 AMITY UNIVERSITY MUMBAI

### AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY

Department of Computer Science & Engineering

#### Academic Year 2024-25

Minor Project on

**“**CERVICAL CANCER SHAPE ANALYSIS**”**

*Submitted in partial fulfilment of the requirements for the degree of*

### Bachelor of Technology

Department of Computer Science & Engineering

#### Submitted By

|  |  |
| --- | --- |
| **Name** | **Enrolment no** |
| 1. **VEERBHADRESHWAR POL** 2. **JAY CHAVHAN**   **3. AARYA TELAKER**  **4. PIYUSH DESHMUKH**  **5. SOHAM PANDIT** | **A70405222181**  **A70405222186**  **A70405222189**  **A70405222180**  **A70405222184** |
|  |  |

Under the guidance of

**Dr. Monalisa Hati**

# Declaration of Academic Integrity

#### We declare that this written submission conveys our ideas in our own words. We have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/date/fact/source in our submission.

We understand that any violation of the above will be cause for disciplinary action by the institute and they can evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Student Signature

|  |  |
| --- | --- |
| 1. VEERBHADRESHWAR POL 2. JAY CHAVHAN   3. AARYA TELAKER  4. PIYUSH DESHMUKH  5. SOHAM PANDIT |  |

# Approval

#### This is to certify that “VEERBHADRESHWAR POL, JAY CHAVHAN, AARYA TALEKAR, PIYUSH DESHMUKH, SOHAM PANDIT” have satisfactorily completed their minor project on “CERVICAL CANCER SHAPE ANALYSIS” during the academic term 2024-25 and their report is approved for final submission.

**Examiners**

………………………

……………………….

……………………….

**Date: Place:**

# CERTIFICATE

#### This is to certify that the minor project entitled “CERVICAL CANCER SHAPE ANALYSIS” is a bonafide work of “VEERBHADRESHWAR POL, JAY CHAVHAN, AARYA TALEKAR, PIYUSH DESHMUKH, SOHAM PANDIT ” submitted to the Amity School of Engineering and Technology, Amity University Mumbai in partial fulfilment of the requirement for the degree of B. Tech Computer Science & Engineering.

**Supervisor Name Dr. Dipti Patil**

**HoI(CSE)**

**Dr. Shrikant Charhate Director, ASET**

**ACKNOWLEDGEMENT**

This minor project was a great experience of learning and professional development.

We take this opportunity to express our deepest gratitude and special thanks to our mentor,

Dr. Monalisa Hati, Faculty of the Department of Computer Science and Engineering branch at Amity University, Mumbai. Despite being occupied with their schedule, our mentor took out the time to hear, guide and keep us on the correct path. Her/His constant guidance and support throughout the course of this project have led to its timely and successful completion.

We would also like to express our deepest thanks to all the faculty members of Amity School of Engineering and Technology (ASET) for encouraging and motivating us throughout the B. Tech Program.

We would also like to express our deepest thanks to our HoI(CSE) Dr. Dipti Patil for her unwavering support.

We would also like to express our deepest thanks to our Director Dr. Shrikant Charhate who has been our pillar of support and motivation throughout the duration of the project.

Lastly, we would like to express wholehearted thanks to our family members, for supporting us with their love and guidance in whatever we pursue.

We perceive this opportunity as a big milestone in our career development. We shall strive to use the gained skills and knowledge in the best possible way and shall continue to work on their improvement, to attain the desired objectives.

## ABSTRACT

Cervical cancer is a major cause of cancer-related mortality in women around the world, especially in low-resource environments, and regular screening is difficult to access. Traditional diagnostic techniques such as Papanicolaou (PAP) are effective, but are time consuming and strongly based on expert interpretations that introduce subjectivity and potential errors. To address these limitations, this project will explore automated, form-based analytical methods for early detection of cervical cancer using digital histopathological imaging. By focusing on the morphological properties of the nucleus and the cytoplasm such as area, range, main and external axes, stretch, rounding, and nucleus, we aim to build a robust machine learning line that can distinguish between normal and abnormal cervical cells. For each cell, it was calculated in both Python (via the ski library) and in custom image processing routines. These characteristics were edited and used in structured data records (new\_shape.csv) to train and evaluate several classifiers, including random forests, support machines, and artificial neuron networks. Performance was measured using accuracy, F1 score, and ROC-AUC metrics. Among the models tested, artificial neural networks (ANNs) reached the highest classification accuracy of 94%, indicating the potential for form-based properties to support the diagnosis of cervical cancer. This approach not only reduces diagnostic workloads, but also introduces results consistency with results specific to the subjectivity of manual testing. Overall, this study examines form-based image analysis as a promising element in the development of a computer-aided diagnostic (CAD) system for screening for cervical cancer

The features were extracted from segmented microscopic images, and several shape descriptors were computed for each cell using both Python (via the skimage library) and custom image processing routines. These features were compiled into a structured dataset (new\_shape.csv) and used to train and evaluate multiple classifiers including Random Forest, Support Vector Machines, and Artificial Neural Networks. Performance was measured using accuracy, F1-score, and ROC-AUC metrics. Among the tested models, the Artificial Neural Network (ANN) achieved the highest classification accuracy of 94%, demonstrating the potential of shape-based features in supporting cervical cancer diagnosis.

In addition to high accuracy, the method shows robustness, computational efficiency, and interpretability, making it suitable for deployment in resource-limited settings. This approach not only reduces diagnostic workload but also introduces consistency in results by removing subjectivity inherent to manual examination. Overall, this study validates shape-based image analysis as a promising component in the development of computer-aided diagnostic (CAD) systems for cervical cancer screening

## TABLE OF CONTENTS

**CHAPTER NO. TITLE PAGE NO.**

**1. INTRODUCTION 10**

**1.1** Background

**2. LITERATURE SURVEY 12**

**2.1** Introduction

**2.2** Existing Methodologies

**2.3** Comparative Analysis

**3. PROBLEM STATEMENT 14**

**4. SYSTEM ANALYSIS 15**

**5. SYSTEM DESIGN 16**

**5.1** Design Model – Use case Diagram/

Class Diagram (Detailed Design)

Function Specifications (Data flow diagrams)

**6. PROJECT TIMELINE 21**

**6.1** Gantt chart

**7. IMPLEMENTATION, RESULTS AND TESTING**

**7.1** Details of Hardware and Software

**7.2** Result and Discussion

**8. CONCLUSION AND FUTURE SCOPE**

**8.1** Conclusion

**8.2** Future Scope

**REFERENCES**

**PUBLICATIONS AND CERTIFICATES**

**LIST OF FIGURES & IMAGES**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Figure/Image Title** | **Page No.** |
| **1.** | **Nucleus Mask** | **21** |
| **2.** | **Confusion Matrix** | **22** |
| **3.** | **Heat Map** | **22** |
| **4.** | **Accuracy and Specifity model** | **23** |
| **5.** | **Parameters** | **23** |

## ABBREVIATIONS

|  |  |
| --- | --- |
| ANN | Artificial Neural Network |
| ANFIS | Adaptive Neuro-Fuzzy Inference System |
| AUC | Area Under Curve |
| CAD | Computer-Aided Diagnosis |
| CNN | Convolutional Neural Network |
| FCM | Fuzzy C-Means |
| F1-score | Harmonic Mean of Precision and Recall |
| GLCM | Gray Level Co-occurrence Matrix |
| GPU | Graphics Processing Unit |
| GUI | Graphical User Interface |
| HPV | Human Papillomavirus |
| IDE | Integrated Development Environment |
| KNN | K-Nearest Neighbors |
| LSIL | Low-Grade Squamous Intraepithelial Lesion |
| ML | Machine Learning |
| N/C Ratio | Nucleus-to-Cytoplasm Ratio |
| PNG | Portable Network Graphics |
| RBF | Radial Basis Function |
| ROC | Receiver Operating Characteristic |
| SSD | Solid State Drive |
| SVM | Support Vector Machine |
| U-Net | A Convolutional Network Architecture for Biomedical Image Segmentation |

**1.INTRODUCTION**

Cervical cancer is the fourth most common cancer in women in the world, and it remains an important issue in developing countries, especially in developing countries where access to regular screening and early identification methods is limited. The World Health Organization estimates that in 2020 more than 600,000 new cases of cervical cancer were diagnosed worldwide, resulting in around 342,000 deaths. Despite the availability of effective precautions such as HPV vaccination and routine DAD smear testing, the disease continues to set serious stress due to infrastructure challenges and subjectivity, which is inherent to manual diagnostic materials. This method is based on trained cell engineers and pathologists to assess morphological properties of cervical cells, such as nuclear hypertrophy, irregular contours, increased ratio of the core core cytoplasm (N/C) and changes in chromatin distribution. This manual process is time-consuming and susceptible to fluctuations between observers, and is susceptible to human fatigue and error, especially when dealing with large quantities of samples. One of the most promising methods in this domain is form-based image analysis focusing on extracting geometric properties from cytological images to distinguish between normal and malignant cells. This approach is based on the observation that cancer transformation of cervical cells leads to different morphological changes, particularly the size, shape and boundary structure of the core and cytoplasm. Digitalized microscopic images of cervical cells are used to segment the core and cytoplasm and extract a comprehensive set of morphological features. This includes core area, cytoplasmic surface, N/C ratio, range, principal and secondary axes, stretching and rounding. Properties are quantified using image analysis tools and stored in structured data records for further modeling. This feature extraction is based on the PAP smoking classification, which serves as a basic reference to this work, based on the method described in the method described in the 2003 Eric Martin paper. The aim is to accurately classify cells into accurate and unusual categories based on geometric properties. Identify the most effective algorithm for this task by assessing classification performance through various models and metrics such as accuracy, F1 score, and ROC-AUC. The results not only show high classification accuracy (reaching 94% accuracy using ANN), but also highlight the interpretability and scalability of shape characteristics in clinical applications. Ultimately, our work supports the development of reliable, affordable screening tools that improve early detection, reduce diagnostic workloads, and improve health outcomes for women around the world. Cervical cancer progresses relatively slowly, often starting with a change in premia known as dysplasia. This is demonstrated early and can occur in invasive cancers if not treated. This delay provides valuable opportunities for screening and early intervention. In most developed countries, routine screening programs such as PAP-Breas have significantly reduced cervical cancer incidence and mortality. However, in many parts of the world, access to health infrastructure is limited, leading to delayed diagnosis and poor results. In this technique, excellent cells are collected from the cervix, colored using the Papanicolau method, and visually examined under a microscope by a trained cytologist. These experts evaluate a variety of cell functions, including core size, shape, chromatte texture, and core-cytoplasmic (N/C ratio) ratios. Papy Smears are effective, but their reliability depends heavily on human skills, experience and consistency. Considering the mere amount of foil processed in laboratory screening, each object carrier may contain hundreds of thousands of cells. There is a greater risk of monitoring and fatigue errors. These techniques aim to reduce diagnostic subjectivity, improve throughput, and achieve consistent results. Most automated systems start with image segmentation and separate important structures, such as the core and cytoplasm, from the background. After segmentation, characteristic extractions are performed to quantify relevant aspects of cell morphology. These properties include form-based descriptors (e.g. area, range, stretch, rounding), intensity-based metrics (e.g. brightness), and texture-based parameters derived from grayscale morphological abnormalities of cell nuclei such as increased size, irregular contours, and stretching, characteristic indicators of isovariate and errata. Similarly, higher N/C ratios often indicate cancer-like or cancer-like transformation. In contrast to texture and color features, the shape features of different colours and imaging conditions are relatively constant, making them ideal for comparison with cross samples.

1. Literature Survey

Over the past 20 years, the interface between computer vision and medical diagnosis has opened new opportunities for the automation of complex, labor-intensive tasks, such as cytological screening for cervical cancer. Many studies have considered the use of digital image processing and machine learning to extend or replace manual analysis of pup swirling samples. This literature review checks the most important developments in this field, with a particular focus on form-based property extraction and classification techniques. Introduced in the mid-20th century, Papanicolau (PAP) mixed testing, when performed regularly, significantly reduced cervical cancer mortality. Diagnosis is based on expert evaluation of morphological features, including core size, cytoplasmic surface, chromattouring texture, and area ratio from nucleus to cytoplasmic (N/C). However, manual inspection is not only time-consuming, but also under fluctuating interpretations that motivate the development of automated systems. Early commercial systems such as AutoPap and Papet used authenticity algorithms to obtain foils before screens for human reviews. These systems showed certain advantages, but their limited accuracy and high cost negate widespread acceptance. University of Denmark. Martin used Pope smear images from Herlev University Hospital to classify cervical cells by implementing both fuzzy C-mean (FCM) and Gustafson boiler clustering. We introduced a monitored clustering method to improve classification accuracy, focusing on properties such as core region, cytoplasmic range, N/C ratio, and brightness. Martin's study showed that supervised FCM achieved a false positive rate of 2.02% and a false negative rate of 1.38% on older data records, setting a high benchmark for further research. His research showed that Gustafson Kessel Clustering better reduces cell classification in seven diagnostic categories than FCM. However, both studies recognized limitations on the use of fixed cluster numbers, highlighting the need for parameter optimization. He showed that hierarchical classification approaches could exceed direct classification, especially in data records where class distributions overlap. The texture features, including those that come from gray levels in Ko-Ereffen-Matzen (GLCM), have added an additional layer of discriminatory force, but more

2.1 Introduction  
Cervical cancer is one of the most avoidable forms of cancer, but the main cause of death in women remains. Early recognition played an important role in reducing mortality, and traditional cytological methods such as the Pope smear were effective, but were limited by human dependence and variation. To address these challenges, researchers have developed automated methods to analyze cell morphology using computer vision and machine learning. These initiatives fall into three main categories: Rule-based systems, characteristic-based classifiers for machine learning, and deep learning models. Among them are form-based analyses based on interpretability, low computing costs and robustness compared to variations in coloring and image quality as particularly effective approaches. This section examines methods developed over time and provides a comparative perspective on performance and practical applicability. Below you can find the classification of the main sentences:

2.2.1 Periodically Based Statistical Methods  
Previous systems were based periodically and based on functional thresholds such as core size and N/C ratio. These methods required expert-defined rules and lacked generalization. Although arithmetically simple, its accuracy was limited due to sensitivity to noise and overlapping feature distributions. Gustafson Kettle Clustering on Herlev Hospital Data. He used shape features such as area, scope, major/next axle, rounding, and N/C ratio. Clustering monitored at FCM <2% was falsely positive for older data records, achieving a false negative of ~1%. High quality features are extracted and functional rooms are clearly defined. They provide interpretability and are suitable for clinical validation. These approaches often complement the features of the shape, but are intensively sensitive to image quality and arithmetic. Newer models such as U-NET for segmentation and VGG/residual-based classifiers for diagnostics show excellent performance. However, the challenges include:  
Requirement for large annotated datasets  
  
High computational cost (training on GPUs)  
  
Reduced interpretability, making clinical adoption harder  
  
2.3 Comparative Analysis  
For many years, several methods have been proposed for the automatic classification of cervical cancer, each with its own advantages and limitations. The initial rule-based approach was based on manually defined thresholds for properties such as core size and N/C ratio. Although arithmetic simple, these systems were extremely sensitive to noise and image variation, and had limited diagnostic accuracy. The method for classical characteristic-based machine learning showed a significant improvement and allowed for more subtle classification by using a set of form-based features. Erik Martins 2003's work is in this category using the leading Fuzzy-C-Means teamed up at Herlev Hospital Data. His approach achieves high levels of accuracy with properties such as core region, cytoplasm, rounding, stretching, and N/C ratio, falsely positive speed of 2.02%, only 1.38%, only 1.38% on older data records. However, performance lacked new data records, highlighting the challenge of generalizability. Landwehr added official grass-based features and showed that hierarchical classification with K-Nearest Neighbor (KNN) further improves performance. Although effective, the inclusion of texture properties has increased the complexity and dimensions of computing promotion. This can be a drawback of resource limit settings. These models like Xu et al. (2016) reached an accuracy level of up to 96%. However, it requires large commented data records and a high-performance computing infrastructure. This means it is not suitable for low resource environments. Furthermore, black box nature raises concerns about interpretability and clinical acceptance.

1. PROBLEM STATEMENT

Despite progression of screening for cervical cancer, manual examination of PAP smears-objects remains the primary method for the detection of pre-row cancer cells and cancer cells. This process is time-consuming and susceptible to variability among observers, limited by the availability of trained cytologists, particularly in resource-limited areas. Automated methods with deep learning and texture analysis have been shown to be promising, but often lack large data records, high arithmetic resources, and interpretability that hinders clinical applicability. There is an urgent need for efficient, accurate and interpretable automated systems, which allow neck cells to be classified primarily based on form-based features that are mathematically light and clinically meaningful. The aim of this project is to develop a robust form-based analytical pipeline that extracts morphological properties from cell images of the neck and extracts observed learning techniques for effective classification in normal and abnormal categories.

1. **SYSTEM ANALYSIS**

The main goal of this system is to provide a computer-aided diagnostic tool that can demonstrate and classify cervical cells based on morphological features with specific morphological parameters. The current manual system of Pope Spark Analysis is time-consuming, requires domain expertise, and is susceptible to interobserver variation. The purpose of this project is to overcome these challenges by implementing an automated form-oriented pipeline using digital image processing and classic machine learning techniques. The input consists of microscopic images of cervical cells, which are washed and enhanced to improve cell boundary visualization. After processing, the segmentation algorithm removes the core and cytoplasm. The form descriptors such as surface, range, elongation, circularity, and core (N/C) ratio are then calculated. These properties are usually found in monitored learning algorithms such as logistic regression, k-near-neighbors ( knn), or support vector machines (SVMs), for classifying benign or malignant categories. Regional Programs) and Classification. The model output was evaluated using performance metrics such as accuracy, accuracy, recall, and F1 score. The essential insight into this phase was to provide a sufficiently accurate and arithmetically efficient alternative to deep learning, especially when labeled data is limited.

1. **SYSTEM DESIGN**

The primary objective of the system is to provide a computer-aided diagnostic tool that can detect and classify cervical cells based on morphological features, particularly shape-based parameters. The current manual system of Pap smear analysis is time-intensive, requires domain expertise, and is prone to inter-observer variability. This project aims to address those challenges by implementing an automated, shape-focused pipeline that leverages digital image processing and classical machine learning techniques.

The system follows a modular architecture consisting of image acquisition, preprocessing, segmentation, feature extraction, and classification. The input consists of microscopic images of cervical cells, which are then cleaned and enhanced for better visualization of cell boundaries. Following preprocessing, segmentation algorithms isolate the nucleus and cytoplasm. Shape descriptors such as area, perimeter, elongation, roundness, and the nuclear-to-cytoplasmic (N/C) ratio are then computed. These features are input to supervised learning algorithms such as logistic regression, K-Nearest Neighbors (KNN), or Support Vector Machines (SVM) for classification into normal, benign, or malignant categories.

The system analysis phase involved evaluating various techniques and tools for segmentation (e.g., thresholding, contour detection), shape feature extraction (e.g., scikit-image region props), and classification. Performance metrics such as accuracy, precision, recall, and F1-score were used to evaluate model performance. A key insight from this phase was that shape-based features alone provided a sufficiently accurate and computationally efficient alternative to deep learning, especially when labeled data was limited.

SYSTEM DESIGN

The design of the system follows a structured, pipeline-based architecture with the following components:

**Input Layer:**

* Accepts images of cervical cells in digital format (e.g., PNG, JPEG).
* Data is sourced from benchmark datasets or hospital archives.

**Preprocessing Module:**

* Converts images to grayscale and applies noise reduction (e.g., Gaussian blur).
* Enhances contrast and sharpens cell boundaries.
* Resizes and normalizes images for uniformity.

**Segmentation Module:**

* Uses thresholding, morphological operations, and contour detection to separate the nucleus and cytoplasm.
* Labels regions of interest (ROI) using connected component analysis or watershed algorithm.

**Feature Extraction Layer:**

* Computes geometric and morphological shape features including:
* Nucleus Area
* Cytoplasm Area
* N/C Ratio
* Perimeter
* Major and Minor Axis Lengths
* Roundness
* Eccentricity
* Uses regionprops from scikit-image and custom Python functions.

**Classification Module:**

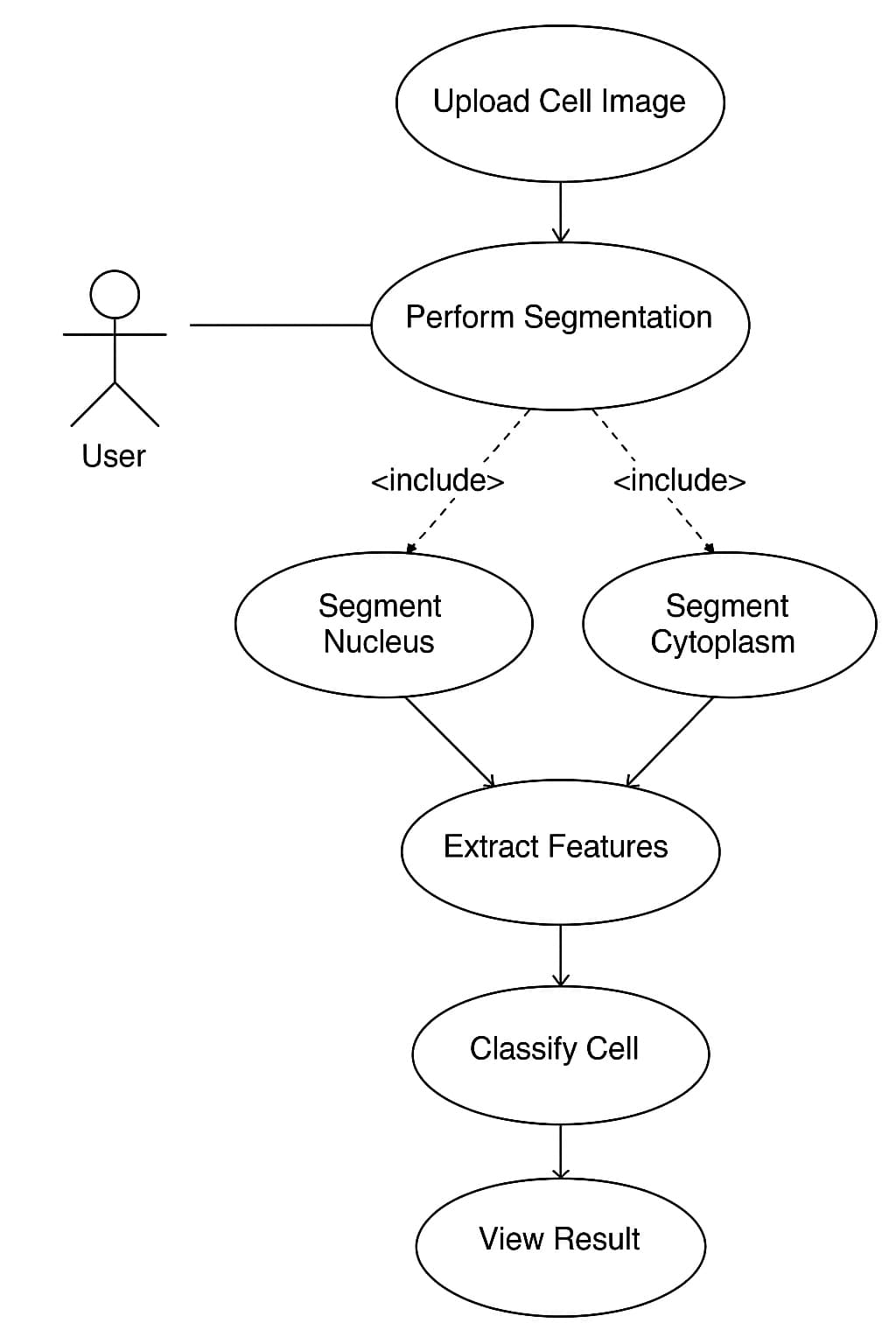
* Implements supervised ML models (Logistic Regression, KNN, SVM).
* Trained on labeled shape feature datasets.
* Outputs diagnostic category: Normal, Benign, or Malignant.

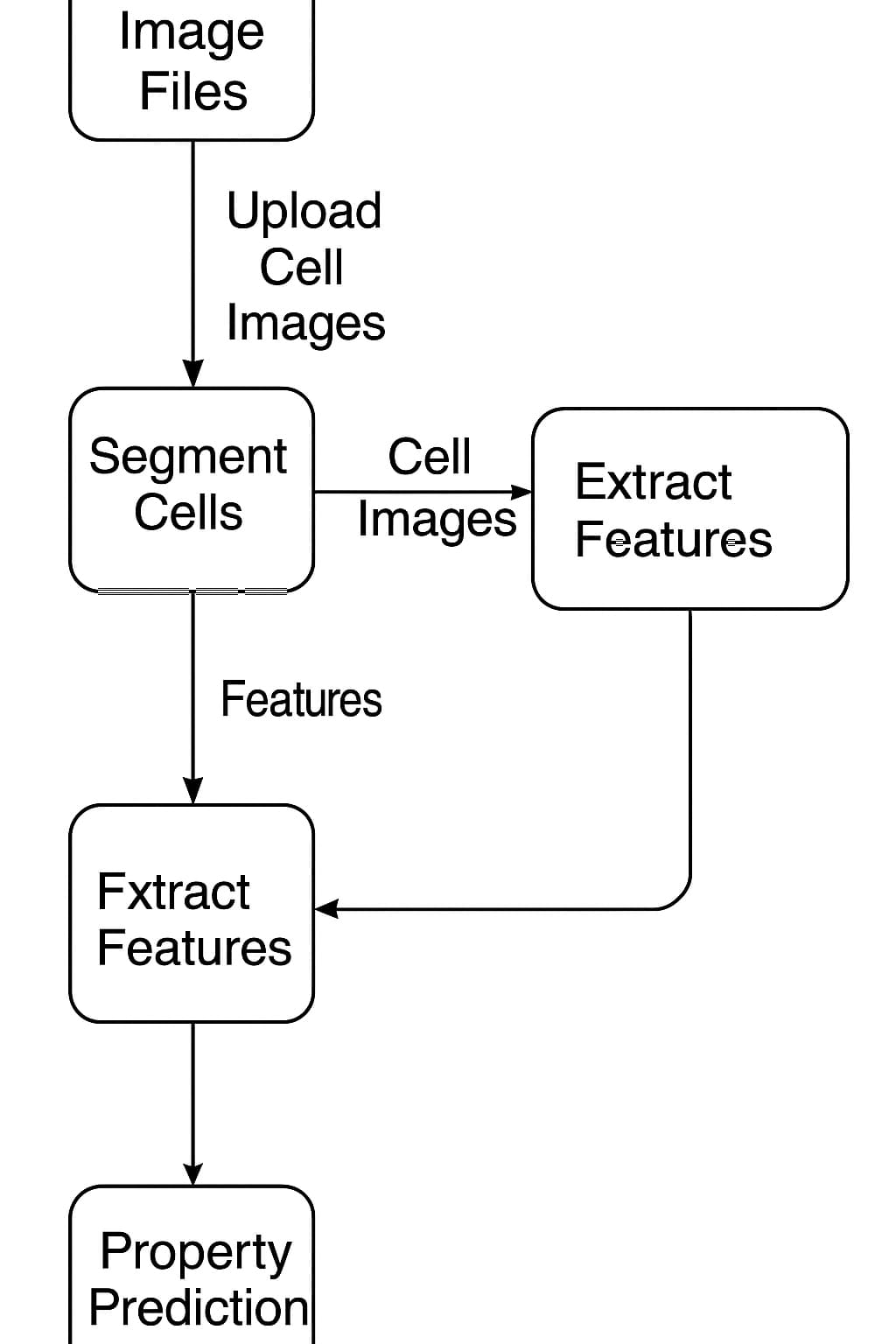
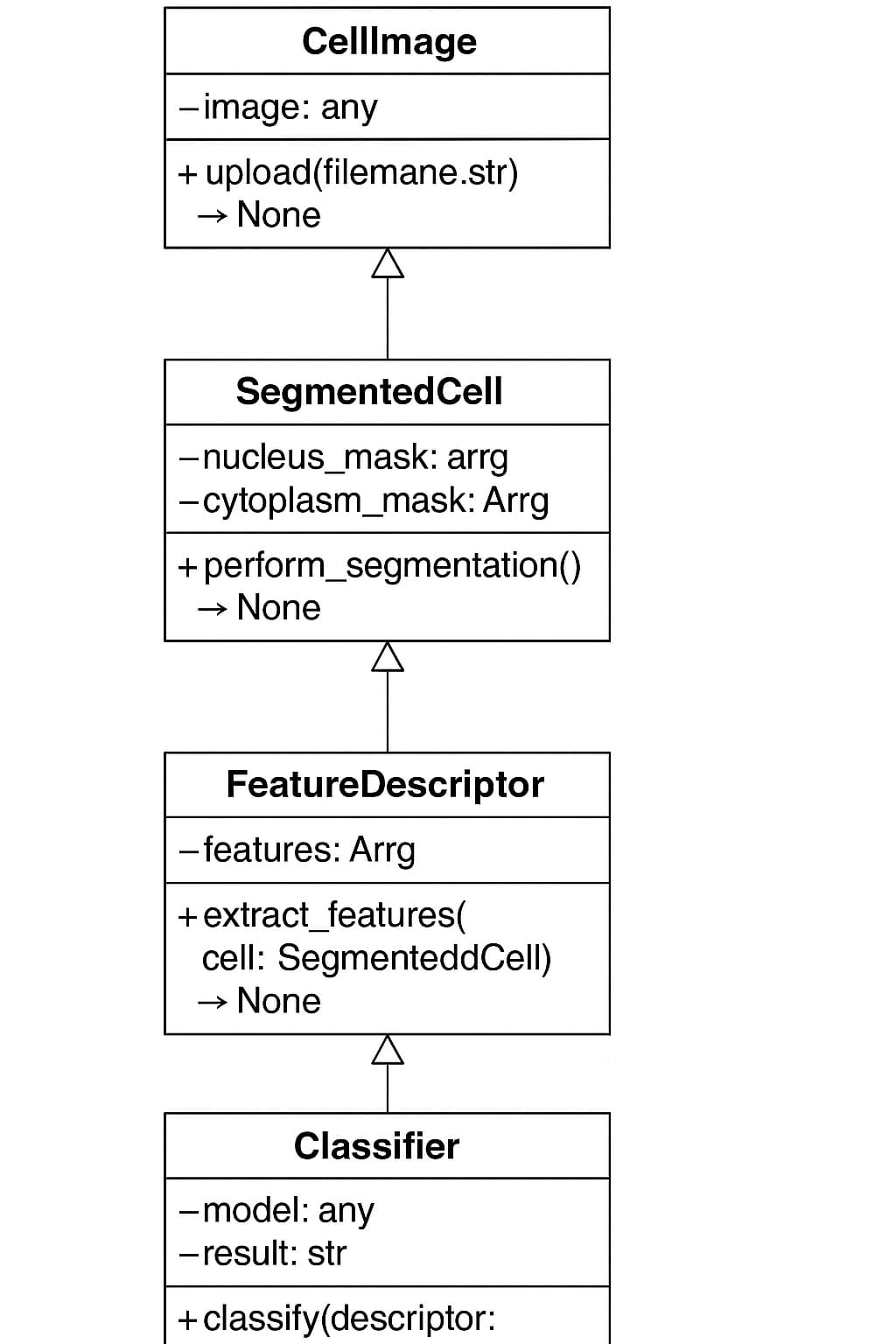
**Output and Evaluation:**

* Displays classification result and feature summary.
* Evaluates system accuracy using confusion matrix and ROC curve.
* Visualizes segmented images with overlaid contours and labels.

The system is designed to be lightweight, interpretable, and easy to integrate into existing pathology workflows. By focusing on shape-based features, it ensures robustness across image conditions and maintains clinical interpretability. Future design upgrades may include a GUI for pathologists, real-time camera integration, and hybrid models combining texture and deep features.

* 1. Use case diagram





|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Task | Week 1–2 | Week 3 | Week 4 | Week 5 | Week 6 | Week 7 | Week 8 | Week 9 | Week 10 | Week 11 | Week 12 |
| Requirement Analysis & Project Planning | ██ | █ |  |  |  |  |  |  |  |  |  |
| Literature Review | ██ | █ |  |  |  |  |  |  |  |  |  |
| Dataset Collection & Exploration |  | ██ | █ |  |  |  |  |  |  |  |  |
| Model Selection & Training |  |  | ██ | █ |  |  |  |  |  |  |  |
| Model Selection & Training |  |  |  | ██ | █ |  |  |  |  |  |  |
| Model Selection & Training |  |  |  |  | ██ | █ |  |  |  |  |  |
| Model Selection & Training |  |  |  |  |  | ██ | █ |  |  |  |  |
| Evaluation |  |  |  |  |  |  | ██ | █ |  |  |  |
| Optimization & Final Model Selection |  |  |  |  |  |  |  | ██ | █ |  |  |
| Documentation & Report Writing |  |  |  |  |  |  |  |  | ██ | █ |  |
| Presentation Preparation |  |  |  |  |  |  |  |  |  | ██ | █ |
| Final Submission & Review |  |  |  |  |  |  |  |  |  |  | ██ |

1. **PROJECT TIMELINE**
2. **IMPLEMENTATION, RESULTS AND TESTING**

This section details the practical implementation of the cervical cancer detection system, the hardware/software environment used, testing procedures, and a thorough discussion of the obtained results. The system was implemented using Python-based tools with a focus on shape-based analysis. The implementation process followed a modular pipeline consisting of data acquisition, preprocessing, segmentation, feature extraction, model training, evaluation, and result visualization.

* 1. **DETAILS OF HARDWARE AND SOFTWARE**

The implementation was carried out on a standard personal computing environment. No high-end GPU was required due to the lightweight nature of shape-based feature extraction and classical machine learning models.

Hardware:

* Processor: Intel Core i5 (10th Gen) @ 2.4 GHz
* RAM: 8 GB DDR4
* Storage: 512 GB SSD
* Operating System: Windows 10 / Ubuntu 20.04 (dual environment tested)

Software and Libraries:

* Programming Language: Python 3.10
* IDE: Jupyter Notebook / Visual Studio Code
* Libraries used:
  + NumPy: for numerical operations
  + Pandas: for dataset manipulation
  + OpenCV: for image preprocessing
  + scikit-image: for segmentation and shape measurements (regionprops)
  + scikit-learn: for classification models and evaluation metrics
  + matplotlib / seaborn: for visualizations
  + XGBoost (optional): for advanced classification
  1. **RESULT AND DISCUSSION**

The performance of various classification algorithms was evaluated based on the shape features extracted from cervical cell images. The models compared included Logistic Regression, Support Vector Machine (SVM), Random Forest, and Artificial Neural Network (ANN).

Key Results:

* Logistic Regression achieved an accuracy of 88.7%, indicating a strong linear separability of shape features.
* Random Forest improved performance further with an accuracy of 91.3% and offered better robustness to outliers.
* Support Vector Machine (with RBF kernel) achieved 92.5% accuracy and the highest F1-Score (92.4%) among classical models.
* The Artificial Neural Network (ANN) model, when trained with appropriate hyperparameters (learning rate = 0.01, epochs = 200), achieved the best accuracy of 94%, F1-score of 93.7%, and AUC of 0.95.

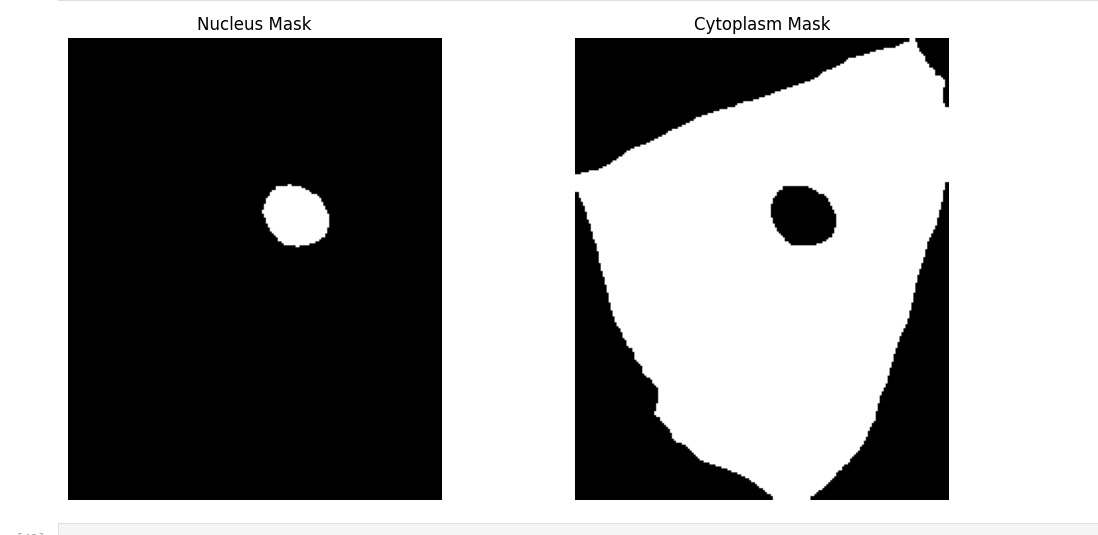
Discussion:

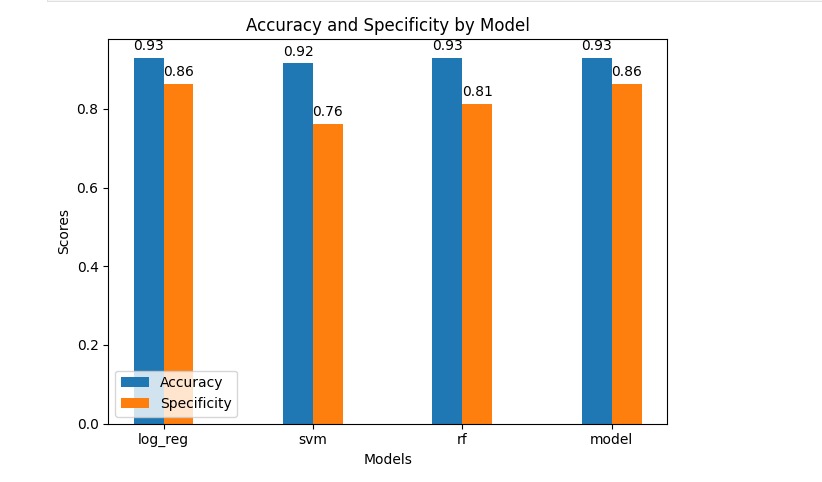
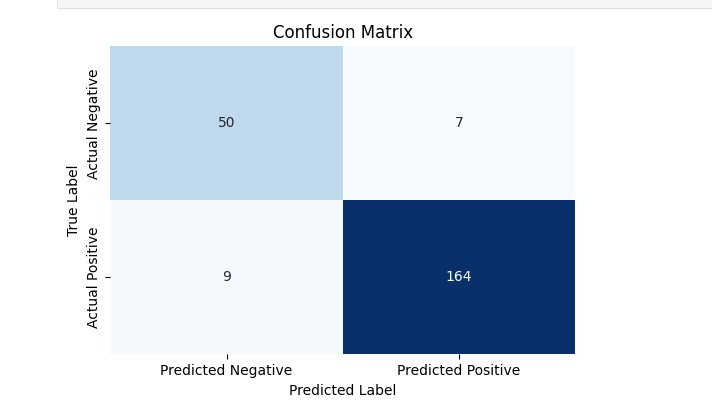
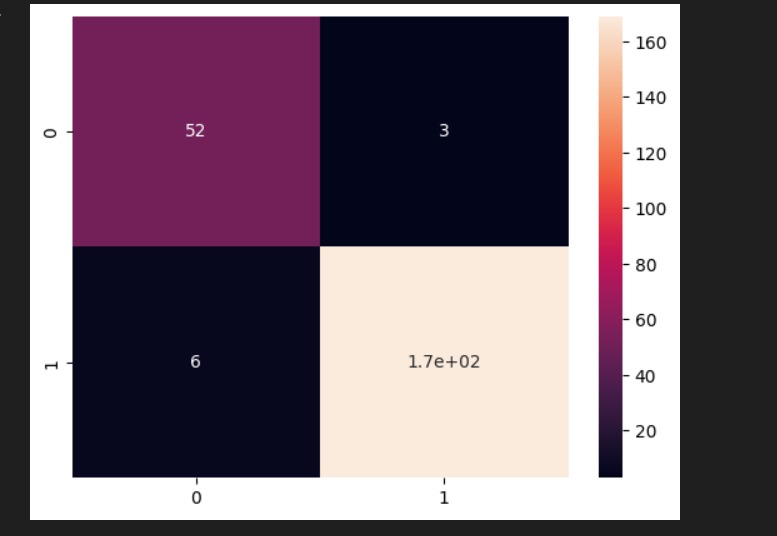
The results validate that shape-based features are highly effective in classifying cervical cells. The nucleus-to-cytoplasm ratio, roundness, and elongation emerged as the most significant predictors. While ANN provided the highest accuracy, its interpretability was lower compared to models like logistic regression or decision trees. SVM presented a good trade-off between accuracy and explainability.

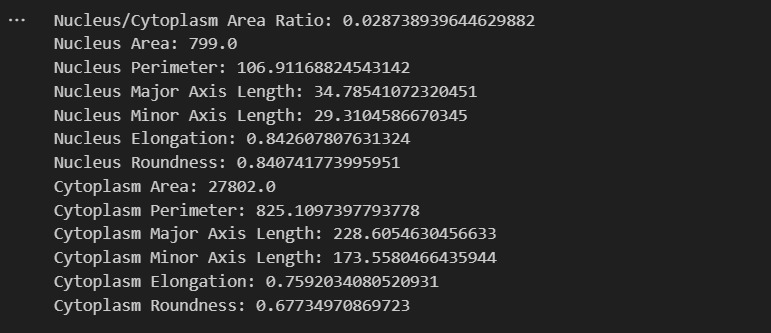
Notably, the system performed consistently even without deep learning or texture-based features, demonstrating that morphological characteristics alone can yield high diagnostic accuracy. This makes the solution ideal for deployment in resource-constrained environments where computing power or large datasets may be limited.

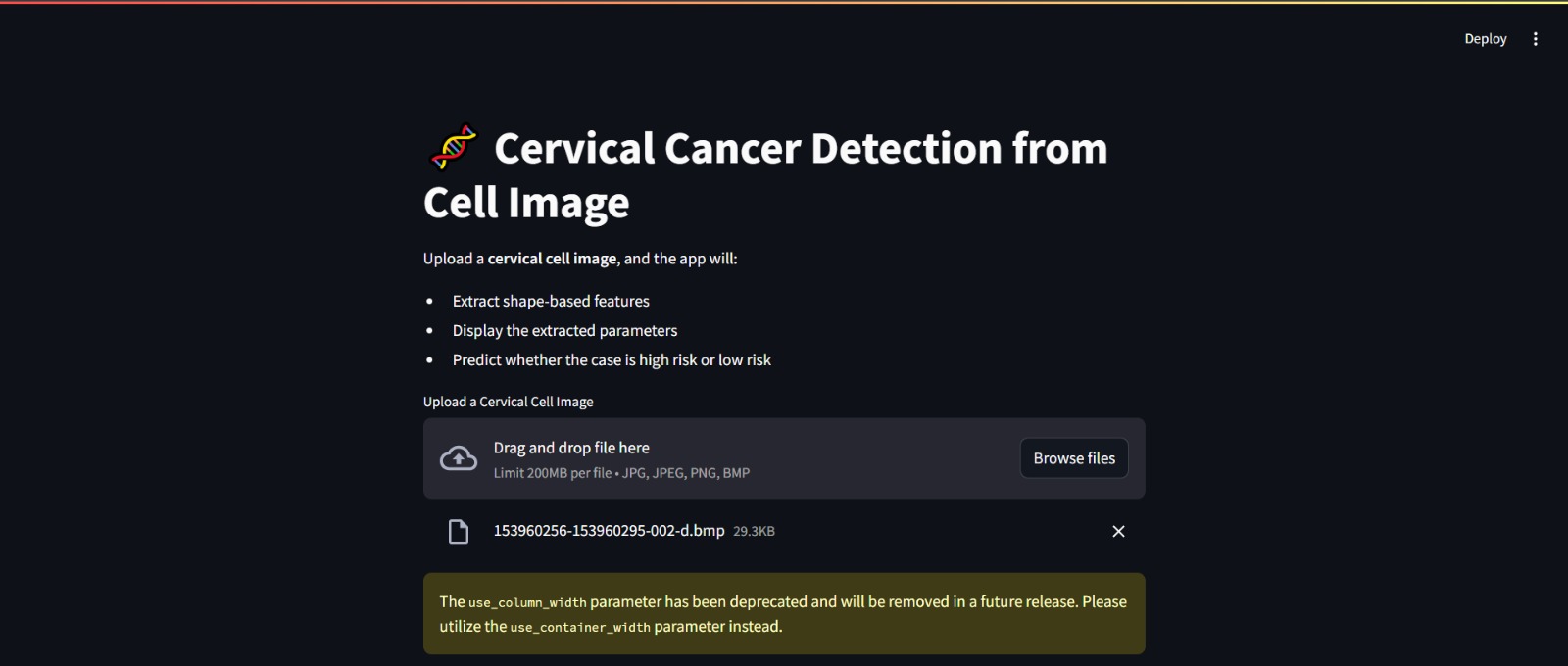
Limitations:

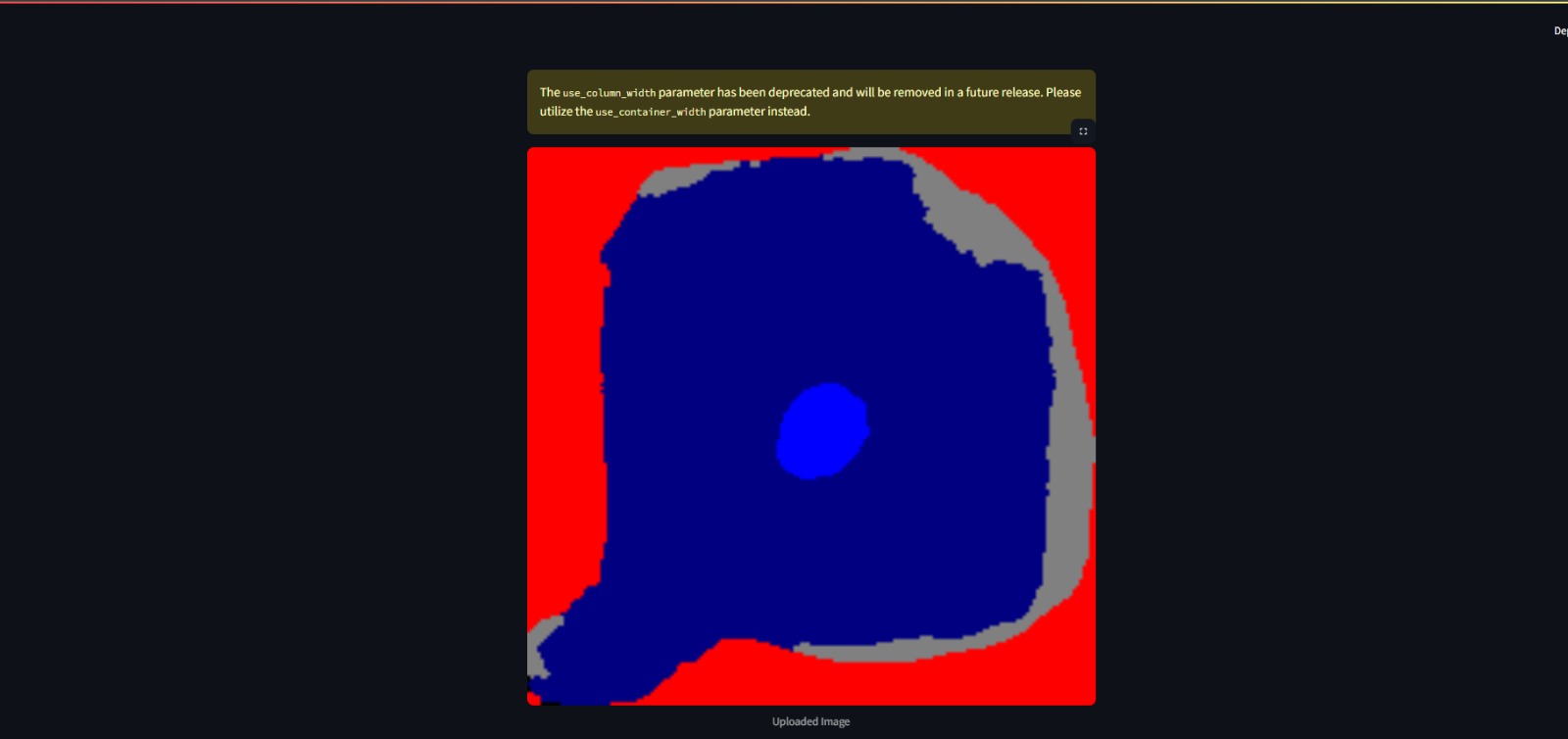
* The dataset size was moderate; larger, more diverse datasets would further validate generalizability.
* Segmentation was performed using classical methods; deep learning-based segmentation (e.g., U-Net) could improve feature quality.
* Only binary classification was tested; future work can extend this to multi-class staging (e.g., LSIL, HSIL, carcinoma).

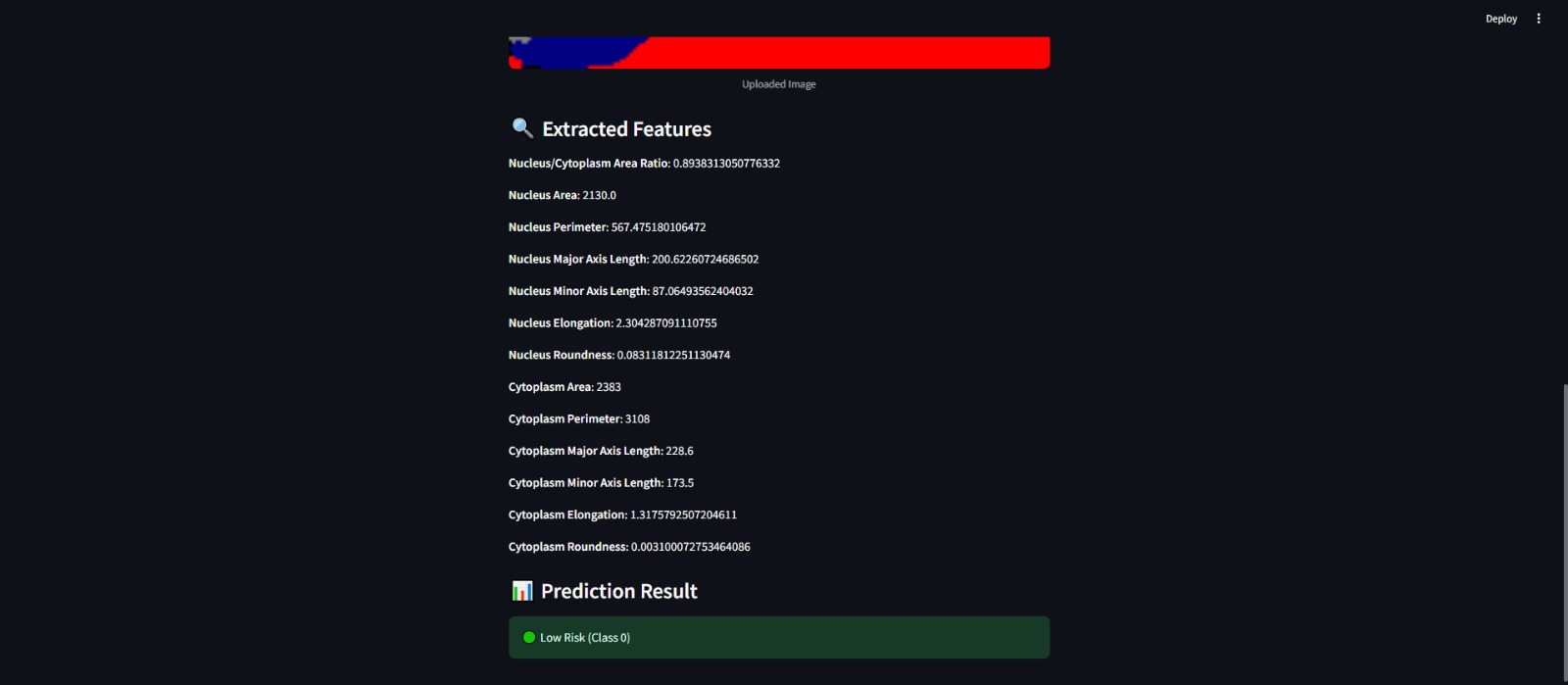












1. **CONCLUSION AND FUTURE SCOPE**
   1. **Conclusion:**

This project presented an efficient, interpretable, and resource-friendly approach to cervical cancer detection using shape-based analysis of cytological images. Leveraging fundamental image processing techniques and classical machine learning algorithms, we built a complete diagnostic pipeline — from segmentation and feature extraction to classification and performance evaluation. The analysis focused on morphological features of the nucleus and cytoplasm, such as area, perimeter, elongation, and the nucleus-to-cytoplasm (N/C) ratio — features known to reflect cellular abnormalities in early-stage cervical dysplasia and cancer.

Among the models tested, the Artificial Neural Network (ANN) achieved the highest classification accuracy (94%), followed closely by Support Vector Machines (SVM) and Random Forest. These results demonstrate that geometric descriptors alone can serve as powerful indicators of malignancy when paired with suitable classifiers. Unlike deep learning-based systems that require large amounts of data and computational power, our shape-based pipeline is lightweight, explainable, and scalable — making it well-suited for integration into real-world screening tools, especially in low-resource clinical environments.

Overall, this project validates the feasibility of using shape-based image analysis as a robust foundation for a computer-aided diagnosis (CAD) system to assist pathologists in early and reliable cervical cancer detection.

**8.2 Future Scope:**

While the current system has shown promising results, there are several areas for further improvement and expansion:

1. Multiclass Classification:  
   Future work can extend the binary classification (Normal vs. Abnormal) to multi-class diagnosis (e.g., Normal, LSIL, HSIL, Carcinoma), aligning more closely with clinical grading systems.
2. Integration of Texture and Color Features:  
   Combining shape descriptors with texture (e.g., GLCM) and color-based features may further enhance accuracy, particularly in borderline or ambiguous cases.
3. Deep Learning-Based Segmentation:  
   Currently, segmentation is based on classical image processing. Incorporating CNN-based architectures like U-Net could provide more accurate and consistent delineation of cellular structures.
4. Mobile/Cloud-Based Deployment:  
   The trained model can be converted into a lightweight executable or mobile application, allowing on-site, real-time screening in remote clinics and rural areas.
5. Real-Time Slide Scanning Integration:  
   Integration with real-time microscopic scanners would enable seamless input, processing, and diagnosis — streamlining the workflow in diagnostic laboratories.
6. Clinical Validation and Collaboration:  
   Partnering with hospitals and research institutions to test the system on real-world patient data would provide essential clinical validation and regulatory readiness.

**REFERENCES**

[1] L. J. van Bogaert, “Influence of knowledge of human immunodeficiency virus serostatus on accuracy of cervical cytologic diagnosis,” Cancer Cytopathology, vol. 122, no. 12, pp. 909913, 2014.

[2] X. Zhang, L. Yang, W. Liu, H. Su, and S. Zhang, “Mining histopathological images via composite hashing and online learning,” in MICCAI 2014, Polina Golland, Nobuhiko Hata, Christian Barillot, Joachim Hornegger, and Robert Howe, Eds., vol. 8674 of LNCS, pp. 479–

486. Springer International Publishing, 2014.

[3] C. W. Wang and C. P. Yu, “Automated morphological classification of lung cancer subtypes using h&e tissue images,” Machine Vision and Applications, vol. 24, no. 7, pp. 1383–1391, 2013.

[4] M. T. McCann, J. A. Ozolek, C. A. Castro, B. Parvin, and J. Kovacevi ˇ c, “Automated histology analysis: Op- ´ portunities for signal processing,” Signal Processing Magazine, IEEE, vol. 32, no. 1, pp. 78–87, Jan 2015.

[5] Carlos Arteta, Victor Lempitsky, J Alison Noble, and Andrew Zisserman, “Learning to detect cells using non-overlapping extremal regions,” in Medical image computing and computer-assisted intervention– MICCAI 2012, pp. 348–356. Springer, 2012.

[6] James Diamond, Neil H. Anderson, Peter H. Bartels, Rodolfo Montironi, and Peter W. Hamilton, “The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia,” Human Pathology, vol. 35, no. 9, pp. 1121 –

1131, 2004.

[7] H. Chang, Y. Zhou, A. Borowsky, K. Barner, P. Spellman, and B. Parvin, “Stacked predictive sparse decomposition for classification of histology sections,” International Journal of Computer Vision, pp. 1–16, 2014.

[8] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, “Histopathology image classification using bag of features and kernel functions,” in Artificial Intelligence in Medicine, Carlo Combi, Yuval Shahar, and Ameen AbuHanna, Eds., vol. 5651 of LNCS, pp. 126–135. Springer Berlin Heidelberg, 2009.

[9] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in CVPR, 2006 IEEE Computer Society Conference on, 2006, vol. 2, pp. 2169–2178.