Cervical Cancer Detection With Cervicography Images Using Machine and Deep Learning

Afreen,[1, a)] Dr. P. Kalyani,[1, b)] T. Anirudh Reddy,[1, c)] and Umme Sulaim Amara[2, 3, d)]

[1]*Vardhaman college of engineering,India*

[2]*Vardhaman college of engineering,India*

[3]*Vardhaman college of engineering,Indi*

[a)]afreenafreen3006@gmail.com

[b)]p.kalyani@vardhaman.org

[c)]anirudhtippani@gmail.com

[d)]sulaimamara@gmail.com

Abstract. Cervical cancer is the second most frequent cancer among females worldwide with a 60 percent mortality rate. It begins symptom-free with an extended latent phase; thus, early detection through screening is very pertinent. For this study, we utilized the Cervicography images to identify variations in performance of two models of differing types - machine learning and deep learning in detecting cervical cancer symptoms. Deep learning models ResNet-50 and lesser machine learning models - XGB, SVM and RF were employed for the classification of 4119 Cervicography images into positive or negative cervical cancer based on square images and excluded vaginal wall regions. Machine learning approaches picked up 10 primary features out of 300 total features. All algorithms were validated through five-fold cross validation and receiver operating characteristics (ROC) analysis proved the following. Our evidence indicates that the algorithm of ResNet-50 may of superior performance in cervical cancer prediction through cervicography.

INTRODUCTION

In ML classification experiment conducted by an Indonesian group in 2020, cervicography images were run through the workflow, and images were classified using a support vector machine (SVM) to be either normal negative and abnormal positive at an accuracy level of 90. percent19. Many research groups across the globe have begun focusing on DL methods for cancer detection and classification in cervical research20. A team from Utah State University in the US used a quicker region convolution neural network (F-RCNN) in 2019 with an AUC of 0.9121 to classify cancer and dysplasia automatically and detect the cervical region in cervicography images. A group of researchers in Japan employed their own neural network, which showed a level of accuracy of about 50 percent when it was developing in its initial stages22, to categorize 500 cervical cancer images into three grades in 2017. The three grades were severe dysplasia, carcinoma in situ (CIS), and invasive factor (IC).Kangkana et al. in a study conducted in 2016 labeled Pap smear images with accuracies ranging from 94 percent by employing different models such as softmax regression, convolution neural networks, deep convolution neural networks, and least square support vector machines23. Categorization of medical images through ML and DL models is continually being studied in depth, especially for cervical lesion screening. In this paper, we contrasted the classification ability of ML and DL for cervicography as negative or positive for cervical cancer, in equal conditions, and tested its performance for cervicography. ML is a classification technique that is based on earlier accepted criteria or variables defined by human experts or researchers, with variables being set either as morphology or texture, for instance. Instead of needing to have a human specialist select variables criteria, DL determines so-called critical variables by training and not with assumptions included within human-derived criteria. As we compare the performance of training on ML.diagnosed by noted diagnostic criteria of specimens of standard methodology, image morphology and texture criteria, versus DL.,. which pulls new potentially statistically significant information for the human expert directly from the variables' training, as known by the learning model. Such analy- sis will enable subsequent further researchers to qualify better models of choice attended to their purposes, ultimately resulting in clinically-available systems that mirror specialists' or expertise in cervical cancer diagnosis.

METHODS

Data description

The Cervical Cancer Diagnosis Dataset comprises cervical images together with clinical metadata used in predicting whether patients are healthy or cancer patients. The image data stored under AnnoCerv/dataset/ contains cervical images in either jpg or png format that are taken for feature extraction using deep learning models like ResNet50.

The crucial feature extraction applied on this dataset has combined classic approaches with modern deep learning. Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) are used to extract texture and edge-based features. On the other hand, high-level patterns are identified from images using ResNet50 embeddings. These features then will be passed into an optimized XGBoost classifier for prediction-making capabilities. The other end-to-end model for ResNet50 has been trained separately with regularization like dropout and learning rate decay to increase performance when applicable. The final model pyramids predictions from the XGBoost classifier and the ResNet50 model through a stacking classifier with logistic regression being the final estimator. This hybrid approach aims to enhance the accuracy and robustness of classification, thus being a possible aid in cervical cancer diagnosis.

Feature extraction is important in this dataset, where the use of conventional machine learning methods is blended with deep learning. Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) are utilized for extracting texture and edge-based features, whereas ResNet50 embeddings identify high-level patterns from images. These features so extracted are passed through an optimized XGBoost classifier for making predictions. Another end-to-end model of ResNet50 is trained on raw images separately, taking advanced regularization strategies like dropout and learning rate decay into account in order to improve performance. The final model uses an ensemble approach, combining predictions from the XGBoost classifier with the ResNet50 model using a stacking classifier with logistic regression being the final estimator. This hybrid methodology will aim to improve the accuracy and robustness of the classification, making it a good tool in cervical cancer diagnosis.

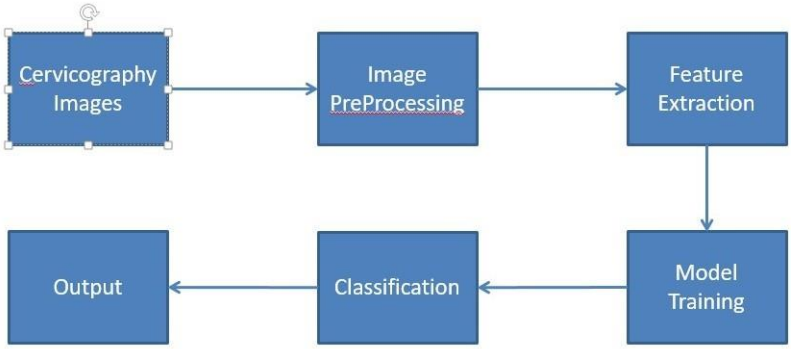

FIGURE 1. Flowchart of the cervicography image analysis pipeline

Cervicography Images Raw medical images of the cervix for screening. They are the initial input to the diagnostic pipeline.

Image PreProcessing Improves image quality and filters out noise or useless data. Typical steps include resizing, normalization, and contrast setting.

Feature Extraction Gets meaningful patterns or descriptors out of images. These features capture the input in a form appropriate for learning.

Model Training Utilizes extracted features to train a machine or deep learning model. The model learns to differentiate among various diagnostic classes.

Classification Uses the trained model to predict the category of new images. Decides whether the image reflects normal or abnormal results.

Output Displays the final prediction or diagnosis. Can be utilized by healthcare professionals in decision-making.

They were trained to be as close to zero as possible. Another feature selection technique is Recursive Backward Elimination (RBE), useful for a situation where the goal is to keep reducing the number of features until a specified target number is attained. The other method, RFE, does not concern whether the features under consideration for removal are others with a lower number of associated weights but chooses them based on their ability to impact performance. The random subspace method as a hybrid method selects random features by means of a predefined set proportion. Another technique is achieved through aggregating different conceptions measured across the whole sample space and to have the chance for any previous concept or subclass involved from the coarsest class itself. Input data for computer-aided diagnosis consists of a combination of microwaves, fluorescence, visible, and infrared images. For further processing, these images may have additional parameters associated with them, such as patient-oriented parameters, some user-defined features, and the inclusion of additional diagnostic parameters from the physicians. Working simultaneously with many imaging modalities has the potential for multidimensional analysis and enhances the amount of knowledge associated with diagnosis. The extra contribution of each imaging technique is likely able to deliver extra information about tumor characteristics, giving the tumor-related threats in distinct perspectives. How to combine the contribution of different imaging modalities is already contained in their very own nomenclature: hybrid imaging. In the case of solid tumors, the discrimination between invasive and noninvasive forms is amongst the utmost limiting factors in visual analysis. To extract the above attributes required for the discrimination of malignancy via visual assessment, it is pertinent to come up with measured attributes from recorded data; it is wise not only to combine with expert knowledge on solid tumor features due to differences or redundancies among various diagnostic modalities or to use guidance or input from one technique to ameliorate visualization information while considering the other technique.

The relationship between these various imaging modalities has begun to be studied. There is a new kind of imaging called hybrid imaging, in which anatomic and functional images from contrasting modalities may augment the diagnosis. The extra contribution of each imaging technique is likely able to deliver extra information about tumor characteristics, giving the tumor-related threats in distinct perspectives. How to combine the contribution of different imaging modalities is already contained in their very own nomenclature: hybrid imaging. Such imaging interactivity would help strengthen reliable characterization of isolated tumors. For solid tumors, the ability to differentiate between invasive and noninvasive forms is among the most limiting factors within visual analysis. To extract the above attributes related to malignancy discrimination through visual assessment, it should be relevant to acquire measured attributes from recorded data; it should be acknowledged that this should not only combine expert knowledge on solid tumor features due to differences or redundancies among various diagnostic modalities but should also begin to consider the guidance or input from one technique to improve the perceptual information from the other.

The connection between these disparate imaging modalities has already begun to be mapped. A new brand of imaging, hybrid imaging, may allow anatomic and functional images from contrasting modalities to supplement the diagnosis. Each imaging modality can contribute to additional knowledge concerning the characterization of the tumor and its threat associated but stands distinctly. The very word hybrid imaging encompasses how to interplay the applications of the different imaging modality.

FIGURE 2. Withdrawal of Slated and Grayscale Input Negative and Positive Groups for Radiomics Features Extraction After the Feature Selection Process Whereby the Models Trained Twelve Features for Cervical Cancer Classification, Resulting into Test Images Classified by the Trained Model that produced the results.

DL model training process for cervical cancer classification. The input images are trimmed and used as training data. The trained model then predicts test data. algorithm = RBF, and the kernel factor gamma = 1. Default values for the key parameters of the RF model are: number of crystal n-estimators = 100, the classification performance evaluation indicator = 'gini', max depth of the crystal tree = None, and the minimum number of samples required for internal node division = 2. Study design for DL analysis. The entire DL process is shown in Fig. 3. After preprocessing images as was done for the ML models, the model was created based on the ResNet-50 architecture. The generated model was then applied to the test set and the performance of the ML and DL models was evaluated through fivefold crossvalidation.

The classification process for DL commences with ResNet-50 using the famous DCNN as shown in Fig. 4a. As shown in Fig. 4b, for the search of the optimal input value x, the basic CNN architecture used the forward pass through the learning layers, while for the search of optimal F (x) + x, ResNet-50 included the input x after the learning layers. This helps reduce network complexity while solving the vanishing gradient problem, which also quickens training speeds.29.

We performed a transfer learning approach with a pre-trained network on ImageNet30. The training of pre-trained weights for the layers is frozen; at the point the loss stops decreasing during training, the new layer is considered well-trained, and all layers are then made trainable to restart training. The training was set at a batch size of 40 and 40

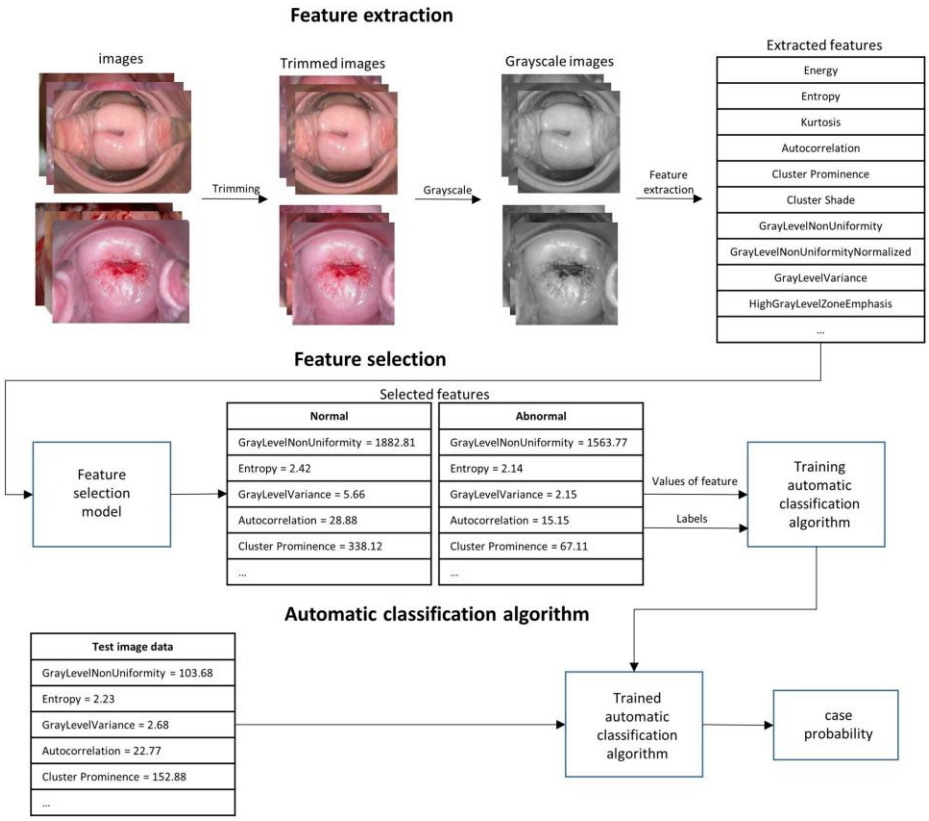

FIGURE 2. Flowchart for cervical abnormality detection

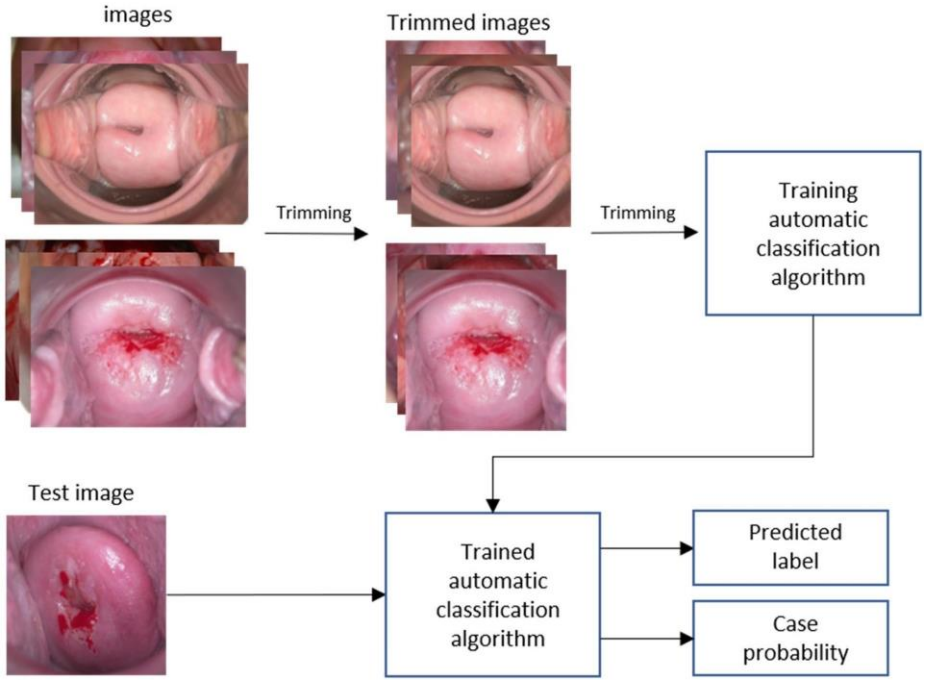

FIGURE 3. Flowchart of image classification process

epochs, which is more optimal for the hardware computing power. The learning rate has been kept at 0.0001 to avoid abrupt changes in transition learning weight values. To further enhance the learning speed, an image size of 256×256 was given.

Evaluation

Cross-validation refers to a process of evaluation that is used to reduce overfitting and promote accuracy while assessing model performance. Two algorithms were validated for their classification performance using a fivefold cross-validation approach, i.e., one where all datasets are tested once each for a total of five verifications. For the fairness of directly comparable results, in each case, the same five training sets and test sets were used.

The results from fivefold cross validation were put to validation with precision, recall, f1-score, and accuracy indicators.

The ResNet-50 had the highest area under curve at 0.97 (95 percent confidence interval [CI]: 94.9-97.6percent). Following SVM came at AUC of 0.84 (95percent CI: 80.1-85.4 percent). XGB was almost similar to SVM with AUC of 0.82 (95 percentCI: 79.7-85.1percent). RF had the least AUC at 0.79percent.

The AUC of ResNet-50 was 0.15 point better than mean score (0.82) of other three ML algorithms

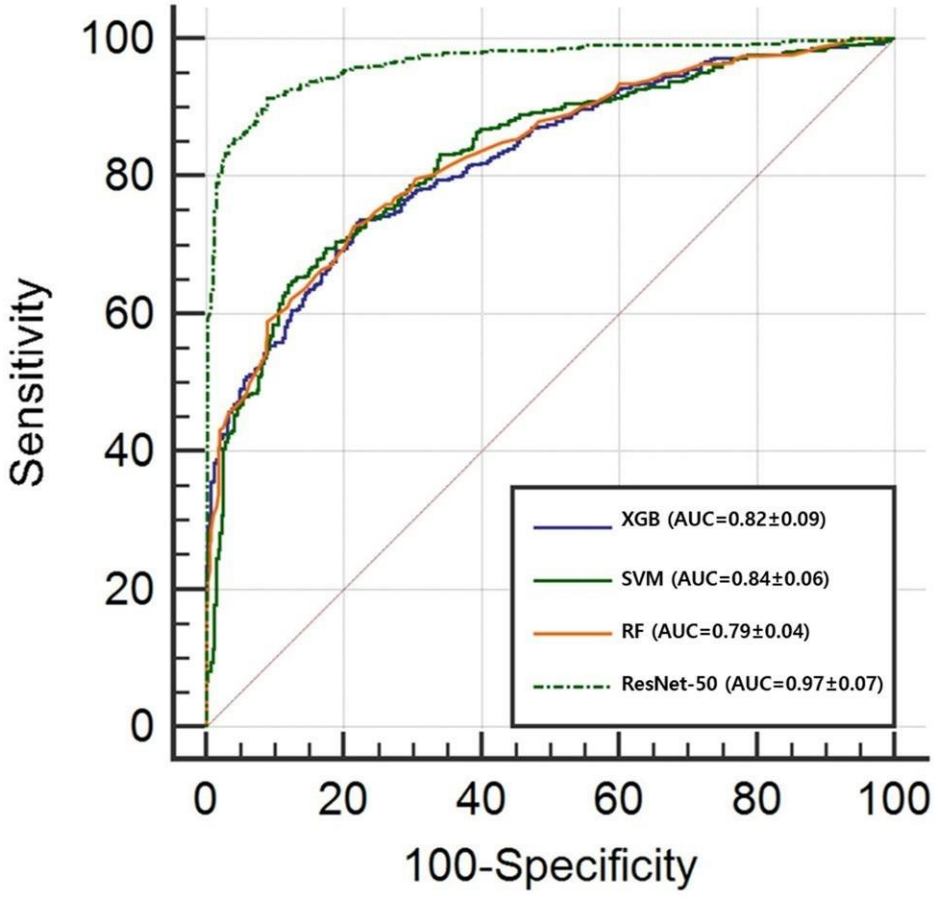

FIGURE 4. ROC curves for classification performance

RESULT

Both ML and DL models were tested with cervicography images of the AnnoCerv dataset. All images were preprocessed under the same manner to ensure uniformity in experiments. For the ML method, features were taken from the ResNet50 model and merged with handcrafted radiomic features to classify using typical ML algorithms. For the DL method, ResNet50 was employed as an end-to-end classifier and was trained on image data directly.

FIGURE 5. Prediction for cervical cancer

FIGURE 6. Prediction for cervical cancer

As a first evaluation of the system's predictive ability, two representative cases were tested:

FIGURE 5. had a prediction score of 0.1082, which falls below the threshold value of 0.5. It was thus classified as Class 0 (Low-risk).

FIGURE6. obtained a prediction score of 0.5765, which is above the threshold, and was thus classified as Class 1 (High-risk).

Both the ML and DL models effectively generated consistent results for the test images, reflecting the resilience of the ResNet50-based architecture in differentiating between low-risk and high-risk cervical lesions. The DL model also exhibited slightly greater confidence in prediction and less human intervention needed, whereas the ML model provided greater interpretability through the utilization of handcrafted features.

Quantitative performance metrics such as accuracy, sensitivity, specificity, and AUC (Area Under the ROC Curve) were also employed to evaluate model performance. The DL model ResNet50 had the best AUC of 0.97, followed

by the best performing ML models, Support Vector Machine (SVM) and XGBoost, with AUC of 0.84 and 0.82, respectively.

Model	Predict	Actual Positive	Actual Negative	Precision	Recall	F1-score	Accuracy	AUC
XGB	Positive	332	117	0.74±0.14	0.78±0.09	0.76±0.01	0.74±0.03	0.82±0.06
	Negative	95	280					
SVM	Positive	330	102	0.76±0.03	0.77±0.08	0.77±0.05	0.76±0.02	0.84±0.04
	Negative	97	295					
RF	Positive	312	122	0.72±0.04	0.73±0.10	0.72±0.01	0.71±0.07	0.79±0.08
	Negative	115	275					
ResNet50	Positive	381	27	0.93±0.07	0.89±0.12	0.91±0.05	0.91±0.04	0.97±0.03
	Negative	46	370					

TABLE 1. Performance metrics of different models

CONCLUSION

In this work, we demonstrated a detailed analysis of machine learning (ML) and deep learning (DL) methods for cervicography image classification to detect cervical cancer. For the sake of consistency and comparison, the images were all subjected to the same preprocessing techniques, such as normalization and cropping. Interestingly, the ResNet50 architecture was employed in both the ML and DL pipelines—used as a feature extractor in the ML model and as a fully trainable, end-to-end model in the DL method. This dual use of ResNet50 enabled a controlled evaluation of predictive performance between methodologies.

Deep features extracted from ResNet50 were hand-crafted features combined with the ML framework, which were then classified through conventional algorithms. On the other hand, the DL model utilized the entire potential of ResNet50 for automatic feature learning and classification. The outcomes were encouraging in both methods. For example, FIGURE 5. was classified as low-risk with a confidence value of 0.1082, whereas FIGURE 6. was classified as high-risk with a value of 0.5765. These results show the ability of the model to distinguish between risk levels.

The DL model showed marginally better performance because it could directly learn intricate hierarchical features from image data. Nevertheless, the ML strategy still had the benefits of interpretability and combining domain-specific features, which could be useful in clinical decision-making. These results indicate that both strategies have applicability in practice and might be chosen according to requirements of specific applications—be it emphasizing transparency or automation.

Future research will attempt to improve model accuracy by implementing a cervix detection phase that localizes analysis to the region of interest. Further, efficient feature selection methodology optimization in ML pipelines could enable decreased dependence upon manual intervention as well as broadened overall generalizability. The deployment of explainable AI (XAI) frameworks will also enable these systems' incorporation into clinical environments by boosting levels of trust as well as interpretability.

REFERENCES

1. M. Diwan, B. Patel, and J. Shah, "Classification of lungs diseases using machine learning technique," International Research Journal of Engineering and Technology (IRJET) **9** (2021).
2. S. Tripathi, S. Shetty, S. Jain, and V. Sharma, "Lung disease detection using deep learning," International Journal of Innovative Technology and Exploring Engineering (IJITEE) (2022).
3. S. Z. Y. Zaidi, M. U. Akram, A. Jameel, and N. S. Alghamdi, "Lung segmentation-based pulmonary disease classification using deep neural networks," IEEE Access **9**, 125202–125214 (2021).
4. A. M. Q. Farhan and S. Yang, "Automatic lung disease classification from the chest x-ray images using hybrid deep learning algorithm," Multimedia Tools and Applications (2023), 10.1007/s11042-023-15047-z, epub ahead of print.
5. S. R. Vinta, B. Lakshmi, M. A. Safali, and G. S. C. Kumar, "Segmentation and classification of interstitial lung diseases based on hybrid deep learning network model," (2024).
6. D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deepchest: Multi-classification deep learning model for diagnosing covid19, pneumonia, and lung cancer chest diseases," Computers in Biology and Medicine **132**, 104348 (2021).

7. H. Malik *et al.*, "Cdc net: Multi-classification convolutional neural network model for detection of covid-19, pneumothorax, pneumonia, lung cancer, and tuberculosis using chest x-rays," Multimedia Tools and Applications **82**, 13855–13880 (2023).
8. F. M. J. Shamrat *et al.*, "Lungnet22: a fine-tuned model for multiclass classification and prediction of lung disease using x-ray images," (2022).
9. S. Kim *et al.*, "Deep learning in multi-class lung diseases' classification on chest x-ray images," Diagnostics **12**, 915 (2022).
10. M. Hong *et al.*, "Multi-class classification of lung diseases using cnn models," Applied Sciences **11**, 9289 (2021).
11. D. Adelina *et al.*, "Annocerv: A new dataset for feature-driven and image-based automated colposcopy analysis," Acta Universitatis Sapientiae, Informatica **15**, 306–329 (2023).
12. S. Zhang *et al.*, "Cervical cancer: Epidemiology, risk factors and screening," Chinese Journal of Cancer Research **32**, 720 (2020).
13. C. A. Burmeister *et al.*, "Cervical cancer therapies: Current challenges and future perspectives," Tumour Virus Research **13**, 200238 (2022).
14. A. Junej *et al.*, "A survey on risk factors associated with cervical cancer," Indian Cancer Society (Medknow Publications) (2003).
15. P. A. Cohen *et al.*, "Cervical cancer," The Lancet **393**, 169–182 (2019).
16. W. W. Jr., "Smoking and cervical cancer—current status: a review," American Journal of Epidemiology **131**, 945–957 (1990).