BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Developing multimodal cervical cancer risk assessment and prediction model based on LMIC hospital patient card sheets and histopathological images

Kelebet Chane Jemane[1*], Muktar Bedaso Kuyu[2,3] and Geletaw Sahle Tegenaw[4]

Abstract

Cervical cancer remains a significant global health burden, particularly in low- and middle-income countries (LMICs) where access to early diagnostic tools is limited. In Ethiopia, cervical cancer diagnosis often relies on manual interpretation of biopsies, which can be time-consuming and subjective. This study aims to develop a multimodal machine learning model that integrates histopathological images and associated patient clinical records to improve cervical cancer risk prediction and biopsy detection. The dataset comprises 404 biopsy images and corresponding clinical records from 499 patients, collected at Jimma Medical Center. The preprocessing of histopathological images and clinical records involved image enhancement, data augmentation, imputation of missing values, and class balancing techniques. Subsequently, (I) a pre-trained convolutional neural network deep learning (VGG16) model was applied on the histopathological dataset, (II) a Random Forest classifier was trained on the patient clinical records, and (III) a late fusion strategy was employed to integrate the outputs of both classifiers for multimodal analysis. Recursive Feature Elimination was used to identify key predictive factors from the patient data, and the model's performance was thoroughly validated using accuracy, AUC-ROC curves, and confusion matrices, ensuring reliability across all classes. As a result, convolutional neural networks and Random Forest classifiers achieved accuracies of 91% and 96%, respectively. The integrated multimodal model achieved 92% accuracy, demonstrating enhanced robustness and clinical relevance by combining complementary data sources. These findings suggest that multimodal approaches hold promise for improving cervical cancer diagnostics in resource-limited settings. Future work will focus on validating the model with diverse datasets and integrating it into clinical workflows to support healthcare providers in LMICs.

Keywords Cervical cancer, Risk analysis, Deep learning, Predictive model, Machine learning, Multimodal model, LMICs

*Correspondence:
Kelebet Chane Jemane
Kelebet.chane@ju.edu.et
[1]Kelebet Chane Jemane, Faculty of Computing and Informatics, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia
[2]Muktar Bedaso Kuyu, Faculty of Computing and Informatics, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia
[3]Gdansk University of Technology, Gdansk, Poland
[4]Geletaw Sahle Tegenaw, Insight Research Ireland Centre for Data Analytics, Dublin City University, Dublin, Ireland

Introduction

Noncommunicable diseases (NCDs) are responsible for approximately 41 million deaths each year, accounting for 74% of all global mortality [1]. Among these, cancers represent a major contributor, especially in low- and middle-income countries (LMICs), where 77% of NCD-related deaths occur due to limited access to healthcare services and early detection tools. In Ethiopia, the NCD mortality rate is estimated at 554 per 10,000 individuals, with neoplasms, including breast, cervical, colorectal, and hematological cancers, playing a major role in this burden.

Cervical cancer is the fourth most prevalent cancer among women globally and remains a leading cause of cancer-related deaths in low and middle-income countries (LMICs). According to the World Health Organization (WHO), nearly 94% of cervical cancer deaths occur in resource-constrained settings [2]. The disease typically develops slowly over several years, progressing from precancerous cellular changes to malignant tumors. Key risk factors include persistent Human Papillomavirus (HPV) infection, smoking, early sexual debut, immunosuppression, and coexisting sexually transmitted infections (STIs). In Ethiopia, the incidence rate is estimated at 24.2 per 100,000 women, with alarmingly low screening coverage. For instance, Visual Inspection with Acetic Acid (VIA) screening revealed that 10.1% of women exhibit high-grade lesions, while 4.5% present low-grade lesions [4].

Although cervical cancer is treatable if detected early, primarily through surgery, chemotherapy, or radiation therapy [3], accurate and timely diagnosis in LMICs remains a challenge. Health systems often suffer from a lack of trained pathologists, limited diagnostic infrastructure, and an overreliance on subjective visual inspection of biopsy samples. At Jimma Medical Center (JMC), for example, cervical cancer accounts for approximately 45% of all cancer-related admissions. Yet, prediction and risk stratification remain inconsistent due to the fragmented nature of the data and the absence of clinical decision-support tools.

In response to these limitations, artificial intelligence (AI) and computer-aided diagnostic (CAD) systems have emerged as powerful tools for early cancer detection and risk analysis [5–6]. Recent studies have applied machine learning (ML) and deep learning (DL) techniques, particularly convolutional neural networks (CNNs) such as InceptionV3, EfficientNet, and VGG19, for classifying cervical cancer images by dysplasia stage or cell type [7–10]. A recent review also underscores the effectiveness of deep learning in improving diagnostic outcomes [11]. Furthermore, risk prediction of antiretroviral therapy (ART) duration and Human Immunodeficiency Virus (HIV) viral load using structured data achieved accuracies up to 90% [12]. Moreover, deep learning applied to histopathology images reported classification accuracies as high as 94.5% [13].

Despite progress in using ML and DL for cervical cancer diagnosis, most existing studies focus on a single data modality, either clinical records or histopathological images, limiting diagnostic accuracy and real-world applicability. Moreover, few models are developed using real patient data from LMICs like Ethiopia, where resource constraints, fragmented data, and late-stage diagnosis are prevalent. As a result, there remains a critical gap in developing clinically relevant, interpretable, and integrated diagnostic tools that mirror actual medical workflows.

To address this gap, this study proposes a multimodal cervical cancer risk assessment and prediction framework that fuses structured patient history (clinical records) with histopathological imaging, aiming to improve early diagnosis and support clinical decision-making in LMICs.

The main contributions of this work are as follows:

1. **Multimodal fusion**: We propose a novel ensemble model that integrates structured clinical data and biopsy images to reflect real-world diagnostic practice in the LMIC context.
2. **Robust preprocessing**: We handle missing values, class imbalance, and feature selection using techniques such as mean imputation, Synthetic Minority Oversampling Technique (SMOTE), Tomek Links, and Recursive Feature Elimination (RFE).
3. **Deep learning-based image analysis**: We fine-tune VGG16 and ResNet50 networks on curated histopathological images collected from LMIC settings to enhance diagnostic accuracy in resource-limited environments and contribute to the broader scientific community.
4. **Model interpretability**: We identify clinically meaningful features (e.g., post-coital bleeding, sexually transmitted infection (STI) history) that contribute to model decisions, increasing transparency and clinical relevance.
5. **Contextual relevance**: We develop and evaluate the model using real-world data from Jimma Medical Center, demonstrating its potential for deployment in Ethiopian and other LMIC healthcare settings.

Moreover, this paper is organized into five sections. Section 2 presents related work, identifying current gaps in unimodal cervical cancer diagnosis. Section 3 details the dataset, preprocessing pipeline, and model architecture. Section 4 reports the experimental setup, evaluation metrics, and results. Section 5 concludes with key findings, limitations, and directions for future research.

Related work

In recent years, there has been an increasing amount of literature on automated cervical cancer detection systems, focusing on advances in AI techniques and their applicability in LMICs. Researchers have demonstrated the effectiveness of DL models such as CNNs, including VGG19, EfficientNet, and ResNet variants, achieving high accuracy (>90%) in classifying cervical histopathological images by cancer subtype, dysplasia stage, or malignancy presence [7, 8, 9, 10, 13]. Additionally, clinical risk prediction models using Random Forest (RF) ML have identified key predictors such as duration on antiretroviral therapy (ART), viral load, HIV clinical stage, tuberculosis preventive therapy (TPT), and family planning methods, attaining classification accuracies around 90% in high-risk populations [12].

However, despite their promising results, most existing models are unimodal, focusing exclusively on either imaging or clinical data. This unimodality limits their ability to capture the multifactorial and complex nature of cervical cancer progression. Moreover, it often oversimplifies clinical decision-making and may hinder generalizability, especially where patient data are frequently fragmented, incomplete, or inconsistent.

Recent advances in cancer subtype classification have applied optimization algorithms to multi-omics data, yielding promising results in handling high-dimensional and heterogeneous datasets. For instance, a Cat Swarm Optimization (CSO) based feature selection framework combined with Support Vector Machines (SVM) achieved a promising result after optimal feature reduction [14]. Similarly, Quantum Cat Swarm Optimization (QCSO) improves clustering and feature selection, resulting in enhanced accuracy and interpretability [15]. However, it remains underexplored for cervical cancer diagnosis in LMICs, which pose unique challenges related to infrastructure and data availability. Furthermore, a review highlights significant progress in integrating multimodal datasets, yet al.so identifies ongoing challenges in data standardization, model transparency, and clinical translation [16]. These challenges highlight the importance of developing methods that balance computational sophistication with real-world feasibility.

In response, several recent studies have explored multimodal models combining clinical and imaging data for cervical cancer diagnosis. Ensemble frameworks that integrate clinical features with CNN-extracted image representations have demonstrated improved diagnostic accuracy and interpretable risk stratification [17]. Ming et al. [17] proposed a deep learning-based multimodal image analysis framework for cervical cancer detection using FDG-PET and CT imaging modalities. Specifically, their adaptive fusion strategy aligned structural (CT) and functional (PET) imaging features prior to detection, yielding an average improvement of 6.06% over single-modality PET and 8.9% over other multimodal fusion techniques. Abinaya and Sivakumar [18] also proposed a 3D CNN combined with a Vision Transformer (ViT) for cervical cancer classification, achieving an accuracy of 98.6%. However, the model did not incorporate structured clinical features, limiting its applicability in real-world LMIC settings where multimodal data integration is critical. Nonetheless, there remains a scarcity of frameworks tailored to the realities of LMICs, where data are often noisy, incomplete, and heterogeneous.

In parallel, emerging research demonstrates the advantages of interpretable models that combine feature engineering and explainable AI tools such as SHAP (Shapley Additive Explanations). For instance, Double Conglomerate (D-CongNet) and Triple Conglomerate (T-CongNet) achieve high accuracy in cardiovascular disease and breast cancer classification without relying on synthetic oversampling [19]. Similarly, in diabetes mellitus prediction, SHAP-based feature selection has been used to enhance ensemble model performance and improve interpretability, despite limited and imbalanced datasets typical of LMIC contexts [20]. In all, it provides clinically meaningful explanations by identifying key feature interactions that align with established medical knowledge. However, implementing AI diagnostic tools in LMICs encounters several barriers, including limited access to high-quality data, infrastructural constraints, and fragmented medical record systems [21]. Moreover, many DL models function as "black boxes," limiting interpretability and clinician trust. Enhancing transparency through explainable AI (XAI) techniques is essential for clinical adoption [22].

Limitations of unimodal models and the case for multimodal integration

Clinical diagnosis of cervical cancer inherently synthesizes multiple information sources, including biopsy results, patient histories, and laboratory findings. Recognizing this, recent research has emphasized the need for multimodal AI frameworks that integrate heterogeneous data sources such as electronic health records, patient card-sheets, imaging, and genomic data to better mimic real-world clinical workflows [23, 17]. For instance, Colposcopic Multimodal Temporal Convolutional Neural Network (CMT-CNN) [25] was proposed to fuse sequential cervigram images (acquired after saline, acetic acid, and Lugol's iodine applications) with structured clinical variables (HPV test results and patient demographics) to detect cervical intraepithelial neoplasia. The same author in [26] employed a pretrained CNN (AlexNet) for cervical cancer classification. In both scenarios, a multimodal approach was demonstrated by combining cervigram imaging with structured clinical variables, yielding

promising results. Nevertheless, multimodal integration introduces new challenges related to data preprocessing, harmonization of feature types, missing data handling, and maintaining model interpretability. Therefore, there is a need for developing a clinically grounded, interpretable, and efficient multimodal model to support personalized cervical cancer diagnosis in low-resource environments.

Research gaps and justification for the proposed approach

While existing multi-omics optimization-based classification methods and multimodal ensemble models have advanced cancer diagnosis, significant gaps remain. Optimization-based approaches have yet to be fully integrated with clinical and imaging data fusion specifically for cervical cancer diagnosis in LMICs. Furthermore, many multimodal cervical cancer models rely heavily on data augmentation techniques, which may compromise model stability and interpretability. To address these gaps, we propose a multimodal ensemble framework that combines clinical data from patient card sheets with deep learning-based histopathological image analysis. By integrating Random Forest classifiers for clinical risk factors with CNN (VGG16 and ResNet50) models for image feature extraction and applying rigorous preprocessing (mean imputation, SMOTE, RFE), the framework aims to enhance diagnostic precision and emulate clinical workflows adapted for LMICs contexts. Furthermore, our approach emphasizes interpretability, clinical relevance, and robustness, building upon insights from recent advances in multi-omics optimization and augmentation-free explainable AI models.

Methods

As shown in the detailed proposed methodology in Fig. 1, we followed four major steps to design a multimodal cervical cancer risk assessment and prediction model: (I) Multimodal data acquisitions and collections, (II) Data processing, (III) Model building, and (IV) Model evaluation.

Step 1: data collection and acquisitions

A set of steps, including retrospective design, study site selection, purposive sampling, and ethical considerations, were followed to capture histopathological images and collect patient card sheets.

Research design

A retrospective design was employed to collect and analyze the patient card-sheet records and histopathological images to examine the potential relationships between risk factors and cervical cancer outcomes. Data was collected from hospital archives at Jimma Medical Center (JMC) for patients who underwent cervical cancer screening between 2021 and 2023. The selection criteria included patients with complete medical records containing demographic details, clinical history, and histopathological findings.

Study site and populations

The study was conducted at JMC in Ethiopia, which is one of the centers for cervical cancer treatment next to Tikur Anbessa Specialized Hospital. JMC is a leading healthcare institution in Jimma, Oromia Region, Ethiopia. The center provides a comprehensive range of medical services, including outpatient care, surgical procedures, and maternal and child health services, while also playing a vital role in medical education and research in the region. The target population included all female patients who underwent cervical cancer screening at JMC between 2021 and 2023.

Sampling

Purposive sampling was employed to select relevant patient records. This sampling strategy ensured that the most informative cases were chosen, optimizing the accuracy of cervical cancer risk assessments. All records supplemented by the domain experts, including patient card sheets and microscopic images of biopsied tissue, were used. Given the study's objective of examining the relationship between risk factors and cervical cancer

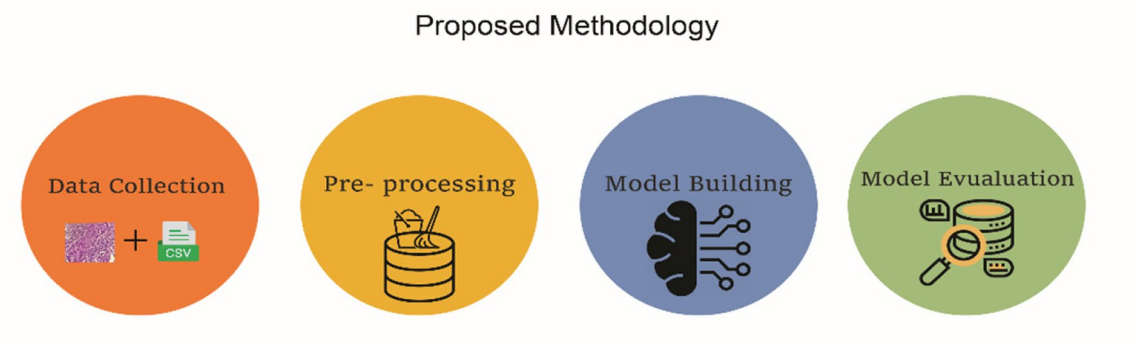

Fig. 1 Proposed methodology

outcomes, purposive sampling ensured that only records containing clinical and histopathological data were included.

The selection process was guided by domain experts (oncologists and pathologists) to ensure that the dataset contained complete patient histories, diagnostic reports, and microscopic images of biopsied tissue. The inclusion criteria prioritized patients with detailed medical records, histopathological findings, and confirmed cervical cancer diagnoses, while incomplete or ambiguous records were excluded. By using this approach, the study maximized the reliability of findings in assessing risk factors and cancer progression.

Ethical approval

Ethical clearance was obtained from the Jimma Institute of Technology, Research, and Ethical Review Board with reference no. RPD/JiT/183/15. Data confidentiality was strictly maintained by excluding personal identifiers and using histopathology biopsy images exclusively for research purposes.

Data collection

To collect the patient card sheets dataset, we used a paper-based pre-digital template sheet. We converted the paper-based data into an MS Excel CSV file format, whereas histopathological images were captured using a Samsung M13 smartphone mounted on a light microscope. As determined by domain experts, images were captured at varying magnification powers (4x, 10x, and 40x) and saved in high-resolution PNG format (2686 × 2686 pixels) with an 8-bit RGB color model.

Step 2: data preprocessing

Patient card sheet preprocessing

The preprocessing of structured clinical records involved three main stages: handling missing data, addressing class imbalance, and feature selection.

Missing value imputation To address incomplete entries in numerical variables, mean imputation was employed. The percentage of missing values ranged from 0.2% (e.g., virginal bleeding) to 2.4% (e.g., occupation), with features such as virginal discharge containing no missing values. Given the low overall proportion of missing data (< 5%), mean imputation was selected for its simplicity, computational efficiency, and ability to preserve the central tendency of each variable without introducing significant bias. This method has been recommended for datasets with minimal and randomly distributed missingness, where more complex imputation strategies may not yield substantial benefits [23].

Handling class imbalance with SMOTE and tomek links To mitigate the effects of class imbalance in risk factor data, a hybrid resampling approach was adopted, combining the Synthetic Minority Over-sampling Technique (SMOTE) with Tomek Links. SMOTE generates synthetic instances of the minority class by interpolating between existing minority samples and their nearest neighbors, thus increasing class representation and reducing model bias [23]. However, oversampling alone may introduce noisy or borderline examples. To counteract this, Tomek Links were applied post-SMOTE to remove ambiguous samples that are closest to the decision boundary, effectively cleaning overlapping instances between classes. This sequential application of SMOTE followed by Tomek Links has been shown to enhance classifier performance by balancing the dataset while preserving class distinction [23].

Feature selection using recursive feature elimination (RFE) To reduce dimensionality and retain the most predictive variables, Recursive Feature Elimination (RFE) was implemented with a Logistic Regression estimator. RFE iteratively removes features based on their contribution to model performance, resulting in a refined subset of attributes that optimally support classification tasks [24]. The most influential features identified included residential area, vaginal discharge, post-coital bleeding, irregular menstruation, age at menarche, STI: syphilis, and STI: HIV. These variables are clinically meaningful, with established associations to cervical cancer risk through behavioral, reproductive, and socioeconomic pathways.

Histopathological image preprocessing

Preprocessing of histopathological images involved color space transformation, contrast enhancement, data augmentation, and class balancing.

Color space conversion and standardization All images were converted from RGB to YCrCb color space to decouple luminance (Y) from chrominance (Cr, Cb), enabling localized contrast enhancement while preserving structural and color information. This separation is particularly beneficial for histopathological imaging, where subtle texture differences in luminance are diagnostically relevant [17]. All images were uniformly resized to 224 × 224 pixels to align with the input dimensions of convolutional neural networks (CNNs) such as VGG16 and ResNet50, facilitating transfer learning and consistent model input [27].

Contrast enhancement via CLAHE To enhance local contrast and reveal intricate tissue features, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to the luminance channel. CLAHE was implemented with an 8 × 8 tile grid and a clip limit of 2.0, strik-

ing a balance between contrast amplification and noise suppression. This approach preserves essential diagnostic information in histological slides and improves the visual separability of tissue structures [28].

Data augmentation and class balancing To increase dataset diversity and improve model generalization, extensive data augmentation was performed using Keras's ImageDataGenerator. Transformations included random rotations, horizontal and vertical flips, shear deformations, and brightness adjustments [29]. This process simulates variability in real-world imaging conditions and combats overfitting, particularly in deep learning architectures.

In addition, SMOTE was applied to CNN-extracted image features to further address class imbalance at the feature level. This ensured that minority class representations were adequately amplified, contributing to more equitable model learning and performance [23].

Step 3: multimodal model building

A multimodal framework was developed to integrate histopathological image features with structured clinical data from patient card sheets, aiming to enhance diagnostic accuracy for cervical cancer. The overall system architecture is illustrated in Fig. 2.

Histopathological image-based classification

Pre-trained convolutional neural networks (CNNs), VGG16 and ResNet-50, were adopted and fine-tuned on the cervical cancer histopathological image dataset. These models were initially trained on ImageNet and adapted using transfer learning.

VGG16 architecture VGG16 comprises 13 convolutional layers utilizing 3×3 filters and 2×2 max-pooling layers, followed by three fully connected layers. It employs ReLU activation and softmax in the output layer [27]. For fine-tuning, the fully connected layers were replaced with two dense layers (512 and 128 units), followed by dropout (0.3) and a softmax classifier. Training was conducted for 30 epochs with a batch size of 16, using the Adam optimizer (learning rate = 0.001) and categorical cross-entropy as the loss function.

ResNet-50 architecture ResNet-50 integrates residual connections to address vanishing gradient issues in deep networks. It contains bottleneck residual blocks using 1×1 and 3×3 convolutions, batch normalization, and identity mappings [30]. The final layers were fine-tuned to adapt to the domain-specific task of cervical cancer classification. A similar training protocol to VGG16 was followed.

Hyperparameter optimization To achieve optimal model performance, hyperparameters such as learning rate, dropout rate, batch size, optimizer type, activation functions, and loss function were systematically tuned using a comprehensive grid search strategy. This approach enabled robust and efficient exploration of parameter configurations, improving training stability and predictive accuracy [31]. The detailed histopathological processing pipeline is illustrated in Fig. 3.

Patient card sheet classification

To analyze risk factors from structured clinical data, a Random Forest (RF) classifier was utilized. RF constructs an ensemble of decision trees, combining their outputs via majority voting or averaging to improve predictive accuracy and reduce overfitting. Hyperparameters, including the number of estimators (200), maximum tree depth, and minimum samples per leaf, were optimized

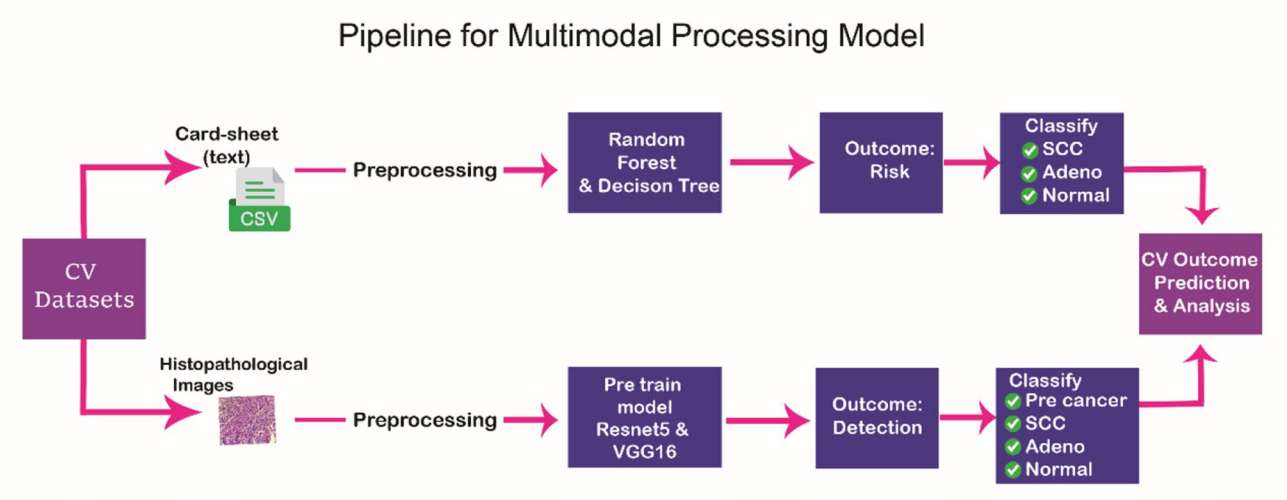

Fig. 2 Pipeline for multimodal processing model

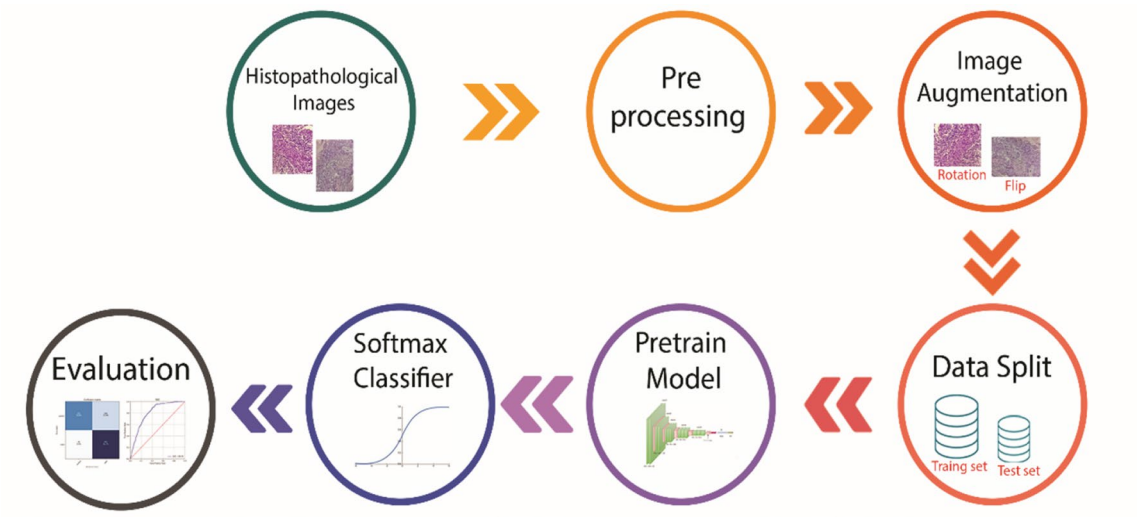

Fig. 3 Histopathological processing pipeline

Fig. 4 Patient card-sheet classification architecture

via grid search. RF was selected based on its proven performance in medical informatics [32]. The processing architecture is shown in Fig. 4.

Fusion strategy for multimodal integration

A late fusion strategy was applied to combine outputs from the two classifiers. Specifically, class probability vectors from the image-based model $P_{image}(C)$ and card-sheet model $P_{card}(C)$ were averaged:

$$P_{final}(C) = \frac{P_{image}(C) + P_{card}(C)}{2} \quad (1)$$

The class C with the highest average probability $P_{final}(C)$was selected as the final prediction. Late fusion was chosen over early fusion to preserve modality-specific interpretability and accommodate heterogeneous data formats. This approach also allows independent optimization of each branch of the model.

Step 4: model evaluation

Evaluation metrics

To comprehensively assess model performance, standard classification metrics were employed:

- **Accuracy**:

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

- **Precision**:

$$\frac{TP}{TP + FP} \tag{3}$$

- **Recall (sensitivity)**:

$$\frac{TP}{TP + FN} \tag{4}$$

- **F1-score**:

$$\frac{2 * Precision + Recall}{Precision + Recall} \tag{5}$$

- **AUC-ROC**: The Area Under the Receiver Operating Characteristic Curve was computed to quantify the trade-off between sensitivity and specificity across different thresholds.

Validation strategy

The dataset was randomly partitioned using an 80/20 stratified train-test split to preserve class balance. All performance metrics were reported on the held-out test set. While k-fold cross-validation was not performed in this study, it will be considered in future work to ensure model robustness.

Ablation studies

Ablation studies were conducted to identify crucial components of the model and various configurations of the dense layers to understand their impact on model performance. These studies involved experimenting with different dense layer configurations, such as 64, 128, 512, and 1024 units, and examining the impact of removing dense layers entirely. The results of these experiments provided valuable insights into refining the model, helping to simplify less impactful elements while focusing on components that significantly improved performance.

Results

A total of 404 cervical histopathology images were collected from two primary data sources (histopathological images and structured patient card-sheet data), covering four diagnostic categories squamous cell carcinoma (RF), adenocarcinoma, pre-cancer, and normal tissues. Additionally, structured records from 499 patients, containing demographic, behavioral, and clinical risk factors, were analyzed to support multimodal predictions.

Histopathology image enhancement and augmentation

Contrast Limited Adaptive Histogram Equalization (CLAHE) significantly improved visibility of nuclear and tissue architecture, benefiting subsequent classification tasks. Figure 5 presents the histopathology images, particularly in regions with subtle pathological features, after the enhancement. To address class imbalance and improve model generalization, the dataset was expanded through data augmentation, including 90° and 180° rotations, flips, and brightness changes, as illustrated in Fig. 6. These transformations increased the image count to 2,024. SMOTE was then applied to balance each class, resulting in 1,238 samples per class, totaling 4,952. The classification system includes four classes: class 0 represents Squamous Cell Carcinoma (SCC), and class 1 corresponds to pre-cancerous conditions of cervical tissue.

Patient Card-Sheet data balancing and feature insights

A more balanced dataset was created using SMOTE (generated new synthetic samples for the underrepresented classes) and Tomek Links (removed borderline samples from the majority class), resulting in 515 samples for squamous cell carcinoma, 363 for adenocarcinoma, and 315 samples for pre-cancer, as shown in Fig. 7.

Age-wise, the highest concentration of cervical cancer cases occurred in women in their 40s and 60s as shown in Fig. 8, with the calculated mean age being approximately 50 years. The average age was computed using the arithmetic mean: $\mu = \Sigma x/N$.

As Table 1 shows, in urban areas, there were 132 cases of squamous cell carcinoma (SCC) and 10 cases of adenocarcinoma, with no cases of pre-cancer. Conversely, rural areas showed a greater prevalence, with 336 cases of squamous cell carcinoma, 18 cases of adenocarcinoma, and three cases of pre-cancer.

Parity, defined as the number of times a woman has given birth, was found to be a significant factor in the distribution of cervical cancer types. As Table 2. shows, among nulliparous women (those who have never given birth), there were only eight reported cases of squamous cell carcinoma and no cases of adenocarcinoma or pre-cancer. In contrast, multiparous women (those who have given birth one or more times) had a markedly higher prevalence, with four hundred sixty-one cases of squamous cell carcinoma, twenty-eight cases of adenocarcinoma, and two cases of pre-cancer. Additionally, 96% of participants reported experiencing symptoms such as vaginal bleeding and discharge, with only a small number of cases lacking these symptoms.

Model training and evaluation

Baseline model

The baseline Random Forest model achieved 83% training accuracy and 65% test accuracy, establishing a reference point for comparison (Fig. 9).

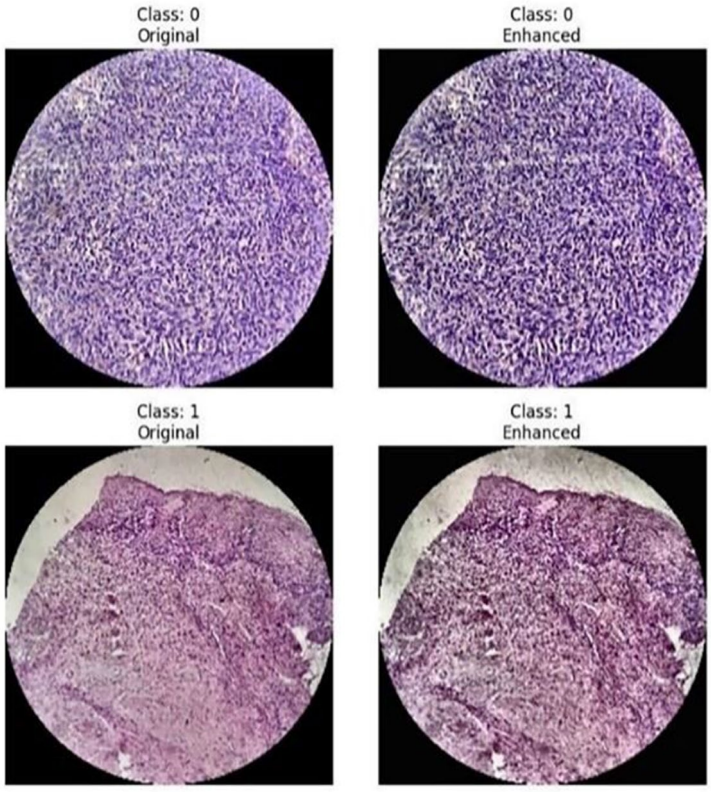

Fig. 5 Sample histopathology images (subtle pathological features)

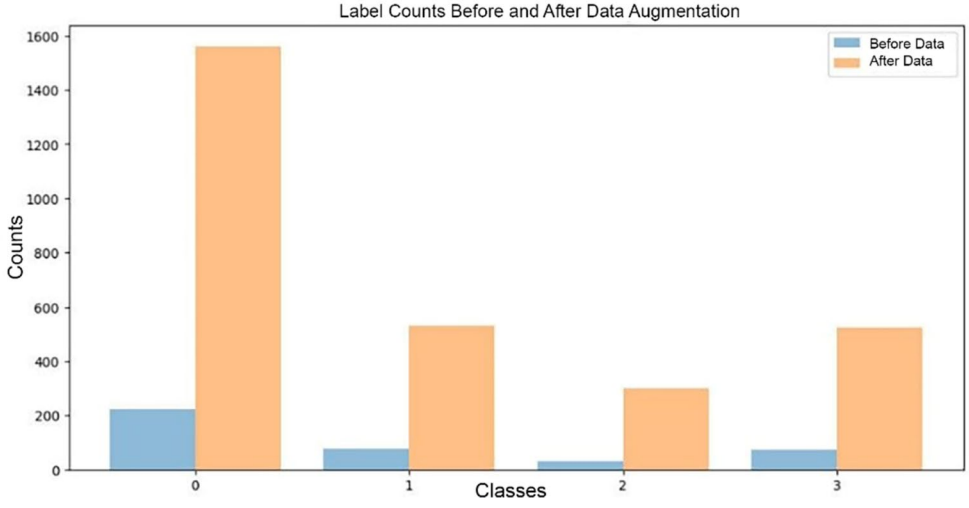

Fig. 6 Class visualization before and after data augmentation

Image classification

VGG16 model

The fine-tuned VGG16 model recorded a best validation loss of 0.2792. It achieved 91% test accuracy, with precision, recall, and specificity values of 90.7%, 90.4%, and 95.9%, respectively. The ROC-AUC score averaged 98.5% across classes (Figs. 10, 11 and 12). Figure 13 also presents the detailed results of the confusion matrix. showing classification results across cervical cancer classes. Class labels are defined as: 0 = SCC (Squamous Cell Carcinoma), 1 = Adenocarcinoma, 2 = Pre-cancer, and 3 = Normal. The Y-axis represents the actual class, and the X-axis shows the predicted class. Color intensity indicates the number of samples classified into each category.

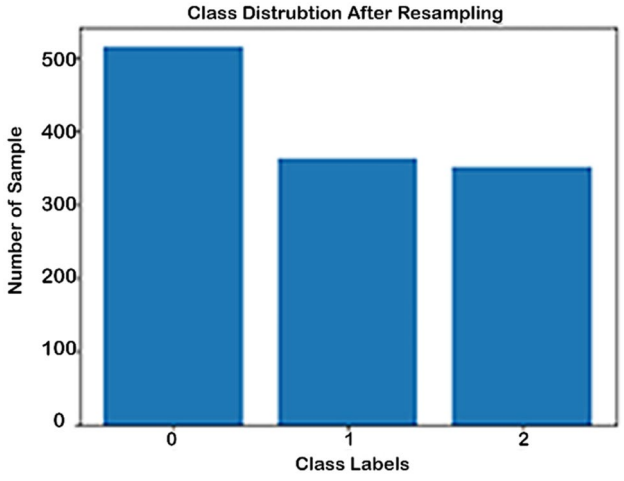

Fig. 7 Patient card-sheet class distribution after resampling

ResNet50 model

ResNet50 achieved a slightly higher training accuracy (95.6%) but underperformed during testing with an accuracy of 89%. It recorded a validation loss of 0.4317 at epoch 5. The model's recall, precision, and AUC were 86.7%, 85.9%, and 97.5%, respectively (Figs. 14, 15, 16 and 17).

Ablation study

VGG16 emerged as the most effective model, achieving an accuracy of 0.91, outperforming the baseline and ResNet50 as illustrated in Table 3. Consequently, the model was developed based on the VGG16 convolutional Neural Network.

In our ablation studies on the VGG16 model, removing all dense layers caused both training and test accuracy to

Table 1 Distributions of patient residential cross-cervical cancer

Class	Resident Area					
	Urban			Rural		
	Total	Mean	SD	Total	Mean	SD
SCC	132	0.929	0.256	336	0.941	0.235
Adeno	10	0.070	0.256	18	0.050	0.218
Pre cancer	0	0	0	3	0.0084	0.092

Table 2 Distributions of parity across cervical cancer

Class	Parity					
	Nulliparous			Multiparous		
	Total	Mean	SD	Total	Mean	SD
SCC	8	1	0	461	0.9391	0.0104
Adeno	0	0	0	28	0.0570	0.0105
Pre cancer	0	0	0	2	0.0041	0.0029

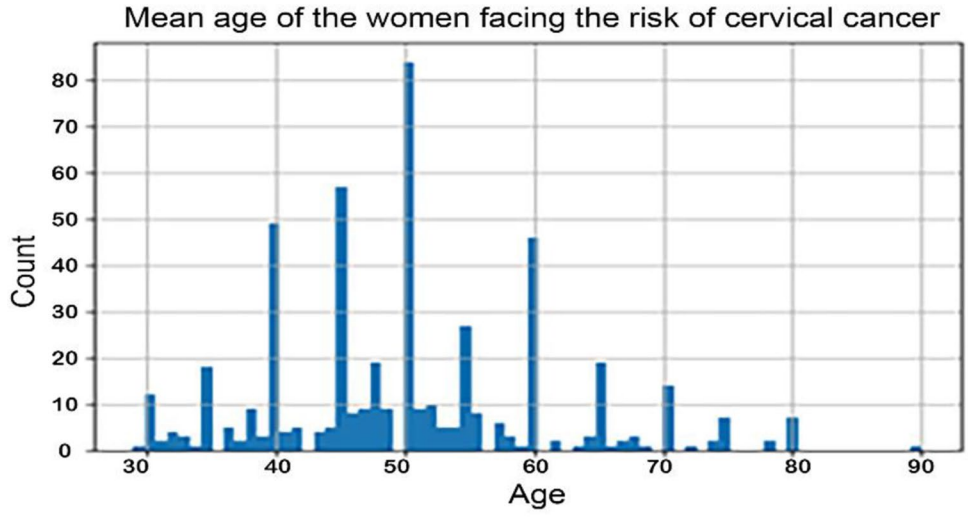

Fig. 8 Mean age of the women facing the risk of cervical cancer

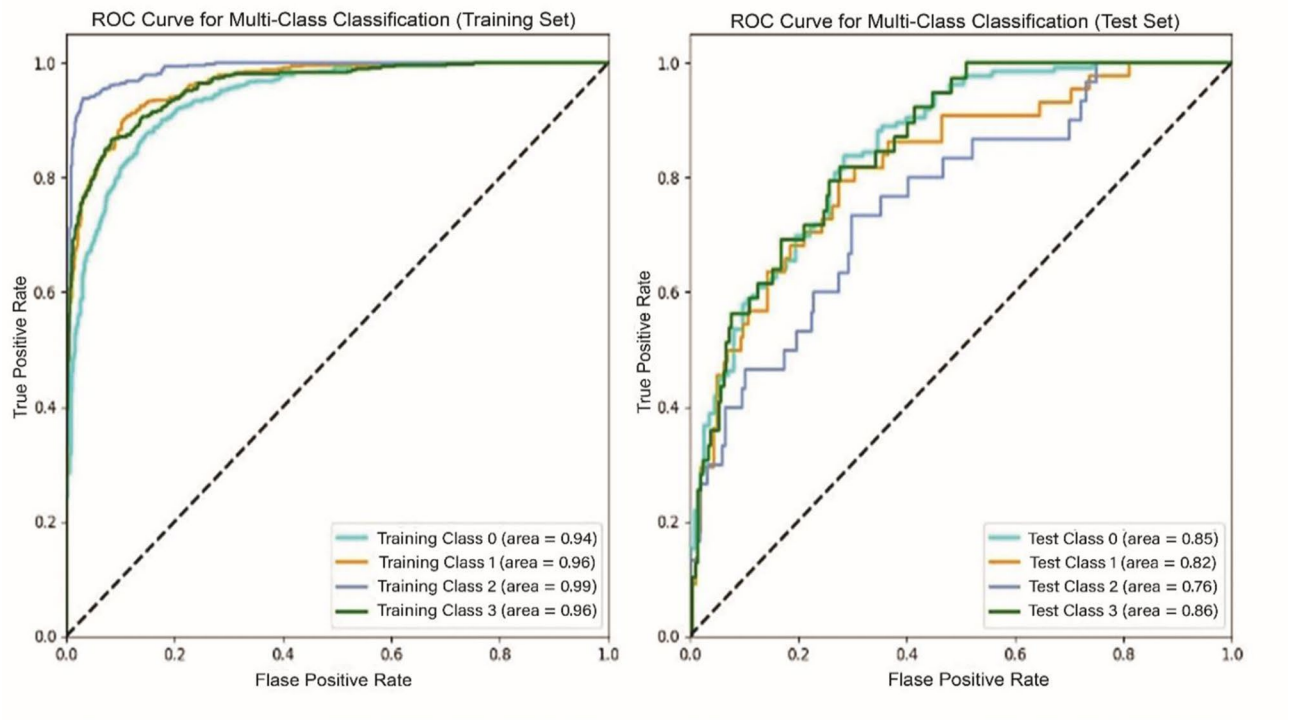

Fig. 9 ROC curve for multi-class classification of training and testing sets

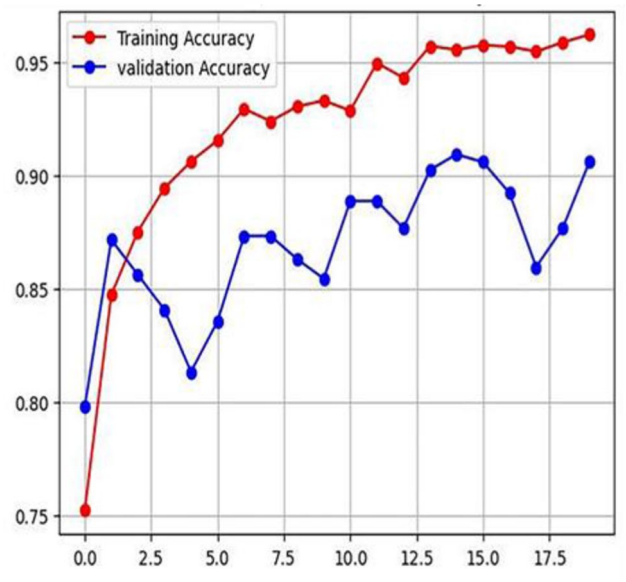

Fig. 10 Training and validation accuracy

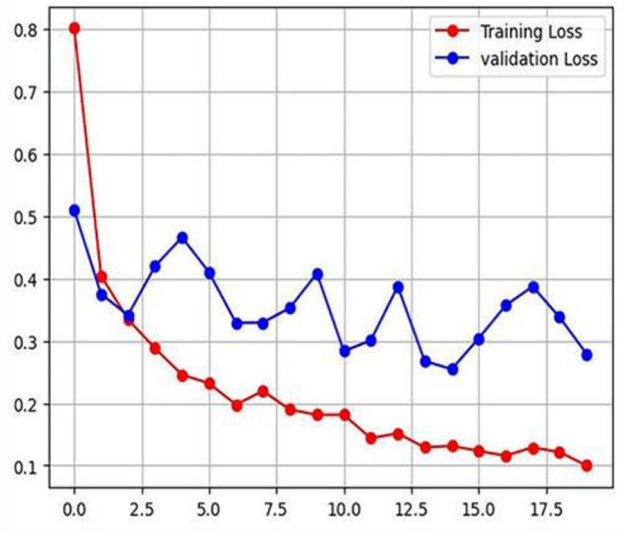

Fig. 11 Training and validation loss

drop to 83%, highlighting the importance of dense layers in capturing complex features. Increasing the dense layer size to 1024 improved training accuracy to 96%, but test accuracy only rose slightly to 85%. Reducing the dense layer size further to 256 and 64 resulted in a training and testing accuracy of 92% and 86%, respectively. Moreover, we found an accuracy of 93% and 89% for 256 and 64 units, respectively. Ultimately, a dense layer size of 512 provided the best balance, achieving 91% test accuracy alongside strong generalization.

Patient card sheet prediction

Random forest

We found RF demonstrated an accuracy of 96% on the patient card sheet dataset. More information on the detailed results and report of the RF confusion matrix is shown in Fig. 18; Table 3. Overall, the average AUC for the classification model across all classes is approximately

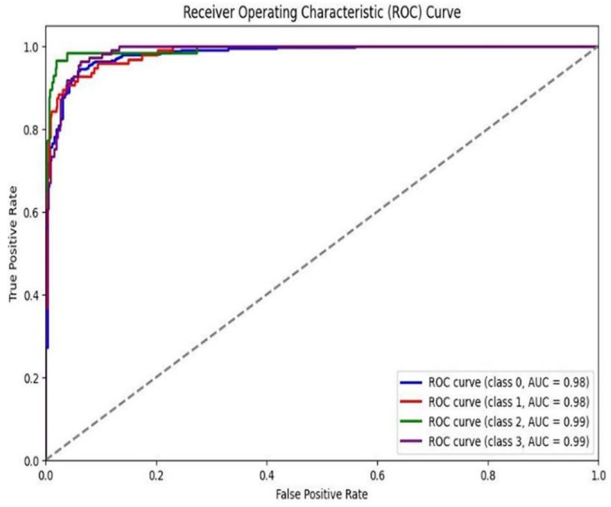

Fig. 12 Roc curves for VGG16

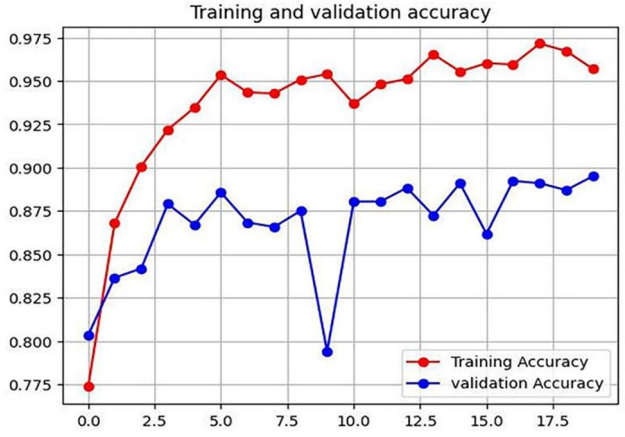

Fig. 14 ResNet50 model training and validation accuracy

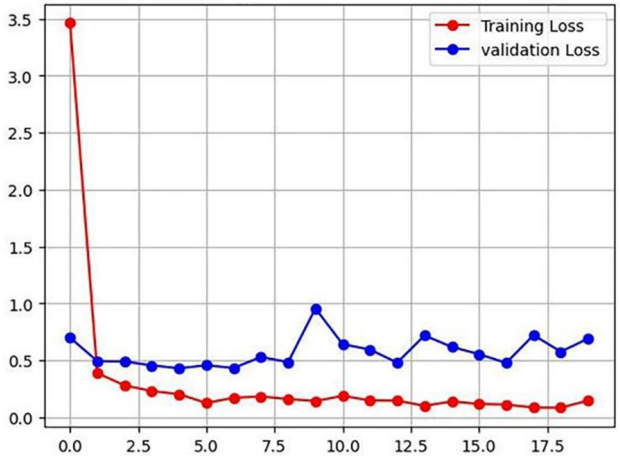

Fig. 15 ResNet50 model training and validation loss

98%. Figure 19 shows further information on the ROC curve of RF.

Decision tree

The decision tree classifier was applied to model the risk factors and obtained an accuracy of 95%. More information on the detailed results and report of the decision tree confusion matrix is shown in Fig. 20; Table 3. We found a 96% average AUC for the classification model across all classes. Figure 21 illustrates the experimental result of the decision tree ROC curve in detail.

Multimodal model

The integrated multimodal model achieved an accuracy of 92%. The details of the experimentation and

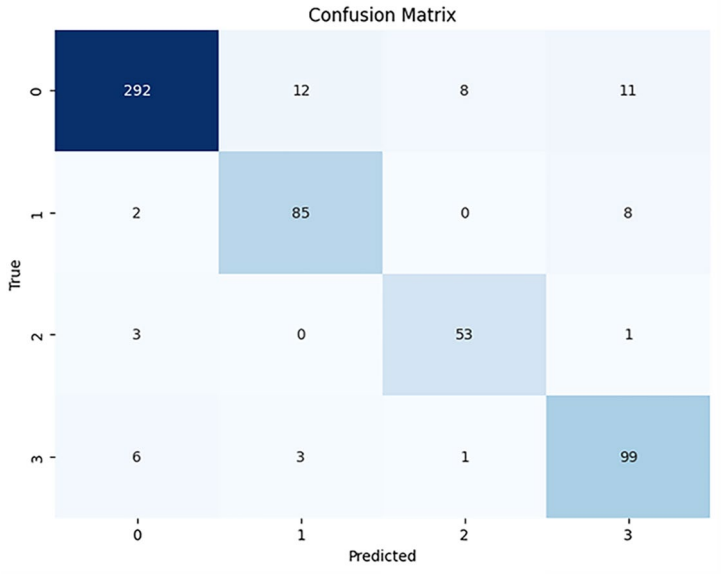

Fig. 13 VGG16 confusion matrix

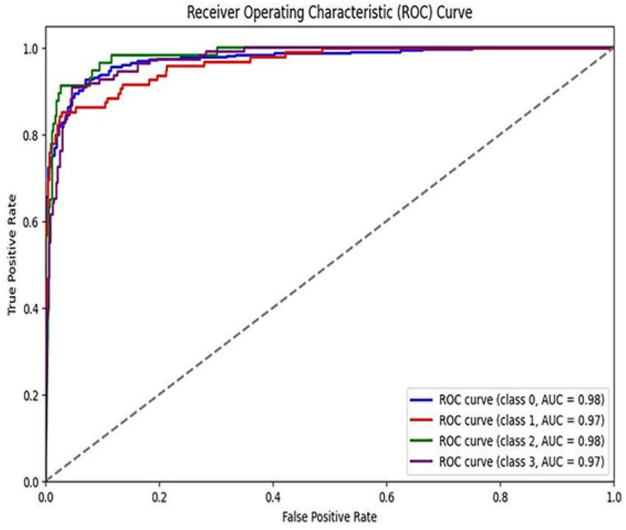

Fig. 16 ResNet50 ROC curve

classification report are shown in Table 3. The model performs particularly well for SCC and pre-cancer, with a precision of 94% and 96%, respectively, highlighting its strong precision and recall for these classes.

Discussion

This study developed and evaluated a multimodal cervical cancer detection model by integrating histopathological images with patient card-sheet data. The multimodal approach leveraged deep learning for image analysis alongside classical machine learning techniques for tabular patient data, yielding a robust diagnostic tool tailored for low-resource settings such as Ethiopia.

The dataset consisted of 404 cervical biopsy images and 499 patient records from Jimma Medical Center, revealing key epidemiological patterns. Women aged 40 to 60 years exhibited a higher risk of cervical cancer, particularly those reporting symptoms such as vaginal bleeding and discharge. Furthermore, women with a higher number of children were at an increased risk, particularly for squamous cell carcinoma, compared to those with fewer children. Another significant finding was the higher incidence of cervical cancer among women residing in rural areas, highlighting potential disparities in healthcare access between urban and rural populations.

Four classification models were developed: VGG16 and ResNet50 for image data, and Random Forest and Decision Tree for patient card-sheet data. Hyperparameters were optimized via grid search, with the best performance achieved using the Adam optimizer, a learning rate of 0.001, a batch size of 16, a dropout rate of 0.3, and 20 epochs.

In the initial experiments, Random Forest was used as a baseline model, providing a benchmark for evaluating the performance of subsequent deep learning and ensemble methods. Among the models tested, the VGG16 model yielded the best performance with 96.2% accuracy on image data, while Random Forest achieved 96% accuracy on patient card-sheet data. The Random Forest model shows more consistent performance across different classes. It achieves higher precision for the Adeno class and slightly better recall for the SCC class. Additionally, the F1-scores, which balance both precision and recall, are higher for the Random Forest in both the SCC and Adenocarcinoma classes. Given that the Random Forest

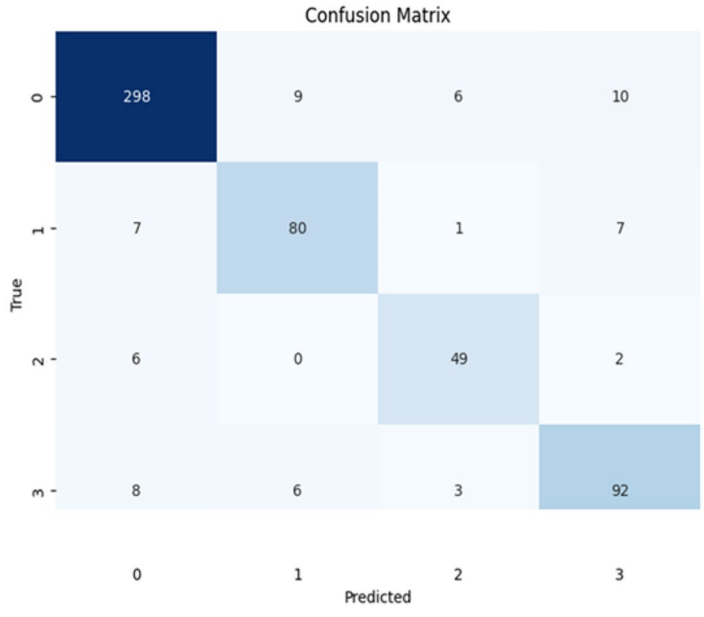

Fig. 17 ResNet50 confusion matrix

Table 3 Model experimentation results and classification report

Model	Category	Classification Report			
		Precision	Recall	F1-score	Support
VGG16	SCC	0.96	0.90	0.93	323
	Adeno	0.85	0.89	0.87	95
	Pre cancer	0.85	0.93	0.89	57
	Normal	0.83	0.91	0.87	109
	Accuracy	-		0.91	584
	Macro avg	0.88	0.91	0.89	584
	Weighted avg	0.91	0.91	0.81	584
Resnet50	SCC	0.93	0.92	0.93	323
	Adeno	0.84	0.84	0.84	95
	Pre cancer	0.83	0.86	0.84	57
	Normal	0.83	0.84	0.84	109
	Accuracy			0.89	584
	Macro avg	0.86	0.86	0.86	584
	Weighted avg	0.89	0.89	0.89	584
Random forest	SCC	0.96	0.95	0.96	116
	Adeno	0.95	0.98	0.97	59
	Pre cancer	0.96	0.96	0.96	71
	Accuracy			0.96	246
	Macro avg	0.96	0.96	0.96	246
	Weighted avg	0.96	0.96	0.96	246
Decision Tree	SCC	0.98	0.91	0.94	116
	Adeno	0.89	0.98	94	59
	Pre cancer	0.95	0.99	0.97	71
	Accuracy			0.95	246
	Macro avg	0.94	0.96	0.95	246
	Weighted avg	0.95	0.95	0.95	246
Model (Vgg16 + Random Forest)	SCC	0.94	0.95	0.94	323
	Adeno	0.88	0.87	0.88	95
	Pre cancer	0.96	0.93	0.95	57
	Normal	0.87	0.84	0.86	109
	Accuracy			0.92	584
	Macro avg	0.91	0.90	0.91	584
	Weighted avg	0.92	0.92	0.92	584

model performs marginally better overall, it was chosen to develop the final model. ROC curves illustrated the strong classification capability of VGG16, with a 98% AUC. The AUC-ROC curve for Random Forest across all classes was approximately 98%. The validation curves, shown in Fig. 14, demonstrate successful model convergence, with the highest validation accuracy achieved at epoch 20 and a low validation loss of 0.2792, indicating effective learning and generalization without overfitting.

In the ablation study on VGG16, we tested dense layer configurations with 64, 128, 512, and 1024 units, and analyzed the impact of removing dense layers. We found that removing layers led to decreased performance, both in training and test accuracy. The 512-unit configuration provided the optimal balance, underscoring the importance of fine-tuning layer sizes to achieve the best performance.

The multimodal model achieved a test accuracy of 92%. It performed especially well for squamous cell carcinoma and pre-cancer, achieving precision rates of 94% and 96%, respectively. While Adenocarcinoma and normal also demonstrated good performance, their precision was slightly lower compared to the squamous cell carcinoma and normal classes. This approach effectively integrated histopathology images with patient card-sheet data, resulting in a more reliable and adaptable model for cervical cancer detection.

Multimodal learning played a crucial role in overcoming the limitations of single-modal models, which often struggle with complex and diverse datasets. By combining image and text data, our model became more adaptable, allowing for more accurate predictions of new data. Previous studies have highlighted the challenges of traditional biopsy methods for cervical cancer diagnosis in Ethiopia, prompting the exploration of automated

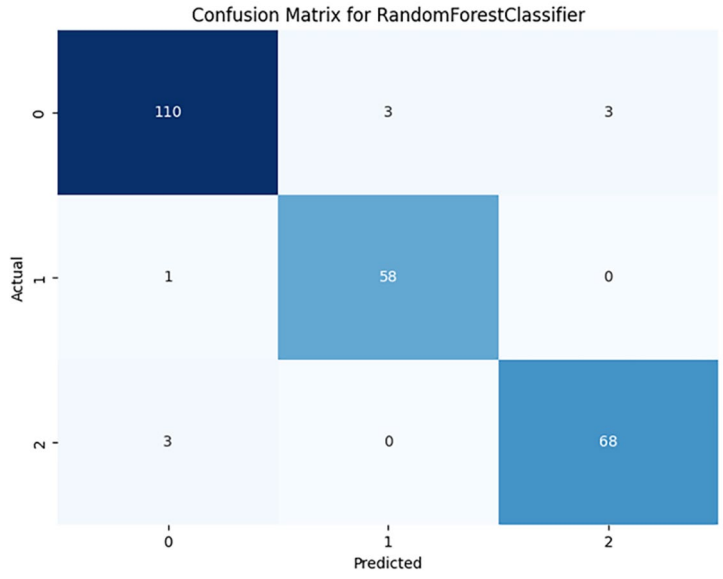

Fig. 18 Confusion matrix for RF model

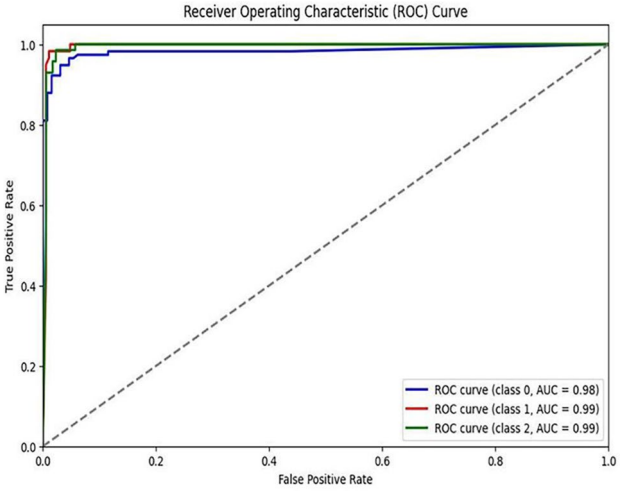

Fig. 19 ROC curve for RF model

diagnostic solutions. These solutions, particularly those combining multiple data sources, have the potential to revolutionize the diagnostic process by providing quicker, more accurate, and less invasive methods.

When compared to previous studies, such as Chen et al. [33]. and Ming et al. [17], our multimodal approach demonstrates comparable performance. Chen et al. introduced MultiFuseNet, achieving 87.4% accuracy in diagnosing cervical dysplasia with multimodal data. In contrast, our model outperformed this with higher classification accuracy, emphasizing the benefits of integrating histopathological images with patient card-sheet data.

Ming et al. focused on cervical cancer detection using FDG-PET/CT images and single-modal approaches. Our study extends this by incorporating both image and text data, providing a more comprehensive detection system. By combining visual and textual information, our model enhances the accuracy and robustness of cervical cancer detection, showcasing the effectiveness of multimodal data integration.

This multimodal approach addresses limitations of single-modal systems by improving adaptability and prediction accuracy in diverse clinical contexts. Automated diagnostics like ours could mitigate delays and errors inherent in manual biopsy interpretation, particularly in resource-constrained environments. The model's clinical utility lies in its potential for earlier detection and streamlined decision-making, which are vital for reducing cervical cancer morbidity and mortality.

The proposed multimodal system offers a promising diagnostic aid for low- and middle-income countries (LMICs). By combining non-invasive clinical data with biopsy images, it supports faster and more accurate screening that is deployable on affordable platforms suitable for rural clinics. This can reduce reliance on specialized pathologists and improve equity in healthcare access.

Conclusion

This study presents a novel multimodal approach for cervical cancer detection by integrating histopathological image analysis with patient card-sheet data. While the Random Forest model based on tabular patient data achieved the highest accuracy (96%), and the VGG16 model trained on biopsy images reached 91%, the proposed multimodal fusion model achieved a balanced accuracy of 92%. Despite this slightly lower accuracy, the multimodal model offers enhanced robustness and

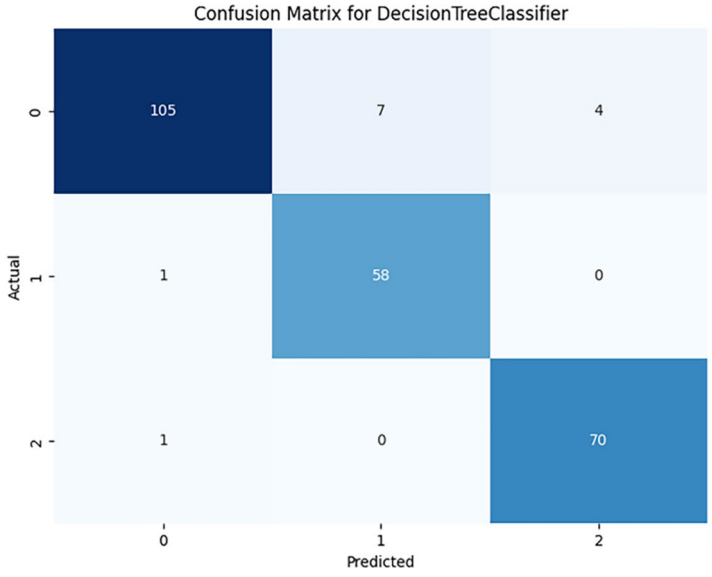

Fig. 20 Confusion matrix curve for the decision tree model

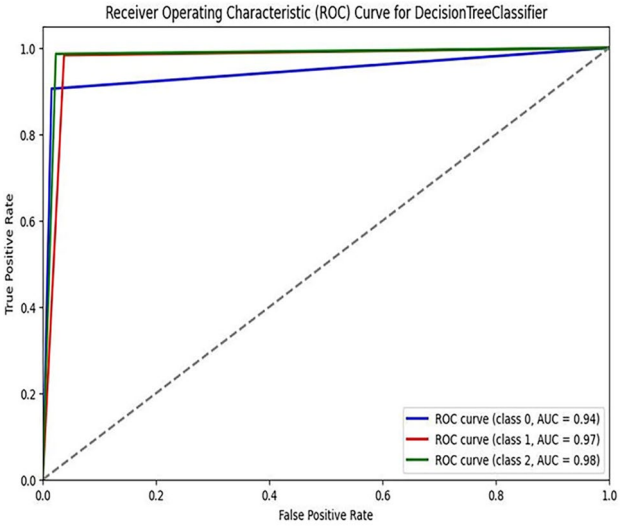

Fig. 21 ROC curve for the decision tree model

generalizability by combining visual pathology with clinical history, reflecting real-world diagnostic practice more closely. Theoretically, this study contributes to the growing body of research on multimodal learning for disease detection by demonstrating how the integration of heterogeneous data sources can support improved clinical inference, particularly in low-resource settings.

This research makes several key contributions. First, it introduces a multimodal diagnostic framework that integrates both image and tabular data to enhance cervical cancer detection. Second, it identifies and ranks critical clinical risk factors such as residential area, vaginal discharge, post-coital bleeding, irregular menstruation, age at menarche, STI: syphilis, and STI: HIV using machine learning based feature selection techniques. Third, it highlights geographic disparities in cervical cancer incidence, offering actionable public health insights into rural–urban differences in healthcare access. Finally, the study conducts ablation studies and comparative analyses, demonstrating the relative strengths of unimodal versus multimodal approaches in a clinical setting.

Despite its strengths, this study has several limitations. The dataset size, particularly for histopathological images, was relatively small, which may limit the model's generalizability to unseen populations. Moreover, the models were trained and validated using retrospective data, and prospective clinical evaluation is needed to confirm real-world utility. Additionally, while SMOTE and data augmentation techniques were employed to handle class imbalance, these synthetic techniques may not fully capture the diversity of real patient data.

To further enhance the effectiveness and applicability of multimodal cervical cancer detection models, future work should consider the following: (I) Dataset Expansion and Cross-Regional Validation: Utilize larger, more diverse datasets from multiple healthcare settings and populations to improve generalizability and model fairness, (II) Model Deployment and Accessibility: Explore lightweight architectures and mobile-based implementations to enable deployment in rural clinics or community health centers across LMICs, and (III) Clinical Integration and User Testing: Conduct usability studies and pilot deployments to assess model performance in real-time clinical workflows and gather feedback from healthcare professionals.

Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DL	Deep Learning
HIV	Human Immune Deficiency Virus
H & E	Hematoxylin and Eosin
HPV	Human Papillomavirus
JMC	Jimma Medical Center
ML	Machine Learning
NCD	Noncommunicable diseases
ResNet	Residual Network Visual
RF	Random Forest
SCC	Squamous cell carcinoma International
VGG	Visual Geometry Group
WHO	World Health Organization

Acknowledgements
We acknowledge Dr. Birhanu, Dr. Milkias, and Dr. Kassa from Jimma Medical Center (JMC), who managed the data collection and selection from beginning to end. The study would not have been possible without their participation and dedication. Furthermore, the resources required for this research were provided by the School of Computing and Informatics, Jimma Institute of Technology, Jimma University, and JMC.

Author contributions
K.C., G.S.T., and M.B. contributed to the conceptualization of the research goals, objectives, and methodology. K.C. was responsible for data curation, formal analysis, investigation, modeling, visualization, and drafting of the manuscript. G.S.T. and M.B. provided supervision. G.S.T. also contributed to the validation, review, editing, and final proofreading of the manuscript. All authors read and approved the final manuscript.

Funding
The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability
The datasets used and analyzed during the current study are available from the corresponding author on reasonable request and with the approval of Jimma Medical Center (JMC).

Declarations

Ethics approval and consent to participate
This study complies with the principles of the Declaration of Helsinki and was approved by the Jimma Institute of Technology, Research, and Ethical Review Board (IRB Reference No. RPD/JiT/183/15). The data used in this research were collected retrospectively from hospital patient card sheets and histopathological image archives. The need for informed consent was waived by the Jimma Institute of Technology, Research, and Ethical Review Board, by national research ethics regulations.

Consent publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 5 May 2025 / Accepted: 27 August 2025
Published online: 01 September 2025

References

1. Noncommunicable diseases, Accessed. Dec. 07, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases
2. Cervical cancer., Accessed. Dec. 07, 2024. [Online]. Available: https://www.who.int/health-topics/cervical-cancer#tab=tab_1
3. Mazeron JJ, Gerbaulet A. The centenary of discovery of radium. Radiother Oncol. Dec. 1998;49(3):205–16. https://doi.org/10.1016/S0167-8140(98)00143-1.
4. Mekuria M et al. Prevalence of cervical cancer and associated factors among women attended cervical cancer screening center at Gahandi Memorial Hospital, Ethiopia, Cancer Inform. 2021;20. https://doi.org/10.1177/11769351211068431
5. Lozano R. Comparison of computer-assisted and manual screening of cervical cytology. Gynecol Oncol. Jan. 2007;104(1):134–8. https://doi.org/10.1016/j.ygyno.2006.07.025.
6. Shahid AH, Singh MP. A deep learning approach for prediction of parkinson's disease progression. Biomed Eng Lett. May 2020;10(2):227–39. https://doi.org/10.1007/s13534-020-00156-7.
7. Battula KP, Chandana BS. Deep learning based cervical cancer classification and segmentation from pap smears images using an EfficientNet. Int. J. Adv. Comput. Sci. Appl. Dec. 2022;13(9):899–908. https://doi.org/10.14569/IJACSA.2022.01309104
8. Huang P, et al. Classification of cervical biopsy images based on LASSO and EL-SVM. IEEE Access. 2020;8:24219–28. https://doi.org/10.1109/ACCESS.2020.2970121.
9. Chandran V, et al. Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images. Biomed Res Int. 2021;2021. https://doi.org/10.1155/2021/5584004.
10. Kurnianingsih A, Allehaibi KHS, Nugroho LE, Widyawan L, Lazuardi L, Prabuwono AS, Mantoro T. Segmentation and classification of cervical cells using deep learning. IEEE Access. 2019;7:116925–41. https://doi.org/10.1109/ACCESS.2019.2936017.
11. Lalasa M, Thomas J. A review of deep learning methods in cervical cancer detection, in Proc. Int. Conf. Soft Comput. Pattern Recognit., Cham: Springer Nature Switzerland. Dec. 2022; pp. 624–633.
12. Namalinzi F, Galadima KR, Nalwanga R, Sekitoleko I, Uwimbabazi LFR. Prediction of precancerous cervical cancer lesions among women living with HIV on antiretroviral therapy in Uganda: a comparison of supervised machine learning algorithms. BMC Womens Health. Dec. 2024;24(1). https://doi.org/10.1186/s12905-024-03232-7
13. Habtemariam LW, Zewde ET, Simegn GL. Cervix type and cervical cancer classification system using deep learning techniques. Med Devices (Auckl). 2022;15:163–76. https://doi.org/10.2147/MDER.S366303.
14. Mohammed MA, Ali AM. Enhanced cancer subclassification using multi-omics clustering and quantum cat swarm optimization, Iraqi J. Comput. Sci. Math. 2024;5(3), Art. no. 37. https://doi.org/10.52866/ijcsm.2024.05.03.035
15. Ali M, Mohammed MA. Optimized cancer subtype classification and clustering using cat swarm optimization and support vector machine approach for multi-omics data, J. Soft Comput. Data Min. Dec. 2024;5(2): pp. 223–244. https://doi.org/10.30880/jscdm.2024.05.02.017
16. Ali AM, Mohammed MA. A comprehensive review of artificial intelligence approaches in omics data processing: evaluating progress and challenges. Int. J. Math. Stat. Comput. Sci. Dec. 2023;2: pp. 114–167.
17. Ming Y, Dong X, Zhao J, Chen Z, Wang H, Wu N. Deep learning-based multimodal image analysis for cervical cancer detection. Methods. Sep. 2022;205:46–52. https://doi.org/10.1016/j.ymeth.2022.05.004.
18. Abinaya K, Sivakumar B. A deep learning based approach for cervical cancer classification using 3D CNN and vision transformer. J Imaging Inf Med. Jan. 2024;37:280–96. https://doi.org/10.1007/s10278-023-00911-z.
19. Ejiyi CJ, Cai D, Eze FO, Ejiyi MB, Idoko JE, Asere SK, Ejiyi TU. Polynomial-SHAP as a SMOTE alternative in conglomerate neural networks for realistic data augmentation in cardiovascular and breast cancer diagnosis. J Big Data. 2025;12(1):1–28. https://doi.org/10.1186/s40537-025-01152-3.
20. Ejiyi CJ, Qin Z, Amos J, Ejiyi MB, Nnani A, Ejiyi TU, Agbesi VK, Diokpo C, Okpara C. A robust predictive diagnosis model for diabetes mellitus using Shapley-incorporated machine learning algorithms. Healthc Anal. 2023;3:100166. https://doi.org/10.1016/j.health.2023.100166.
21. López DM, Blobel B, Hullin C. Challenges and solutions for transforming health ecosystems in low- and middle-income countries through artificial intelligence. Front. Med. 2022;9, Art. no. 958097. https://doi.org/10.3389/fmed.2022.958097
22. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9(4):e1312. https://doi.org/10.1002/widm.1312.
23. Zeng M, Zou B, Wei F, Liu X, Wang L. Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data, in Proc. IEEE Int. Conf. Online Anal. Comput. Sci. (ICOACS). 2016: pp. 225–228, Sep. https://doi.org/10.1109/ICOACS.2016.7563084

24. Senan EM, et al. Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques. J Healthc Eng. 2021;2021. https://doi.org/10.1155/2021/1004767.
25. Mukku L, Thomas J, CMT-CNN. Colposcopic multimodal Temporal hybrid deep learning model to detect cervical intraepithelial neoplasia. Int J Adv Intell Inf. 2024;10(2):317–32.
26. Mukku L, Thomas J. Multimodal early fusion strategy based on deep learning methods for cervical cancer identification, in Proc. Congr. Intell. Syst. Singapore: Springer. Sep. 2023; pp. 109–118.
27. Russakovsky O et al. Sep., ImageNet large scale visual recognition challenge, 2014. Accessed: Dec. 07, 2024. [Online]. Available: http://arxiv.org/abs/1409.0575
28. Ma J, Fan X, Yang SX, Zhang X, Zhu X. Contrast limited adaptive histogram equalization-based fusion in YIQ and HSI color spaces for underwater image enhancement. Int J Pattern Recognit Artif Intell. Jul. 2018;32(7). https://doi.org/10.1142/S0218001418540186.
29. Building powerful image classification models using very little data. Accessed: Dec. 07, 2024. [Online]. Available: https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html
30. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/
31. Team K. GridSearch Tuner, Keras Team. Accessed: Nov. 06, 2024. [Online]. Available: https://keras.io/api/keras_tuner/tuners/grid/
32. Razali N, Mostafa SA, Mustapha A, Wahab MHA, Ibrahim NA. Risk factors of cervical cancer using classification in data mining. J Phys Conf Ser. Jun. 2020;1529(2). https://doi.org/10.1088/1742-6596/1529/2/022102.
33. Chen T et al. Apr., Multi-modal fusion learning for cervical dysplasia diagnosis, in Proc. IEEE Int. Symp. Biomed. Imaging (ISBI). 2019-Apr: pp. 1505–1509. https://doi.org/10.1109/ISBI.2019.8759303

Publisher's note