electronics & Telecommunication, Dr. Babasaheb Ambedkar Technological University, Lonere, India
brijeshiyer@dbatu.ac.in (corresponding author)

Received: 4 April 2025 | Revised: 22 April 2025 and 5 May 2025 | Accepted: 10 May 2025

 DOI: https://doi.org/10.48084/etasr.11277

ABSTRACT

Screening parts of the human body for health analysis is a routine application of biomedical imaging. However, physicians do not rely only on imaging to arrive at a final diagnosis of the disease, but prefer clinical or laboratory results or reports to study the case along with the screened images. This study examined these dependent parameters to design an early prediction system for cervical cancer. A cervix-screened image of a patient is taken as input, along with clinical reports of the same patient. An image-based prediction is proposed by converting the original cervical image into the LAB color space before passing it through a two-channel Deep Convolution Neural Network (DCNN). The selection of the most suitable machine learning algorithm for accurate prediction based on clinical reports was ensured by focusing on recall. Logistic regression was the most effective technique in combining the two predictions for a final decision. The overall recall score of 95.83% shows the importance of the proposed method for early diagnosis.

Keywords-cervical cancer; DCNN; clinical reports; early prediction; multimodal combination

I. INTRODUCTION

Cervical cancer is one of the most dangerous diseases in women and is diagnosed using screening images. It is one of the four most common types of cancer in women across the globe. In [1], the effect of cervical cancer in India was studied by analyzing the national cancer registry program, highlighting the burden of 1.5 million patients in 2025 and revealing that it is the second most common cancer in India. Medical professionals highly refer to clinical data from a patient, along with medical images, to diagnose cancer. The effectiveness of screening techniques is fundamentally dependent on the expertise of the radiologist. Encouraged by this dependency, many researchers proposed Computer-Aided Diagnostic (CAD) systems for the accurate prediction of cervical cancer based on image input. Motivated by such CAD systems and the importance of clinical reporting in medical diagnosis, this study investigates a novel combination of screening images together with clinical test results for the early prediction of cervical cancer.

II. STATE-OF-THE-ART EARLY PREDICTION OF CERVICAL CANCER

Cervical cancer can be cured with early detection followed by appropriate treatment. Cells that affect the surface of the cervical tissue are known as Cervical Intraepithelial Neoplasia (CIN). The detection of these cells is achieved by screening techniques that require expertise. A variety of screening techniques are found in the literature, including Pap smear, colposcopy, cervicography, histology, and in vivo confocal microscopy, with different pros and cons [2]. Many experiments are carried out on medical images for automatic prediction by developing some computer-based algorithms. The literature can be broadly categorized as typical Machine-

Learning (ML) algorithmic flow-based [3, 4], segmentation-based [5, 6], morphology-based [7, 8], histogram operation-based [9, 10], and Deep Learning (DL) -based [11-16]. Table I summarizes the contributions and methods suggested in these categories.

TABLE I. STATE-OF-ART CERVICAL CANCER EARLY DETECTION APPROACHES

Ref.	Approach	Methodology	Remark
[3, 4]	Machine Learning (ML)	Feature extraction from input images passed for classification	The most relevant feature extraction increases accuracy
[5, 6]	Segmentation	Region Of Interest (ROI) extraction	Closed cells are difficult to separate
[7, 8]	Morphology	Extracts internal information by convolution	Attention in the shape of the Structuring Element (SE)
[9, 10]	Histogram	Pixel location-independent intensity adjustments	Instability on large datasets and low speed
11-17	Deep Learning (DL) approaches	Automatic feature extraction applied with different neural network architectures	Restriction to only image-based prediction limits the scope for better performance

State-of-the-art techniques offer a wide range of CAD-based early prediction. However, several gaps exist:

- Existing CAD techniques often lack sufficient accuracy in early prediction, limiting their clinical reliability.
- There is a strong dependence on specific filter selection, which affects the adaptability and robustness of the models.
- The methods exhibit instability when applied to large-scale datasets, indicating scalability issues.
- Most current approaches rely on a single imaging modality (cervix images), which restricts the overall efficiency and diagnostic power of the CAD systems.

To overcome these limitations, this study proposes a hybrid approach. In line with the diagnostic system used by physicians, this approach focuses on the prediction of cervical cancer based on multimodal input. Cervical images and clinical data of the same patient are processed to improve prediction. With this basic principle of image and clinical test report combination, the proposed system identifies the input data as normal (CIN1) or abnormal (CIN2/3) stages. The novelty and main contributions of this study are as follows:

- Effective use of LAB color space images using a DCNN for cervix image-based prediction.
- Use a Logistic Regression (LR) model for clinical data-based cancer prediction.
- Multimodal analysis system in line with the principles obeyed by medical practitioners.

III. METHODOLOGY

Figure 1 shows a graphical abstract of the proposed method. Cervigram and colposcopy images collected from the National Cancer Institute (NCI) dataset [17] are used in this work.

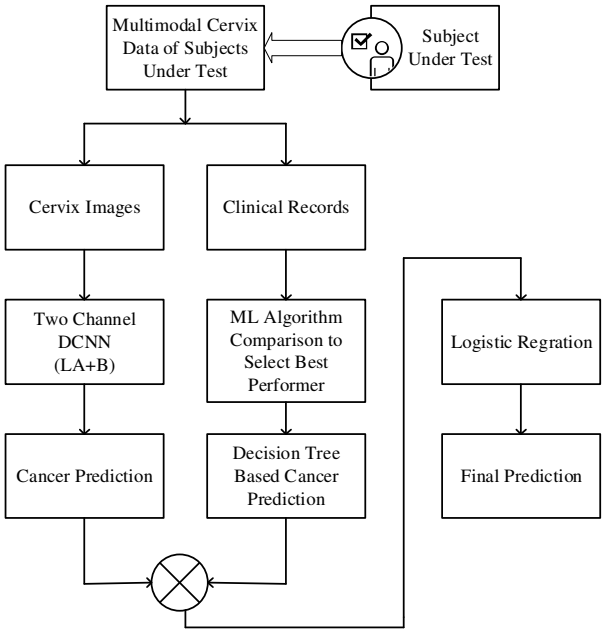

Fig. 1. Graphical abstract of the proposed system.

A. Preprocessing

Out of the different modalities, this study used colposcopy images. The cervical image captured or generated with most of the modalities consists of Specular Reflection (SR), which acts like noise and affects the efficiency of any early detection system. Thus, the adaptive SR removal technique suggested in [18] is applied to all colposcopy images used in this experiment. The SR-free images are passed as input to the next phase.

B. Two-Channel DCNN

DL algorithms perform better than ML techniques. Most previous studies used RGB color models of the original images, as the intensity of RGB channels is modified due to changes in illumination [19]. Thus, to focus on color-related features, few studies promoted the use of the CIE LAB transformation on cervix images, which works in Cartesian space. The L channel of the LAB color space represents the Lightness, the A channel depicts redness-greenness color opponents, whereas the B channel portrays the blueness-yellowness color concentration [20]. Upon separation of these three channels of abnormal or cancerous cervix images, it is observed that the infected area of the cervix is more highlighted in channel B compared to channel A. As DL techniques perform better in the Cartesian framework, this study chose a DCNN to predict cancer on cervix images.

When applying a kernel in a DCNN, some information gets passed from that kernel or filter to the layers. Many researchers suggested approaches to change the kernel per layer. Focusing on the basic principle that the next layer should get as much useful information as possible, an initial convolution was applied on separate branches, i.e., the LA and the B channels. This design, shown in Figure 2, concatenates the output of these two branches and then moves into further layers.

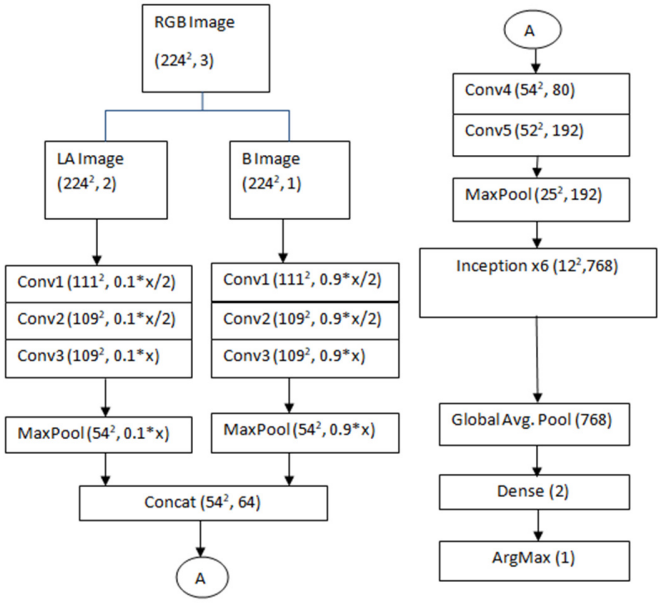

Fig. 2. Proposed network architecture.

Considering the major abnormality highlighted in channel B, a significantly larger number of filters was used for this channel. This architecture employed three different combinations of filters, namely 30-70, 20-80, and 10-90 for the LA-B channels, respectively. By observing exponential growth in terms of performance measures, such as precision, recall, and F1-score, 90% of the total filters were used for the B channel and 10% of the filters were distributed to the LA band. Thus, the total count of LA|B kernels in the first three layers is 3|29, which is given as 0.1**x* and 0.9**x*, where *x* is the number of kernels. The convolution layer used ReLU activation and batch normalization with 2D convolution. The training data was shuffled in each epoch. Transfer learning was not used in model training, i.e., training was done from scratch. This system was designed for automatic prediction of input cervical images as normal or abnormal. Thus, two dense layers are used in the training and testing of the DCNN model.

C. Prediction Based on Clinical Laboratory Data

The NCI dataset [17] consists of cervix images along with some clinical test report data that medical professionals use for actual diagnoses. Colposcopy and cervigram images used in this work have the same four clinical attributes of individual patients, namely age group, WRST_AFTER_DT, and HPV status. These three attributes determine the stage of cervical cancer named WRST_HIST_AFTER, which is considered the target variable. All these attributes are in numeric format. The target variable is the result of the diagnosis, which indicates the stage of cervical cancer. Due to an unbalanced dataset in terms of quantity for all stages, the data were classified for initial prediction, i.e., normal (up to CIN1) or abnormal (CIN2 to CIN4). Thus, all images in this dataset are represented in these two classes.

These clinical reports are represented in terms of known values (numbers) that make it possible to apply different ML algorithms for cancer prediction. This study applied and compared seven different popular ML techniques. Details on the hyperparameters in each type of classifier are given in Tables II and III. Tree-based classifiers used 100 estimators with the Gini diversity index as a learner. Large-scale bound-constrained optimization (LBFGS) was used as a solver in LR. Each ensemble classifier was used with different methods for sampling and some specific loss types. The KNN model was trained with 5 neighbors fetched by Minkowski distance with a leaf size of 30.

TABLE II. HYPERPARAMETERS FOR TREE-BASED AND REGRESSION ALGORITHMS

Classifier	**Variant**	**Maximum number of splits/estimators**	**Split criteria/ learners**
Tree Based	Decision tree	100	Gini's diversity index
	Random forest	100	
Regression	Logistic regression	100	LBFGS

TABLE III. HYPERPARAMETERS FOR ENSEMBLE AND NEAREST NEIGHBOR-BASED ALGORITHMS

Classifier	**Variant**	**Method**	**No of estimators**	**Loss type**
Ensemble	AdaBoost	-	50	Exponential
	CatBoost	Bayesian (for sampling)	1000	Log loss
	XGBoost	Friedman_MSE	100	Log loss
Classifier	**Variant**	**Leaf size**	**No of neighbors**	**Distance weight**
Nearest Neighbor	K-Nearest Neighbor	30	5	Minkowski

D. Multimodal Combination-Based Prediction

The two-way predictions of the input data, explained above, produce a probability of cancer. These multimodal predictions were combined to improve efficiency. Separate datasets with labels 0 as normal and 1 were considered in this multimodal test. The techniques analyzed in this multimodal approach are mean, weighted average, and LR [21], implemented using a meta-classifier approach. In this, a dual-channel DCNN predicts the likelihood of a disease on a cervix image. Clinical data were used to train a model to achieve the best results in predicting disease. A metaclassifier combines these two predictions to produce a final prediction that is more accurate than either individual. The ensemble methods of mean and weighted mean averaging were compared with LR to determine the efficiency of combining predictions of multiple base classifiers.

E. Performance Metrics

Classification-based algorithms are characterized and evaluated by various performance metrics, such as accuracy, precision, recall, and F1-score [21]. An F1-score greater than 0.9 designates outstanding performance, and between 0.8-0.9 is desirable. In medical image classification, although classification accuracy is important to test the system, more attention should be paid to the correct identification of positive instances. Therefore, this study focused on the recall metric.

IV. RESULTS AND DISCUSSION

The proposed multimodal approach was tested by carrying out massive experiments using Python.

A. Dataset

A total of 462 images from the NCI dataset [17] were used, including cervigrams and colposcopy images. This dataset comprises 269 images of the normal category and 193 images of the abnormal stage. This set was selected to train the DCNN model, and clinical data of the same images were used for analysis using ML algorithms.

B. Two-Channel DCNN

The efficiency of the proposed two-channel DCNN architecture was evaluated using precision, recall, and F1-score, as shown in Figure 3. The results show that the proposed DCNN performed well in terms of precision and F1 score. The recall of predicting abnormal images is high (0.92).

C. Clinical Data-Based Prediction

Clinical reports of the same patients are also available in the NCI dataset [17]. Results were collected for the same 462 images used in the above test. The performance of all ML algorithms tested with this clinical dataset was measured with the same metrics, namely precision, recall, and F1-score, along with accuracy, as shown in Table IV. The train-test split used in this analysis was 80:20.

TABLE IV. COMPARISON OF PREDICTIONS BASED ON CLINICAL DATA

Algorithm		Accuracy	Precision	Recall	F1-Score
Logistic Regression	Training	0.7669	0.7805	0.7669	0.7686
	Testing	0.7419	0.7473	0.7419	0.7433
KNN classifier	Training	0.8374	0.8377	0.8374	0.8375
	Testing	0.7957	0.8027	0.7957	0.7969
Decision Tree	Training	**0.9512**	**0.9520**	**0.9512**	**0.9513**
	Testing	**0.9140**	**0.9140**	**0.9140**	**0.9140**
Random Forest	Training	0.9512	0.9530	0.9512	0.9514
	Testing	0.8710	0.8710	0.8710	0.8710
XGB Classifier	Training	0.9485	0.95	0.9485	0.9487
	Testing	0.9032	0.9059	0.9032	0.9036
CatBoost Classifier	Training	0.9322	0.9345	0.9322	0.9245
	Testing	0.8817	0.8932	0.8817	0.8825
AdaBoost Classifier	Training	0.8699	0.8833	0.8699	0.8702
	Testing	0.8602	0.8781	0.8602	0.8611

Table IV shows that the Decision Tree classifier achieved the best performance, producing the most efficient observations in terms of all four metrics. Its testing performance indicates its efficiency. The Random Forest algorithm also gave promising results but had lower performance in testing. Therefore, the Decision Tree was used in the final decision-making system.

Figure 3 shows its class-wise test results on 93 cases (a 20% test split), highlighting both image-based and clinical-data-based prediction results. The recall value is very promising and ensures a significant possibility of abnormality detection. The F1 score is satisfactory in both cases. The average recall and F1 scores in clinical data prediction were 0.85 and 0.82, respectively.

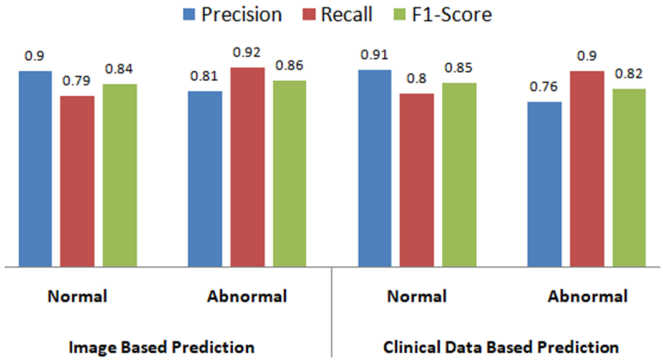

Fig. 3. Performance metrics on cervical cancer prediction.

D. Multimodal Combination System

The proposed system combines a model trained by an image-based prediction system using DCNN with a clinical data-based prediction system using a Decision Tree algorithm. These two prediction models were applied to a new dataset of 48 images that were not used in the above experiments. The statistical mean, weighted mean, and LR of the two predictions were compared with the actual class, where the latter outperformed by exhibiting efficient results in terms of overall accuracy and recall. As discussed previously, the correct identification of the abnormal class (class 1) is the most important, and this study focused more on recall. It was observed that 23 cancer patients out of 24 were correctly predicted by a multimodal combination using LR, indicating a 95.83% recall in abnormality detection. The confusion matrices shown by the mean and weighted mean methods were not promising.

Figure 4 illustrates the overall performance of the methods used in multimodal combination. In terms of the average value of precision, recall, and F1-score, LR's performance in multimodal combination is overwhelming in all metrics against mean and weighted mean combinations. Thus, LR was chosen to combine image-based prediction using DCNN and clinical-based prediction using DT. Table V presents a qualitative comparative analysis of the proposed method with others. Figure 5 compares the accuracy of this with previous studies.

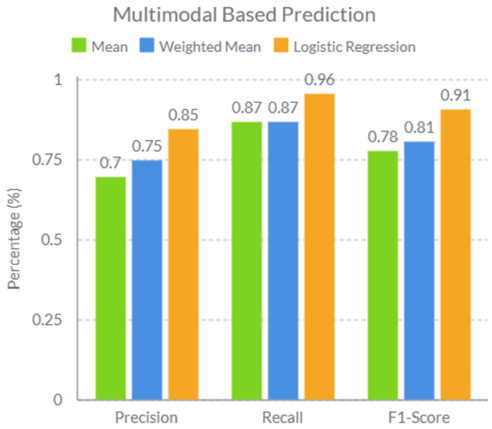

Fig. 4. Performance evaluation of the proposed multimodal combination.

TABLE V. . QUALITATIVE COMPARATIVE ANALYSIS WITH PREVIOUS METHODS

DL Technique	Modality	Dataset	Accuracy	Recall rate	Summary
Graphical CNN [12]	Single (Colposcopy images)	NCI, Costa Rica, US	78.33 %	Not reported	Used a feature encoding network with limited accuracy.
Deep Metric Learning (DML) [13]	Single (Colposcopy images)	NCI, Costa Rica, US	Average 88.5%	Not reported	Three DL algorithms (namely ResNet-50, MobileNet, and NasNet) combined to produce results.
MFEM-CIN [14]	Single (Colposcopy images)	Wanan Medical College, China	89%	Not reported	Integration of CNN and transformer architecture is done to produce a lightweight model. Suggested validation on more datasets.
SWIN transformer and CNN [15]	Single (Colposcopy images)	Colposcopy Image Bank of the International Agency	75%	Not reported	Limited data augmentation possibility and smaller size of precancerous images are the factors affecting the performance.
Proposed approach [DCNN+DT+LR]	Multi-modal (Colposcopy and Cervigram images + Clinical reports)	NCI, Costa Rica, US	89.58%	95.83%	Multimodal method working with the principle of routine medical practice. Promising results for abnormality detection.

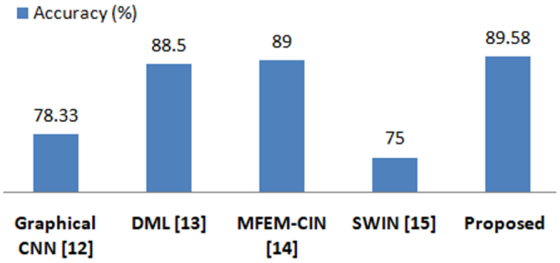

Fig. 5. Performance comparison with earlier techniques.

V. CONCLUSION AND FUTURE SCOPE

This paper presented a data-driven strategy for cervical cancer diagnosis using cervical images and their respective clinical reports. First, cervical cancer is predicted based on input images using DCNN with LAB channel mixing, which exhibited effective performance in terms of accuracy and recall. The clinical data per patient were then considered for a separate prediction by choosing the best ML algorithm. The prediction results were passed again via different statistical methods, with LR offering the best results. The effective use of the LAB color channel with a DCNN is the first contribution of this work, which achieved 92% recall in predicting abnormal images. The addition of clinical data is in line with the regular approach of medical professionals. This contribution achieved approximately 90% accuracy and 96% recall. This multimodal approach was tested on 510 images, which is a sizable amount compared to previous studies. The benefit of this work lies mainly in the multimodal analysis to increase the prediction of abnormal stages. This approach can be extended to predict more stages of cervical cancer, which will require a specific dataset with many images from all stages. In addition, it would be useful to explore different combination mechanisms of images with clinical data processing. Such a multimodal approach can be applied to different diseases, improving results.

ACKNOWLEDGMENT

The authors are grateful to acknowledge Dr. M. Schiffman and the National Cancer Institute team for providing NCI Guanacaste and ALTS project data.

REFERENCES

[1] T. Ramamoorthy *et al.*, "Burden of cervical cancer in India: estimates of years of life lost, years lived with disability and disability adjusted life years at national and subnational levels using the National Cancer Registry Programme data," *Reproductive Health*, vol. 21, no. 1, Jul. 2024, Art. no. 111, https://doi.org/10.1186/s12978-024-01837-7.

[2] M. Schiffman and D. Solomon, "Screening and Prevention Methods for Cervical Cancer," *JAMA*, vol. 302, no. 16, pp. 1809–1810, Oct. 2009, https://doi.org/10.1001/jama.2009.1573.

[3] Q. Ji, J. Engel, and E. Craine, "Texture analysis for classification of cervix lesions," *IEEE Transactions on Medical Imaging*, vol. 19, no. 11, pp. 1144–1149, Aug. 2000, https://doi.org/10.1109/42.896790.

[4] P. Oak, P. S. Deshpande, and B. Iyer, "Hybrid Feature Engineering for Early Prediction of Cervical Cancer Using Machine Learning," *Sensing and Imaging*, vol. 26, no. 1, Mar. 2025, Art. no. 39, https://doi.org/10.1007/s11220-025-00556-y.

[5] P. Mitra, S. Mitra, and S. K. Pal, "Staging of cervical cancer with soft computing," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 934–940, Jul. 2000, https://doi.org/10.1109/10.846688.

[6] W. Mu *et al.*, "A Segmentation Algorithm for Quantitative Analysis of Heterogeneous Tumors of the Cervix With 18 F-FDG PET/CT," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2465–2479, Jul. 2015, https://doi.org/10.1109/TBME.2015.2433397.

[7] M. Staring, U. A. van der Heide, S. Klein, M. A. Viergever, and J. P. W. Pluim, "Registration of Cervical MRI Using Multifeature Mutual Information," *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1412–1421, Sep. 2009, https://doi.org/10.1109/TMI.2009.2016560.

[8] Y. F. Chen *et al.*, "Semi-Automatic Segmentation and Classification of Pap Smear Cells," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 1, pp. 94–108, Jan. 2014, https://doi.org/10.1109/JBHI.2013.2250984.

[9] S. Kaaviya, V. Saranyadevi, and M. Nirmala, "PAP smear image analysis for cervical cancer detection," in *2015 IEEE International Conference on Engineering and Technology (ICETECH)*, Mar. 2015, pp. 1–4, https://doi.org/10.1109/ICETECH.2015.7275029.

[10] T. Chankong, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in Pap smears," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 539–556, Feb. 2014, https://doi.org/10.1016/j.cmpb.2013.12.012.

[11] X. Jiang, Z. Hu, S. Wang, and Y. Zhang, "Deep Learning for Medical Image-Based Cancer Diagnosis," *Cancers*, vol. 15, no. 14, Jan. 2023, Art. no. 3608, https://doi.org/10.3390/cancers15143608.

[12] Y. Li *et al.*, "Computer-Aided Cervical Cancer Diagnosis Using Time-Lapsed Colposcopic Images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3403–3415, Aug. 2020, https://doi.org/10.1109/TMI.2020.2994778.

[13] A. Pal *et al.*, "Deep Metric Learning for Cervical Image Classification," *IEEE Access*, vol. 9, pp. 53266–53275, 2021, https://doi.org/10.1109/ACCESS.2021.3069346.

[14] P. Chen, F. Liu, J. Zhang, and B. Wang, "MFEM-CIN: A Lightweight Architecture Combining CNN and Transformer for the Classification of Pre-Cancerous Lesions of the Cervix," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 5, pp. 216–225, 2024, https://doi.org/10.1109/OJEMB.2024.3367243.

[15] F. A. Mohammed, K. K. Tune, J. A. Mohammed, T. A. Wassu, and S. Muhie, "Early Cervical Cancer Diagnosis with SWIN-Transformer and Convolutional Neural Networks," *Diagnostics*, vol. 14, no. 20, Oct. 2024, Art. no. 2286, https://doi.org/10.3390/diagnostics14202286.

[16] R. Srinivasan, R. Korah, and M. Ravichandran, "Revolutionizing Diagnostic Insights: Exploring Advanced Image Processing Techniques and Neural Networks in Traditional Indian Medicine," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19214–19220, Feb. 2025, https://doi.org/10.48084/etasr.8975.

[17] R. Herrero *et al.*, "Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste Project," *Rev Panam Salud Pública*, pp. 362–374, 1997.

[18] B. Iyer and P. Oak, "Adaptive Specular Reflection Detection in Cervigrams (ASRDC) Technique: A Computer-Aided Tool for Early Screening of Cervical Cancer," in *Advances in Multidisciplinary Medical Technologies — Engineering, Modeling and Findings*, 2021, pp. 215–231, https://doi.org/10.1007/978-3-030-57552-6_14.

[19] J. P. S. Schuler, S. Romani, M. Abdel-Nasser, H. Rashwan, and D. Puig, "Color-Aware Two-Branch DCNN for Efficient Plant Disease Classification," *MENDEL*, vol. 28, no. 1, pp. 55–62, Jun. 2022, https://doi.org/10.13164/mendel.2022.1.055.

[20] I. A. P. F. Imawati, M. Sudarma, I. K. G. D. Putra, and I. P. A. Bayupati, "A Study of Lab Color Space and Its Visualization," presented at the First International Conference on Applied Mathematics, Statistics, and Computing (ICAMSAC 2023), May 2024, pp. 17–28, https://doi.org/10.2991/978-94-6463-413-6_3.

[21] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, Jul. 2011.