International Journal of Research Publication and Reviews, Vol 6, Issue 6, pp 4532-4540 June 2025

WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Implementation of Safety Device for Women

[1]Dr.Chidananda Murthy M V B. E, Ph. D [2] Kushala M, [3] Monisha K, [4] Rahul D, [5]Sahana G K

[1] Professor, [2]UG Student, [3] UG Student, [4]UG Student, [5]UG Student
Department of Electronics and Communication Engineering,
Sri Siddharatha Institute of Technology, Tumkur, India

A B S T R A C T

Cervical cancer is one of the leading causes of mortality among women worldwide, and its early detection is critical for improving patient outcomes. This work focuses on developing a machine learning-based system for the detection of cervical cancer using colposcopy images. By leveraging advanced image processing and classification techniques, the proposed system aims to address challenges such as diagnostic subjectivity, delays, and limited access to skilled clinicians, particularly in low-resource settings.

The proposed work incorporates preprocessing techniques to enhance image quality, balanced dataset creation using ADASYN, and a convolutional neural network (CNN) for feature extraction and classification. Comprehensive performance metrics, including precision, recall, and F1-scores, are used to evaluate the model's reliability and accuracy. The dataset, curated from credible sources, is meticulously pre-processed to ensure relevance and quality. Through an in-depth review of IEEE papers and related research, the system design is optimized for scalability and integration into clinical workflows. The proposed system holds promise for improving diagnostic accuracy, reducing human errors, and enhancing accessibility to advanced cervical cancer diagnostics.

Keywords: Cervical Cancer Detection, Colposcopy Images, Machine Learning, Data Imbalance, ADASYN, Clinical Diagnostics

1. Introduction

Cervical cancer remains one of the leading causes of cancer-related mortality among women, particularly in developing countries. Early detection plays a crucial role in improving survival rates, yet traditional screening methods such as Pap smears and manual inspection of colposcopy images often rely heavily on expert interpretation, which can be subjective and prone to error.

With the advancement of medical imaging and computational technologies, machine learning (ML) has emerged as a promising tool for enhancing the accuracy and efficiency of cervical cancer diagnosis. In particular, ML algorithms can assist in the automated analysis of colposcopy images by learning complex patterns associated with pre-cancerous and cancerous lesions.

In this study, we propose a machine learning-based framework for the detection of cervical cancer using colposcopy images. Our approach involves preprocessing the images to enhance relevant features, extracting discriminative features through CNNs, feature descriptors, and classifying the images into normal and abnormal categories using support vector machines, deep neural networks. The goal is to develop a robust and accurate diagnostic tool that can assist clinicians, reduce diagnostic time, and expand access to cervical cancer screening.

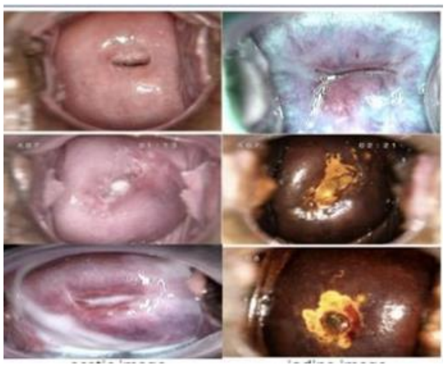

Fig 1. Detection of cancer in cervix

The image presents a comparative visual analysis of cervical tissue under different conditions, captured via digital colposcopy.

Acetic Acid Images (Left Column): After acetic acid application, abnormal epithelial cells undergo temporary whitening due to increased nuclear density, helping to highlight suspicious lesions (acetowhite changes).

Iodine Images (Right Column): Lugol's iodine binds to glycogen in normal squamous epithelial cells, staining them brown. Areas that do not uptake iodine (appearing yellow or non-stained) are indicative of possible precancerous or cancerous changes.

In the first row, the cervix appears relatively healthy, with minor changes.In the second row, white patches are visible after acetic acid and yellow-stained lesions are evident post-iodine application, suggesting dysplasia.

In the third row, clear non-staining areas and irregular lesion patterns are observable indicating higher-grade abnormalities.

The remainder of this paper is organized as follows: Section II reviews existing literature and applications in cervical cancer detection. Section III details the materials and methods, including data acquisition, preprocessing techniques, and the architecture of the proposed model. Section IV presents the experimental results and performance evaluation. Finally, Section V concludes the paper and outlines directions for future work.

2. Problem Statement

Cervical cancer remains a major global health concern, requiring early and accurate detection to improve patient outcomes.

Cervical cancer not only affects the individuals diagnosed but also has a profound impact on their families and society as a whole. The financial strain associated with cancer treatment further deepens social and economic inequalities.

Traditional diagnostic methods, such as Pap smears and HPV tests, are time- consuming, require specialized medical expertise and may not be accessible in resource-limited settings.

3. Proposed System

A. Existing method

Cervical cancer is a malignant neoplasm that originates in the epithelial cells lining the cervix, the lower part of the uterus that connects to the vagina. It typically develops over time from precancerous changes, which can be detected and treated before progressing to invasive cancer. The primary etiological factor is persistent infection with high-risk types of human papillomavirus (HPV), especially HPV-16 and HPV-18. According to the World Health Organization (WHO), cervical cancer remains one of the leading causes of cancer-related mortality among women globally, with a disproportionate burden in low- and middle-income countries due to limited access to preventive screening and treatment services.

Traditional Cervical Cancer Detection Methods:

Historically, the detection and diagnosis of cervical cancer have relied on cytological, histopathological, and clinical examination techniques. The most widely used traditional method is the Pap smear test (Papanicolaou test), in which exfoliated cells from the cervix are collected and examined microscopically for abnormalities in cell morphology indicative of precancerous or cancerous changes. While the Pap smear is effective in reducing cervical cancer incidence through early detection, it is subject to limitations such as human error, subjective interpretation, and variability in diagnostic accuracy.

Another significant method is colposcopy, a visual examination of the cervix using a colposcope to identify suspicious areas that may require biopsy. Histopathological examination of biopsy samples remains the gold standard for definitive diagnosis, providing information about the type, grade, and extent of malignancy. Additionally, HPV DNA testing is increasingly incorporated as a primary or adjunctive screening tool to identify high-risk HPV infections associated with cervical carcinogenesis. However, these traditional methods require specialized equipment, trained personnel, and are often time-consuming, limiting their scalability, especially in resource-limited settings.

Machine Learning-Based Cervical Cancer Detection

Recent advancements in artificial intelligence (AI) and machine learning (ML) have introduced promising alternatives and augmentations to traditional cervical cancer detection methods. Machine learning algorithms, particularly supervised learning techniques, can be trained on large datasets of medical records, cytology images, or clinical risk factors to identify patterns and predict the likelihood of cervical cancer with high accuracy and consistency.

Cervical cancer classification integrates both machine learning (ML) and deep learning (DL) approaches to enhance diagnostic accuracy. The dataset comprises 4,119 cervical images, categorized into negative and positive cases, obtained from different Cervicam models with varying resolutions. Data preprocessing includes grayscale conversion, cropping to focus on the cervical region, and normalization. Feature extraction is performed using pyradiomics, capturing 300 statistical and texture-based features through GLCM, GLRLM, GLSZM, and LoG filters. Key features are selected using the Lasso model to optimize classification performance. The ML classification models—XGBoost, SVM, and Random Forest—are trained and evaluated using a 5-fold cross-validation approach. In parallel, a deep learning-based classification model utilizes a ResNet-50 architecture pre-trained on ImageNet weights, fine-tuned with transfer learning.

B. Proposed Methodologyy

The proposed system aims to develop an efficient and accurate deep learning-based classification model for cervical cancer detection using Convolutional Neural Networks (CNNs). Unlike traditional machine learning approaches, CNNs can automatically learn complex patterns from data without requiring manual feature extraction, making them highly suitable for medical classification tasks.

The process begins with data collection and preprocessing, which includes handling missing values, feature selection, normalization, and splitting the dataset into training and testing sets. Feature selection will help identify the most relevant attributes contributing to cervical cancer classification, improving model performance and reducing computational complexity.

Once the data is pre-processed, a CNN model will be designed and trained using an 80-20 train-test split to learn patterns and relationships within the dataset. The model architecture will include multiple convolutional layers to extract hierarchical features, batch normalization for stable training, and dropout layers to prevent overfitting. Hyperparameter tuning will be performed to optimize the model for better accuracy and generalization.

To evaluate the effectiveness of the CNN model, performance metrics such as accuracy, precision, recall, F1-score, and the confusion matrix will be computed. After achieving satisfactory performance, the trained model will be tested on unseen data to validate its reliability and generalizability. The final model will be saved for future use and can be integrated into a user-friendly system for automated cervical cancer detection.

The ultimate goal of this system is to provide a robust and cost-effective deep learning-based solution for early cervical cancer detection. By automating the classification process, the proposed system aims to reduce manual diagnostic errors, accelerate diagnosis, and improve accessibility to healthcare services, particularly in underdeveloped and remote areas.

Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are:

- Convolutional layer
- Pooling layer
- Fully-connected (FC) layer

The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or pooling layers, the fully-connected layer is the final layer. With each layer, the CNN increases in its complexity, identifying greater portions of the image. Earlier layers focus on simple features, such as colors and edges. As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object.

There are three hyperparameters which affect the volume size of the output that need to be set before the training of the neural network begins. These include:

1. The number of filters affects the depth of the output. For example, three distinct filters would yield three different feature maps, creating a depth of three.

2. Stride is the distance, or number of pixels, that the kernel moves over the input matrix. While stride values of two or greater is rare, a larger stride yields a smaller output.

3. Zero-padding is usually used when the filters do not fit the input image. This sets all elements that fall outside of the input matrix to zero, producing a larger or equally sized output. There are three types of padding:

- Valid padding: This is also known as no padding. In this case, the last convolution is dropped if dimensions do not align.
- Same padding: This padding ensures that the output layer has the same size as the input layer
- Full padding**:** This type of padding increases the size of the output by adding zeros to the border of the input.

Fully Connected Layer:

The name of the full-connected layer aptly describes itself. As mentioned earlier, the pixel values of the input image are not directly connected to the output layer in partially connected layers. However, in the fully-connected layer, each node in the output layer connects directly to a node in the previous layer.

This layer performs the task of classification based on the features extracted through the previous layers and their different filters. While convolutional and pooling layers tend to use ReLu functions, FC layers usually leverage a softmax activation function to classify inputs appropriately, producing a probability from 0 to 1.

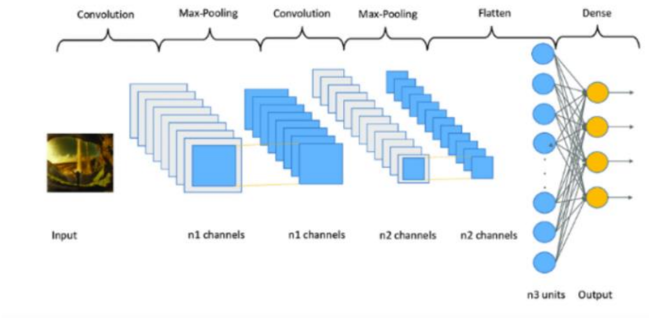

Fig 2 Convolution Neural Networks

C. Block diagram

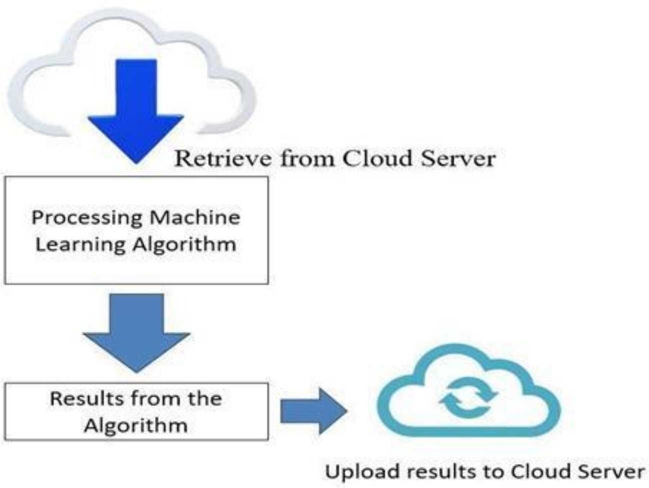

Fig 3. Block diagram of cancer detection

The given workflow diagram illustrates a cloud-based machine learning pipeline designed to facilitate data-driven predictions, such as those used in medical diagnosis systems like cervical cancer detection. The process begins with retrieving data from a cloud server. This data can include medical records, patient demographic details, clinical test results, or images — all securely stored and managed on the cloud. By accessing data from a centralized cloud server, the system ensures consistency, accessibility, and scalability, which are crucial when dealing with sensitive health information and large datasets.

Once the data is retrieved, it is fed into a processing system where a machine learning algorithm operates. This stage involves several sub-processes, including data preprocessing (handling missing values, normalization, and feature extraction), model loading, and predictive analysis. The machine learning algorithm, which may have been trained previously using historical data, processes the new input data and generates results — for example, classifying whether a patient falls into a high-risk or low-risk category for cervical cancer.

After the predictions are generated, the results are stored and uploaded back to the cloud server. This ensures that both healthcare professionals and remote applications can access the diagnostic outcomes conveniently and securely from different locations. The cloud server also enables further actions such as integrating the results with electronic health records (EHRs), notifying clinicians, or triggering alerts for urgent cases. This kind of cloud-integrated machine learning workflow improves efficiency, allows for real-time decision support, and enables scalable, automated diagnostic systems within the healthcare sector.

D. Flowchart

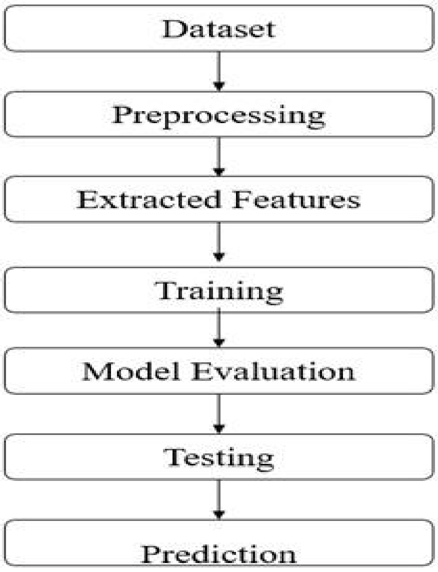

Fig 4. Flowchart of cancer detection

The proposed system follows a structured workflow for cervical cancer classification using deep learning, specifically Convolutional Neural Networks (CNNs). The workflow consists of several key stages, including data preprocessing, feature extraction, CNN model training, model evaluation, testing, and prediction.

1. Dataset Collection and Preprocessing

The process begins with collecting a dataset that contains cervical cancer-related clinical and demographic data. This dataset may include patient details such as age, risk factors, medical history, and extracted imaging features. Before applying the CNN model, the dataset undergoes preprocessing steps, including:

- Handling Missing Values: Missing data is managed using imputation techniques (e.g., mean, median, or mode replacement) or by removing incomplete records.
- Feature Scaling and Normalization: Ensures all features are within a uniform range to improve the efficiency of the model's learning process.
- Dataset Splitting: The dataset is divided into training (80%) and testing (20%) sets to evaluate the model's performance on unseen data.

2. Feature Extraction

Unlike traditional machine learning models that require handcrafted features, CNNs automatically learn feature representations from raw data. However, certain preprocessing techniques enhance the quality of features extracted from images:

- Convolutional Feature Learning: The CNN model extracts spatial and texture features from cervical cancer images without manual intervention.
- Data Augmentation: Techniques such as rotation, flipping, and contrast adjustment are applied to enhance model generalization and improve classification performance.

By leveraging deep learning, the system eliminates the need for manual feature extraction methods like Gray-Level Co-occurrence Matrix (GLCM) or Laplacian of Gaussian (LoG), as CNNs inherently learn meaningful representations directly from images.

3. CNN Model Training

The CNN architecture is designed to efficiently classify cervical cancer cases by learning hierarchical patterns from images. The model consists of multiple layers, including:

- Convolutional Layers: Extract spatial features from the input images using filters that detect patterns such as edges, textures, and shapes.
- Batch Normalization: Stabilizes and accelerates training by normalizing activations within the network.
- Pooling Layers: Reduce the spatial dimensions of feature maps while preserving essential information, improving computational efficiency.
- Dropout Layers: Prevent overfitting by randomly deactivating neurons during training.
- Fully Connected Layers: Combine extracted features to classify images into respective categories.

The CNN model is trained using an 80-20 train-test split. Hyperparameter tuning, including optimizing learning rate, batch size, and number of filters, is performed to enhance model accuracy.

4. Model Evaluation

Once trained, the CNN model is evaluated using performance metrics to assess its effectiveness in cervical cancer classification. The following metrics are used:

- Accuracy: Measures the overall correctness of the model's predictions.
- Precision: Evaluates the proportion of correctly identified positive cases among all predicted positives.
- Recall (Sensitivity): Assesses the model's ability to correctly identify actual positive cases.
- F1-score: The harmonic mean of precision and recall, providing a balanced measure of model performance.
- Confusion Matrix: A table that displays true positives, false positives, true negatives, and false negatives to analyze model errors.

These metrics help in understanding the CNN model's predictive power and effectiveness in real-world applications.

5. Testing Phase

After evaluation, the trained CNN model is tested on the 20% test dataset that was not used during training. This step ensures that the model generalizes well to new, unseen data. If significant performance differences are observed between training and test results, further fine-tuning is performed to address overfitting or underfitting issues.

6. Prediction

Once the final CNN model is trained and tested, it is used to make predictions on new patient data. Given a set of input images, the model predicts whether a patient is at risk of cervical cancer. The predicted output assists healthcare professionals in making informed decisions regarding further diagnosis and treatment.

By leveraging deep learning, the proposed system automates cervical cancer classification, reducing manual errors, improving diagnosis speed, and enhancing accessibility to healthcare services, particularly in underdeveloped and remote areas.

E. Formula

(a) Precision: A high precision score reflects a reduced number of false positives, indicating that most of the positive predictions are indeed correct. Precision is calculated using formula:

Precision=TP/(TP+FP)

where:

TP=TruePositives

FP=FalsePositives

TN=TrueNegatives

FN=FalseNegatives

(b) Sensitivity(Recall): Recall, also known as sensitivity, assesses a model's effectiveness in correctly identifying all actual positive cases. A higher recall value means the model successfully captures more true positives while minimizing false negatives. However, increasing recall can sometimes reduce Precision.

Recall = TP/ (TP+ FN)

(c)Sensitivity(Recall): Recall, also known as sensitivity, assesses a model's effectiveness in correctly identifying all actual positive cases. A higher recall value means the model successfully captures more true positives while minimizing false negatives. However, increasing recall can sometimes reduce Precision.

Recall = TP/ (TP+ FN)

(d)Accuracy: Accuracy represents the overall correctness of the model by indicating how many predictions were correct out of all predictions made. It is calculated using the formula:

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

4. System Components

- **HARDWARE REQUIREMENTS**

Personal computer: To train and test the model and display results.

- **SOFTWARE REQUIREMENTS**

Python idle 3.8: Python is an excellent choice as a first programming language without sacrificing the power and advanced capabilities that users will eventually need.

- **LIBRARIES**

OS: The os module in Python is a built-in library that provides a way to interact with the operating system.

Cv2 : The cv2 module in Python is part of the OpenCV (Open Source Computer Vision) library, which is widely used for image processing, and ML tasks.

Numpy: The numpy module in Python is a powerful library for numerical computing, providing support for large, multi-dimensional arrays and matrices and a collection of mathematical functions to operate on them.

Tkinter: Tkinter is the standard GUI (Graphical User Interface) library for Python, providing a simple and efficient way to create desktop applications.

Pandas: Pandas is a Python library used for data manipulation and analysis, offering data structures like Data Frame for handling structured data efficiently.

Scikit image: The scikit-image module in Python is an open-source library built on top of NumPy that provides a collection of algorithms for image processing and computer vision.

Pillow: Pillow is a widely used Python imaging library that provides extensive

image processing capabilities.

Tensorflow: TensorFlow is an open-source machine learning framework developed by Google that is widely used for applications such as image and speech recognition, natural language processing, time series forecasting, and object detection across various industries.

Keras: Keras is an open-source deep learning library written in Python, designed to provide an easy-to-use and modular interface for building and training neural networks.

5. Working Modules

This wearable safety device is specifically tailored for women, continuously monitoring vital health parameters and environmental factors through integrated sensors. It tracks key metrics such as heart rate, body temperature, and oxygen saturation to detect any signs of distress or irregular conditions.

The device offers both automated and manual emergency response options. If abnormal readings are detected or if the user feels at risk, they can activate the alert system by pressing a dedicated emergency button. Upon activation, the device uses GPS technology to pinpoint the user's exact location and sends this information, along with a distress signal, to a list of pre-selected contacts via a GSM module. Additionally, a loud buzzer is triggered to alert those nearby.

An onboard display provides real-time updates on health metrics and system status. Powered by a rechargeable battery and controlled by a central microcontroller, the device is compact, efficient, and designed for everyday use.

6. Result

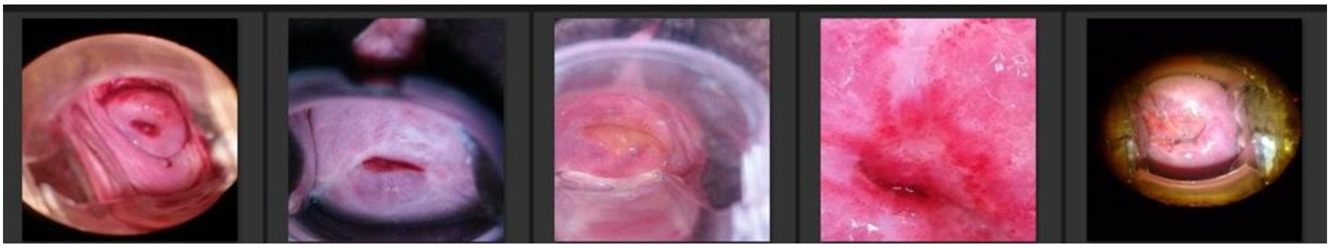

Fig.5 Images from Dataset

Each image in the dataset likely corresponds to a colposcopy view of the cervix, often captured after applying acetic acid. The different coloration, textures, and lesions seen in the images can be key indicators of disease or normalcy. These images are used for developing and training machine learning models to detect pre-cancerous or cancerous lesions on the cervix.

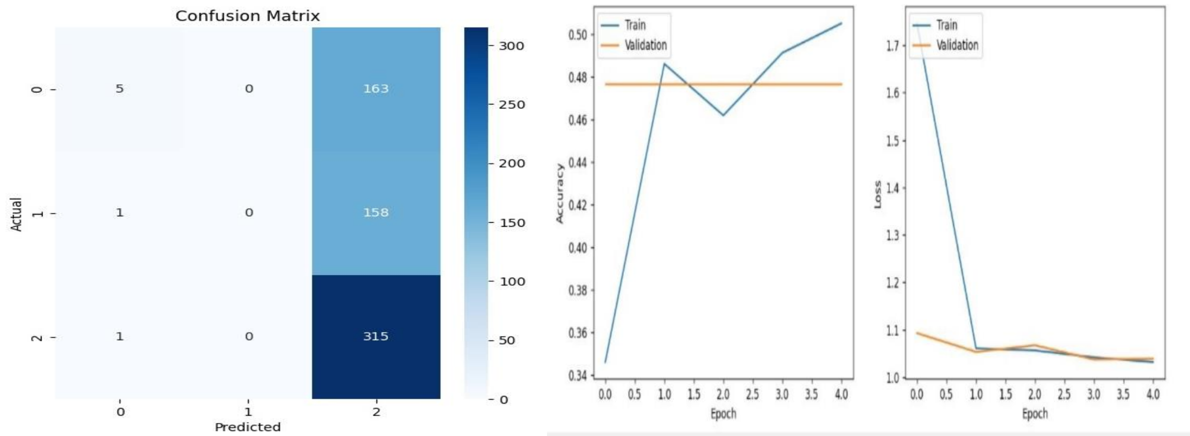

Fig.6 Graph from code

Training accuracy increases steadily, while validation accuracy stays flat around 48%, indicating limited generalization.

Training loss drops sharply at first, then plateaus along with validation loss, this suggests the model is learning the training data quickly.

Validation accuracy and loss are not changing much, means the model is underfitting.

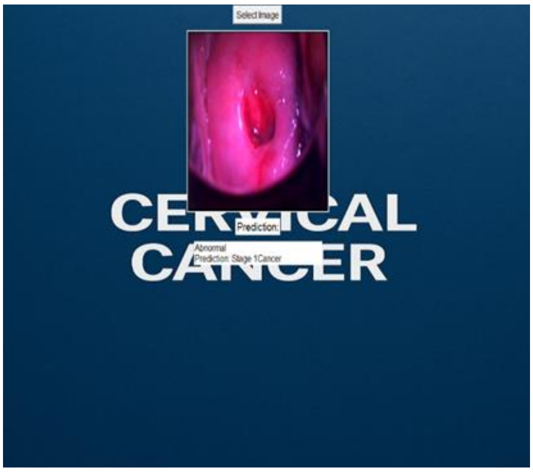

Fig.7 Detection Page

This is the doctor's web page where the cervix image is uploaded and prediction take place. The Predication shows on which stage does the cancer stand.

7. Observation

Confusion Matrix Analysis:

- The confusion matrix shows predictions across 3 classes (0, 1, 2), likely representing different diagnostic categories (e.g., Normal, Low-grade lesion, High-grade lesion or cancer).
- **Severe misclassification is evident**:
    - Almost all samples, including those from classes 0 and 1, were predicted as class 2.
    - Only 5 out of 168 actual class 0 samples and 1 out of 159 actual class 1 samples were correctly predicted.
    - Class 2 has the highest correct predictions (315), but the model is heavily biased toward predicting this class.
- **Imbalance in prediction**: This suggests the model has a **strong bias toward class 2**, possibly due to:
    - Class imbalance in the dataset.
    - Poor feature generalization from classes 0 and 1.

Accuracy and Loss Curve Analysis:

- **Training accuracy** improves steadily over 5 epochs, reaching above 0.5, while **validation accuracy remains flat (~0.47)**.
- **Training loss** drops sharply in the first epoch, then stabilizes, but **validation loss does not show meaningful improvement**.
- This indicates **overfitting**, where the model learns the training data but **fails to generalize to unseen validation data**.

Dataset and Model Context:

- The model is likely trained on colposcopy images, which contain visual patterns like acetowhite lesions or abnormal textures after acetic acid application.
- These visual features are crucial, but the poor performance suggests:
  - The model may not be effectively extracting or learning from those features.
  - More preprocessing, better data augmentation, or a deeper architecture might be needed.

8. Merits

- **Early Diagnosis**: Automated detection allows for quicker identification of cervical abnormalities, even in asymptomatic patients.
- **Decision Support**: Acts as a second opinion for healthcare professionals, improving diagnostic confidence.
- **Remote Screening**: Enables telemedicine-based diagnosis, beneficial for rural and underserved areas.
- **Cost-Effective Screening**: Reduces the need for extensive lab-based testing and specialist consultation in early stages.
- **Reduced Human Error:** Increases diagnostic accuracy and consistency by minimizing subjective interpretation.

9. Conclusion

The training phase of cervical cancer detection model is successfully completed using various machine learning techniques and image processing libraries. The model has shown promising accuracy on training data. However, further testing and validation are essential to ensure its reliability and generalizability on unseen, real-world data.

Future work will focus on:

- Thorough model evaluation on test datasets
- Enhancing accuracy and reducing false positives/negatives
- Potential deployment for practical clinical use

References

T. Šarenac and M. Mikov, ‘‘Cervical cancer, different treatments and importance of bile acids as therapeutic agents in this disease’’ Frontiers Pharmacol., vol. 10, pp. 484 510, Jun. 2019.

S. K. Singh and A. Goyal, ‘‘Performance analysis of machine learning algorithms for cervical cancer detection’’ Int. J. Healthcare Inf. Syst. Informat., vol. 15, no. 2, pp. 1 21, Apr. 2020.

S. M. A. Elsalam, O. Mokhtar, L. Adel, R. Hassan, M. Ibraheim, and A. Kamal, ‘‘Impact of diffusion weighted magnetic resonance imaging in diagnosis of cervical cancer’’ Egyptian J. Radiol. Nucl. Med., vol. 51, no. 1, pp. 1–8, Dec. 2020.

X. Hou, G. Shen, L. Zhou, Y. Li, and T. Wang, ‘‘Artificial intelligence in cervical cancer screening and diagnosis’’ Frontiers Oncol., vol. 12, no. 1, pp. 851367–851380, 2022.

M. Zhao, R.-Y. Gu, S.-R. Ding, L. Luo, Y. Jia, C.-X. Gao, B. Chen, X.-B. Xu, and H. F. Chen, ‘‘Risk factors of cervical cancer among ethnic minorities in Yunnan province, China: A case–control study’’ Eur. J. Cancer Prevention, vol. 31, no. 3, pp. 287–292, 2022.

Q. Wen, X. Wang, J. Lv, Y. Guo, P. Pei, L. Yang, Y. Chen, H. Du, S. Burgess, A. Hacker, F. Liu, J. Chen, C. Yu, Z. Chen, and L. Li, ‘‘Association between involuntary smoking and risk of cervical cancer in Chinese female never smokers: A prospective cohort study’’ Environ. Res., vol. 212, Sep. 2022, Art. no. 113371.

M. Anousouya Devi, J. I. Sheeba, and K. S. Joseph, ‘‘Neutrosophic graph cut-based segmentation scheme for efficient cervical cancer detection’’ J. King Saud Univ. Comput. Inf. Sci., vol. 34, no. 1, pp. 1352–1360, Jan. 2022.