Faculty of Computer Science and Information Technology

COMPARING SEGMENTATION TECHNIQUES FOR SEGMENTING ENDOCERVIX IN COLPOSCOPY IMAGES

Hannah Shi En Yahya

Bachelor of computer science with honours

(Information Systems)

2024

COMPARING SEGMENTATION TECHNIQUES FOR SEGMENTING ENDOCERVIX IN COLPOSCOPY IMAGES

HANNAH SHI EN YAHYA

This project is submitted in partial fulfillment of the
requirements for the degree of
Bachelor of Computer Science and Information Technology

Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK

2024

MEMBANDINGKAN TEKNIK SEGMENTASI UNTUK MEMBENTUK ENDOCERVIX DALAM IMEJ COLPOSCOPY

HANNAH SHI EN YAHYA

Project ini merupakan salah satu keperluan untuk Ijazah
Sarjana Muda Sains Komputer dan Teknologi Maklumat

Fakulti Sains Komputer dan TeknologiMaklumat
UNIVERSITI MALAYSIA SARAWAK
2024

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE: COMPARING SEGMENTATION TECHNIQUES FOR SEGMENTING ENDOCERVIX IN COLPOSCOPY IMAGES

ACADEMIC SESSION: 2024/2025

HANNAH SHI EN YAHYA
(CAPITAL LETTERS)

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (√)

☐ CONFIDENTIAL (Contains classified information bounded by the OFFICIAL SECRETS ACT 1972)

☐ RESTRICTED (Contains restricted information as dictated by the body or organization where the research was conducted)

☑ UNRESTRICTED

Validated by

(AUTHOR'S SIGNATURE) (SUPERVISOR'S SIGNATURE)

Permanent Address
Lot 805, Lorong Tengah C1,
Siburan, 942000 Kuching, Sarawak.

Date: 10 July 2025 Date: 10 July 2025

Note * Thesis refers to PhD, Master, and Bachelor Degree
** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to Almighty God for granting me the strength, patience, and perseverance to complete this thesis. His guidance has been my anchor throughout this journey. I would like to extend my sincere thanks to Dr. Wang Yin Chai as our Programme Coordinator for his guidance and efforts in supporting all final-year students through the completion of our theses.

I wish to acknowledge the Faculty of Computer Science and Information Technology (FCSIT) as well for providing the necessary resources, knowledge, and a conducive learning environment, which were pivotal in completing this project. I am also immensely grateful to Roboflow for making the colposcopy image datasets available, forming the foundation of this research. This work would not have been possible without these invaluable resources.

Finally, I would like to thank my family and friends for their unwavering support, prayers, and encouragement. Your belief in me and your presence during challenging times have been a source of great strength and comfort. I am truly blessed to have you by my side. To all who have contributed to this milestone in my academic journey, I extend my heartfelt appreciation and gratitude.

ABSTRACT

Cervical cancer is a significant global health concern, and early detection through colposcopy plays a vital role in reducing mortality rates. Accurate segmentation of the endocervix in colposcopy images is essential for diagnosis and treatment. This study compares segmentation techniques to determine the most effective method for segmenting the endocervical region. A publicly available, medium-sized dataset of annotated colposcopy images from Roboflow was used. A standardised preprocessing pipeline was applied to resize images and generate binary segmentation masks. Four segmentation models: K-Means clustering, Random Forest, Support Vector Machine (SVM), and U-Net were implemented and evaluated using Intersection over Union (IoU), Dice Coefficient, and F1 Score. Among the models tested, U-Net achieved the highest performance, demonstrating superior ability to capture spatial and contextual features. In contrast, traditional machine learning models showed moderate results but were limited in handling complex anatomical structures. These findings highlight the advantage of deep learning, particularly the U-Net architecture, for accurate endocervix segmentation. This research supports the application of AI-based methods in cervical cancer screening by providing a comparative evaluation of segmentation approaches in medical imaging.

ABSTRAK

Kanser serviks merupakan isu kesihatan global yang serius, dan pengesanan awal melalui prosedur kolposkopi memainkan peranan penting dalam mengurangkan kadar kematian. Pembahagian tepat endoserviks dalam imej kolposkopi adalah penting untuk diagnosis dan rawatan. Kajian ini membandingkan teknik segmentasi untuk menentukan kaedah yang paling berkesan untuk membahagikan kawasan endoserviks. Satu set data bersaiz sederhana yang mengandungi imej kolposkopi beranotasi telah digunakan daripada platform awam Roboflow. Proses pra-pemprosesan standard telah dilaksanakan untuk menyeragamkan saiz imej dan menjana topeng segmentasi binari. Empat model telah dibangunkan dan dinilai: K-Means Clustering, Random Forest, Support Vector Machine (SVM), dan U-Net, menggunakan metrik Intersection over Union (IoU), Dice Coefficient, dan F1 Score. Daripada semua model yang diuji, U-Net menunjukkan prestasi paling tinggi dan terbukti berkesan dalam menangkap ciri spatial dan konteks imej. Sebaliknya, model pembelajaran mesin tradisional hanya mencatat prestasi sederhana dan kurang berkeupayaan mengendalikan struktur anatomi yang kompleks. Dapatan kajian ini menyerlahkan kelebihan pendekatan pembelajaran mendalam, khususnya seni bina U-Net, dalam segmentasi imej endoserviks yang tepat. Kajian ini menyumbang kepada bidang analisis imej perubatan dengan menyediakan penilaian perbandingan terhadap kaedah segmentasi, serta menyokong penggunaan kaedah berasaskan AI dalam saringan kanser serviks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1 INTRODUCTION

1.1 Introduction

Cervical cancer remains a major public health concern, as it is the fourth ranking among the leading causes of cancer-related mortality in women worldwide. According to the World Health Organization, over 300,000 women die each year from cervical cancer (Torode et al., 2021 as cited in International Agency for Research on Cancer, 2018), with the majority of these cases occurring in low- and middle-income countries. Regular screening through colposcopy, a procedure that visually examines the cervix, plays a crucial role in early detection and prevention. However, the traditional process of interpreting colposcopy images is challenging, as it relies heavily on the skill and experience of medical professionals (Xue et al., 2020), leading to variability and potential inaccuracies in diagnosis. Moreover, manual image analysis is often time-consuming and subject to human error, posing barriers to effective and timely screening.

In recent years, advances in image segmentation—a computer vision technique used to partition an image into meaningful regions—have shown potential to improve the accuracy and efficiency of medical image analysis (Xu et al., 2024). Through segmentation, key regions of interest (ROI), such as the endocervical area, can be automatically identified, allowing for faster and potentially more consistent diagnoses. Despite these advances, there is still no consensus on the best segmentation technique for colposcopy images, as models vary in terms of accuracy, computational requirements, and ease of integration into medical workflows (Jiménez Gaona et al., 2022). This gap highlights a critical need for a systematic evaluation of segmentation models to identify the most suitable approach for cervical cancer detection.

This thesis aims to address this need by conducting a comprehensive comparison of different image segmentation algorithms on a curated dataset of colposcopy images. The project will involve creating a well-documented dataset, segmenting endocervical regions, and evaluating each technique's performance based on accuracy, efficiency, and adaptability. The expected outcome is to recommend an optimal segmentation model that can contribute to future research and practical applications, enhancing diagnostic tools for cervical cancer screening and ultimately improving patient outcomes.

1.2 Problem statement

Cervical cancer is one of the leading causes of cancer-related deaths among women worldwide, especially in low- and middle-income countries where access to routine screening is limited. Early detection is critical in improving treatment outcomes, yet existing diagnostic methods such as manual examination of colposcopy images are time-consuming, subjective, and prone to human error. Image segmentation, a key technique in computer vision, offers potential for automating the detection process by accurately isolating and analyzing endocervical regions in colposcopy images. Image segmentation is the process of giving each pixel in an image a label so that pixels with the same label have similar visual traits. In image processing jobs, it enhances image analysis. For example, in this thesis the aim is to extract the segmented endocervix region in the colposcopy image (Saini & Arora, 2014). However, current segmentation models vary widely in effectiveness and applicability, and there is a lack of consensus on the optimal approach for cervical cancer detection. This project aims to address this gap by systematically comparing various image segmentation techniques to identify the most accurate and practical model for detecting endocervical regions in cervical cancer screening. By establishing a well-documented and annotated dataset, this research will provide a foundation for further advancements in image-based diagnostics, ultimately contributing to improved accessibility and accuracy in cervical cancer detection. Accurately segmenting the endocervix in colposcopy images is a challenging task due to the variability in image quality, lighting conditions, and anatomical differences among patients. Traditional image processing techniques often fail to achieve the required precision, while newer deep learning module may be computationally intensive and complex. Thus, there is a need to systematically compare these techniques to determine which module balances accuracy and computational efficiency for use in clinical settings.

1.3 Objectives

The objectives for this project are:

1. To standardize colposcopy images using image processing techniques by normalizing intensity, resizing to a uniform resolution, and enhancing contrast for improved consistency and analysis.
2. To compare segmentation techniques for segmenting endocervix in colposcopy images.
3. To determine the best segmentation technique for segmenting endocervix in colposcopy images.

1.4 Methodology

Figure 1.1 below shows the methodology use for this project which is the Cross Industry Standard Process for Data Mining (CRISP-DM) model.

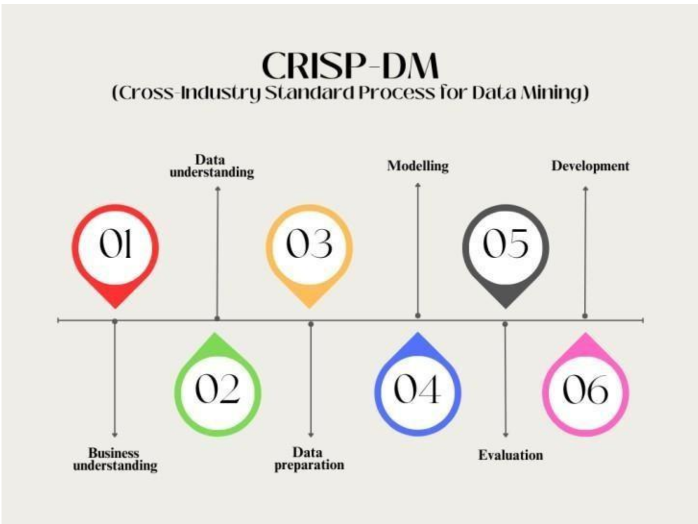


Figure 1.1 CRISP-DM methodology

In this project, CRISP-DM is chosen as the methodology due to its structured, iterative approach that aligns well with the demands of medical image segmentation. CRISP-DM's phases—particularly data understanding, preparation, and modelling, support the systematic experimentation and evaluation of different segmentation techniques, ensuring a comprehensive analysis that informs recommendations for future research in cervical cancer detection. This methodology enables a thorough, adaptable workflow, making it ideal for achieving the project's objective of identifying the best segmentation approach.

1.4.1 Business Understanding

The first phase focuses on understanding the project's objectives and requirements from a business or research perspective. It involves defining the problem, setting clear goals, and determining project constraints, which is essential for aligning technical work with desired outcomes. For this project, the goal is to compare different segmenting algorithm and select the best preforming model that can segment and extract the endocervix region in a colposcopy images.

1.4.2 Data Understanding

In this phase, any relevant data or information related to the project will be collected. With the collection of data found, exploration onto the data will be preform to identify its quality and characteristics. This involves assessing data relevance, completeness, and accuracy to ensure it is suitable for the project goals.

1.4.3 Data Preparation

After choosing the suitable dataset, the next phase is to preform analysis. This including cleaning, transforming, and organizing the data to suit the modelling phase. Since image segmentation is involve in this project, preprocessing steps like resizing, annotation, or normalization need to be done onto the gathered data set, which is the colposcopy images, to ensure its accurate, consistent and usable.

1.4.4 Modeling

In this phase, selection and application of various modelling techniques are applied onto the prepared data. For this project, the modelling phase will include experimenting with five different segmentation algorithms and tuning parameters to improve performance.

1.4.5 Evaluation

After building the models and experimenting, all the models will undergo evaluation as well as comparison of effectiveness. Additionally, the models will be evaluated based on how effectively it can address the project objectives goals. This include addressing the model accuracy, consistency, and relevance to determine the best segmentation method for cervical cancer detection.

1.4.6 Deployment

The last phase involves implementing the final model for practical use. In research-oriented projects, it may include documenting the findings, preparing recommendations, or setting up the model for further study.

1.5 Scope

The scope of this project focuses on using cervical images found from Roboflow to compare various segmentation algorithms for accurately segmenting the endocervix. The scope includes preprocessing the Roboflow VIA dataset, applying different segmentation techniques, and evaluating their performance based on F1 Score and other key metrics. The goal is to identify the best-performing prediction model for this specific dataset. The study is limited to the analysis of Roboflow images and will not explore other data sources or non-segmentation methods.

1.6 Significance of project

This research is important for improving the accuracy and efficiency of cervical cancer diagnosis. By comparing different segmentation techniques for the endocervix in colposcopy images, the project aims to identify the most effective segmentation model. Accurate segmentation can enhance early detection of cervical abnormalities, leading to better patient outcomes. Additionally, the findings may support the development of automated tools for use in clinical practice, also known as Computer-aided diagnosis (CAD). This will ultimately improve the diagnostic processes and making healthcare more accessible, especially in low-resource settings.

1.7 Expected outcome

Comprehensive analysis of different segmentation techniques for cervical cancer detection, providing insights into the strengths and limitations of each approach. A well-documented dataset of colposcopy images will be developed, including segmented endocervix regions that are carefully annotated and pre-processed for reliable experimentation. Based on the comparative analysis, recommendations will be made on the most effective segmentation module, aiming to support both future research and practical applications in cervical cancer detection.

1.8 Project Schedule

Figure 1.2 shows the full gantt chart planned for Final Year Project.

Figure 1.2 Gantt Chart for Final Year Project

1.9 Thesis outline

1.9.1 Introduction

Chapter 1 introduces the research topic, introduction, problem statement, objectives, methodology, scope, significance of project, expected outcome, project outcome, project schedule and thesis outline.

1.9.2 Literature review

Chapter 2 reviews existing research and theories relevant to the topic. It discusses key concepts, previous findings, and identifies gaps in the literature that the current research aims to address.

1.9.3 Methodology

Chapter 3 showcases and explains the research design, data collection methods, and analysis techniques. It explains how the project is conducted, including tools and processes used to achieve the research objectives.

1.9.4 Implementation and Testing

In chapter 4, this section presents the experimental procedures, setup, and implementation. It presents how the methodology is applied in practice and any adjustments made during the experiment.

1.9.5 Conclusion and future work

The final chapter summarizes the key findings, discusses their implications, and suggests possible future research. It concludes the thesis by reflecting on the research objectives and offering recommendations.

1.10 Summary

Chapter 1 provides an overview of the study, including the introduction, problem statement, and objectives. It also briefly discusses the chosen methodology, the CRISP-DM model, which will guide the project's development process. Additionally, this chapter outlines the scope, significance, and expected outcomes of the project, highlighting its aims and potential contributions. A project schedule is presented to serve as a roadmap for the project's implementation. Finally, the chapter concludes with a thesis outline, summarizing the content and focus of each subsequent chapter.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter covers the literature review of cervical cancer, examines how recent technological developments, particularly in machine learning and deep learning are being applied to medical image segmentation to address the challenges faced in segmenting colpocopsy images. Various segmentation algorithms, including K-means, Random Forest, U-net and more are reviewed for their effectiveness in segmenting anatomical structures in medical imagery. Each technique is analyzed for its strengths, limitations, and applicability to endocervical segmentation. The chapter also covers the evaluation metrics used, such as Dice Coefficient, Intersection over Union (IoU), F1 Score, Pixel Accuracy and more that are commonly used to measure segmentation accuracy. Additionally, the chapter also highlights the gaps in existing research, justifying the need for a comparative study of segmentation techniques tailored specifically to colposcopy images. This review lays the groundwork for selecting and implementing effective segmentation models that can enhance diagnostic capabilities in cervical cancer detection by automating and refining the segmentation process. Through this examination, the chapter emphasizes how technological advancements in machine learning and deep learning have the potential to transform cervical cancer screening by increasing segmentation accuracy, reducing diagnostic variability, and ultimately supporting clinicians in making timely and accurate diagnoses.

2.2 Domain background

Cervical cancer remains a significant global health issue and is one of the leading causes of cancer-related mortality among women, particularly in low- and middle-income countries. Based on recent study, an estimated 604 127 new cases of cervical cancer and 341 831 deaths has occurred in 2020 based on the Global Cancer Observatory (GLOBOCAN) 2020 repository (Singh et al., 2022). The disease often progresses silently in its early stages, making early detection crucial for effective intervention and improved survival rates. Cervical cancer is largely preventable through regular screening and early intervention, which can lead to the detection and treatment of precancerous lesions before they progress to invasive cancer.

Screening methods such as the Pap smear, HPV testing, and colposcopy play essential roles in early cervical cancer detection (Hou et al., 2022). Colposcopy, in particular, is a diagnostic procedure used to closely examine the cervix, vagina, and vulva for signs of disease (Burness et al., 2020). During a colposcopy, clinicians can identify abnormal tissues and, if necessary, take biopsies for further analysis. Among these tissues, the endocervical region located within the cervical canal and adjacent to the transformation zone where most cervical cancers originate, is extremely important. Precise and accurate analysis of this area can reveal early signs of abnormal cell growth or lesions. However, effective colposcopy diagnosis requires identifying specific regions of the cervix accurately, a process that traditionally depends on the skill and experience of the clinician. Manual segmentation of the endocervical area, necessary for closer examination and analysis, is labor-intensive and subject to inter-observer variability, potentially leading to inconsistencies in diagnosis (Hou et al., 2022). In addition, colposcopy image analysis presents several other challenges due to the variability in cervix anatomy, lighting conditions, and the presence of artifacts like mucus or blood that can obscure areas of interest (Desai et al., 2021). These challenges often complicate manual segmentation efforts, which can impact diagnostic accuracy and efficiency, especially in high-throughput clinical settings, where a large number of medical images need to be processed quickly, such as busy hospitals or large clinics. Thus, underscores the need for standardized and automated segmentation solutions that can help streamline cervical cancer diagnosis.

Integrating Machine learning or deep learning techniques in medical field offers a promising solution to these challenges by enabling automated segmentation of the endocervical region in colposcopy images. First and foremost, automated segmenting model enhance diagnostic accuracy (Yang et al., 2022). By accurately and consistently segmenting the endocervical area, machine learning models can assist clinicians in targeting regions of interest, reducing the risk of missed diagnoses and supporting early intervention. Next, CAD models also helps reduces the frequency of human error and Inter-observer variability (Tian et al., 2022). Automated methods minimize the subjective variability inherent in manual segmentation, allowing for more standardized assessments across patients and clinicians. With the integration of CAD models, clinical efficiency will also be enhanced. This is because with automated segmentation, the time required to prepare colposcopy images for diagnosis is significantly reduced, allowing clinicians to focus on interpreting results and making treatment decisions. Lastly, automated segmentation models also offer scalability in resource-constrained settings (Nazir et al., 2023). This means, in areas with limited access to expert clinicians, automated segmentation models can serve as valuable tools for healthcare providers, enabling more effective screening and diagnosis even in remote or underserved locations.

Automated segmentation in colposcopy images not only improves the efficiency and accuracy of cervical cancer diagnosis but also supports the broader public health goal of cervical cancer prevention. By facilitating earlier and more reliable identification of precancerous lesions, automated segmentation has the potential to enhance screening programs, increase patient access to early treatment options, and ultimately reduce cervical cancer incidence and mortality.

2.3 Related work

The use of machine learning and deep learning algorithms for medical image segmentation has gained significant attention in recent years, especially for cervical cancer detection. Accurate segmentation of cervical regions in colposcopy images plays a crucial role in early diagnosis and treatment planning. Numerous studies have employed both learning techniques, such as K-Means, Random Forest, U-net and more for classifying and segmenting medical images. Evaluation metrics like Intersection over Union (IoU), Dice coefficient, F1 score, pixel accuracy and more are widely used to assess the effectiveness of segmentation and classification models. By comparing methods like K-means clustering and other segmenting algorithms, this work aims to identify the most suitable approach for colposcopy image segmentation tailored to cervical cancer detection. The Table 2.1, showcases the comparison of related work from recent year covering year 2020 till 2024:

Table 2.1 Comparison of related work

References	Algorithm	Description	Evaluation metrics	Result	Conclusion
(Wijaya et al., 2021)	K-Means clustering	Obtain from Herlev University Hospital. 70 training images and 28 testing images in total.	• Accuracy	• Accuracy: 75.76% (grayscale with low pass filter)	Proposed method achieved segmentation accuracy highest with grayscale and low pass filter.
(Riana et al., 2022)	K-means clustering	Obtain from RepoMedUNM with 6,168 Pap smear images. It includes ThinPrep and Non-ThinPrep images which are two	• Accuracy	• Average accuracy: 93.33% (ThinPrep) 90% (Non-ThinPrep)	Proposed method identifies Pap smear images effectively. Achieved average accuracy of 92% across all classes.

		different preparation of Pap smear samples.		• Overall average accuracy: 92%	
(Widodo et al., 2021)	K-Means clustering	Dataset from 401 gynecological patients examined. 205 patients detected with cervical cancer.	• Accuracy • AUC	• Euclidean method accuracy: 79.30%, • ROC: 79.17%. • Manhattan method accuracy: 67.83%, • ROC: 65.94%.	Euclidean method is the best method for K-Means Clustering on cervical cancer datasets.
(Shinohara et al., 2023)	U-Net	30 actual colposcopic images of acetowhite epithelium. Images	• Accuracy • Precision • F1 scores	• Accuracy, precision, and F1 scores	Proposed method improves lesion segmentation accuracy

		taken before and after acetic acid application (dual images). Dual images refer to the colposcopic images taken before and after the application of acetic acid solution.		improved with both image types.	using dual images. Images before and after acetic acid enhance deep learning performance.
(Jiménez et al., 2022)	U-net	Obtain from Intel Mobile ODT Cervical Cancer Screening public dataset. Private dataset from a rural community in Ecuador.	• Dice Coefficient • Precision • Accuracy • IoU	• Dice coefficient: 50% • Precision: 65% • Accuracy: 80%	The CAD system needs improvement but could be acceptable in environments with limited access to

					clinicians for cervical cancer diagnosis. Better performance is possible with larger datasets and exploration of other deep learning methods.
(Arum et al., 2021)	U-Net	444 IVA images from an Indonesian hospital in Palembang. Cervical images labeled by experts as ground truth.	• Pixel Accuracy • Mean IoU • Mean Accuracy • Dice Coefficient • Precision • Sensitivity	• Pixel Accuracy: 90.86% • Mean IoU: 56.5% • Mean Accuracy: 75.69%	Performance values indicate potential for improvement in detection. Encourages further research in automatic cervical pre-cancer detection.

				• Dice coefficient: 34.09% • Precision: 41.24% • Sensitivity: 56.91%	
(Li et al., 2022)	Random forest (RF)	Retrospective database of 56 postoperative BCa patients. Includes tumor and wall regions for training.	• Dice Coefficient • ASSD	• Mean Dice coefficient: 0.906 (95% CI: 0.852-0.959) • Mean ASSD: 1.190 mm	Proposed method achieves accurate BCa lesion segmentation.

				(95% CI: 1.727-2.449)	
(Hartmann et al., 2021)	Random forest (RF)	PhC-C2DH-U373 dataset and Retinal imaging dataset.	• Dice Coefficient • IoU • Accuracy	• Deep convolutional neural network achieved best results. One random forest approach showed similar high performance.	Random forests are viable alternatives to deep neural networks based on the comparison of both models. High performance achieved without using GPUs.

(Arora et al., 2021)	Support Vector Machines (SVM)	Herlev pap-smear image dataset.	• Dice coefficient	• Segmentation phase achieved 92% match with manual annotations. • Highest classification accuracy reached 95% using polynomial SVMs.	Machine Learning improves cervical cancer detection accuracy. Polynomial SVMs achieved 95% classification accuracy.

(Ahishakiye & Kanobe, 2024)	Support Vector Machines (SVM)	Pap smear image dataset	• Accuracy • Precision • Recall • F1 Score	• Optimized SVM model achieved 100% accuracy. • DGP model achieved 99.48% accuracy.	Proposed models effectively detect and classify cervical cancer. Suitable for deployment in mobile application-based tool.

In summary, based on the previous studies in cervical cancer (refer Table 2.1) have primarily focused on deep learning models for image segmentation, achieving promising results. However, many of these approaches require large datasets and substantial computational resources, making them less feasible in resource-constrained settings. In contrast, machine learning techniques, such as K-means clustering, offer more scalable and accessible alternatives for segmenting cervical images, particularly in environments with limited access to expert clinicians. This work aims to build on existing methods by exploring and comparing the effectiveness of machine learning and deep learning-based segmentation models in improving diagnostic accuracy, while maintaining efficiency and scalability.

2.4 Machine learning algorithm

A machine learning algorithm is a computational method used to analyze data and learn patterns or make predictions without being explicitly programmed for each specific task (Bi et al., 2019). It involves training a model on a dataset, allowing the model to identify underlying structures or relationships within the data. This learning process helps the algorithm generalize from the examples it's given, enabling it to make accurate predictions or perform complex tasks on new, unseen data.

Machine learning algorithms can be classified into several categories, such as supervised, unsupervised, and reinforcement learning, depending on the nature of the data and the goal of the task (Dasgupta & Nath, 2016). Traditional manual segmentation of medical images, such as colposcopy images, is often time-consuming and prone to human error. Leveraging machine learning and deep learning techniques offers a promising solution by automating these processes, enabling precise and efficient segmentation of the endocervical area. In medical imaging and image segmentation, CAD algorithms are commonly used to analyze, classify, and segment specific structures in images. This automation is especially crucial in clinical settings, where timely and accurate analysis is critical.

2.4.1 K-means clustering

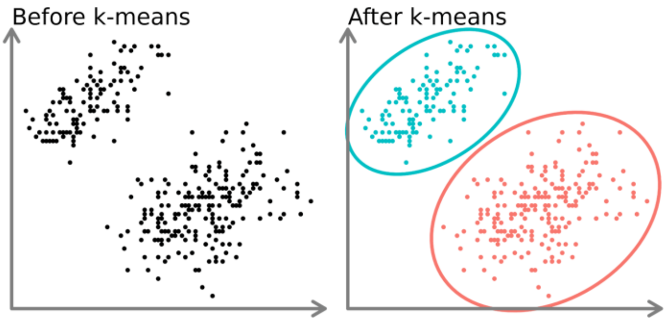


Figure 2.1 K-means clustering (Babitz, 2018)

K-means is an unsupervised learning algorithm used for clustering data into groups, where each group is called a cluster (Vankayalapati et al., 2021). The goal of K-means is to divide the data into K clusters, where K is a number chosen beforehand. The algorithm works by repeatedly assigning data points to the nearest cluster and then adjusting the cluster centers (called centroids) until they stabilize. The process starts by randomly selecting K initial centroids. Then, each data point is assigned to the nearest centroid, based on how close it is. After all points are assigned, the centroids are updated by calculating the average of all points in each cluster. This process of assigning points to centroids and updating the centroids repeats until the centroids no longer change significantly (refer Figure 2.1).

In image segmentation, K-means can group pixels of similar color or intensity into different regions, helping to identify structures or areas of interest. Although K-means is fast and simple, it has some limitations. The number of clusters K must be chosen in advance, and the algorithm assumes that the clusters are roughly spherical and evenly sized, which may not always match the true data structure. In summary, K-means is a useful and efficient algorithm for segmenting images, but it works best when the clusters are well-defined and of similar size.

2.4.2 Random Forest (RF)

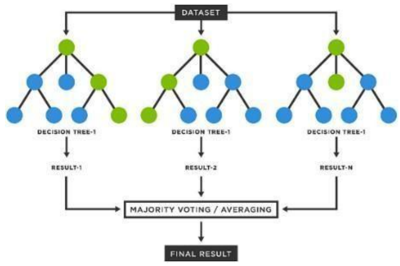


Figure 2.2 Random forest (Gunay, 2023)

Random Forest (RF) is an ensemble learning method used for classification and regression tasks (Sarica et al., 2017). It works by creating multiple decision trees during training and combining their predictions to make a final decision (refer Figure 2.2). Each decision tree is trained on a random subset of the data, and at each node, a random subset of features is considered for splitting. This randomness helps make the model more robust and less prone to overfitting, which can happen in a single decision tree. During prediction, each tree in the forest makes its own prediction, and the final prediction is determined by taking a majority vote (for classification) or averaging the predictions (for regression). This approach helps improve accuracy and generalization, as it reduces the impact of noise or outliers that might affect a single tree.

In image segmentation or classification tasks, Random Forest can be used to classify each pixel based on its features (like color or texture), making it useful for segmenting different regions in an image. One of the key strengths of Random Forest is its ability to handle large datasets and complex relationships between features, while also being less sensitive to overfitting compared to individual decision trees. However, Random Forest can be computationally expensive, especially with many trees, and it may not provide as interpretable results as a single decision tree.

2.4.3 Support Vector Machine (SVM)

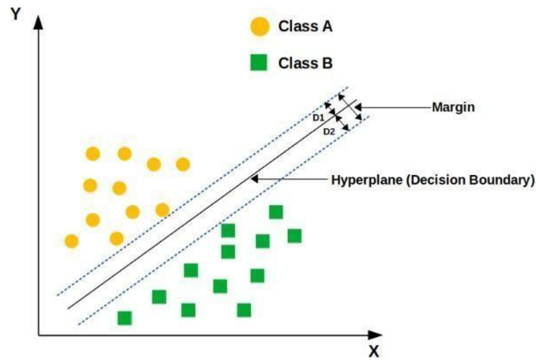


Figure 2.3 Support Vector Machine (Mehta, 2023)

The Support Vector Machine (SVM) is a supervised machine learning algorithm widely used for classification and regression tasks (Mahesh, 2018). It works by finding the optimal hyperplane that best separates data points of different classes in a high-dimensional space (refer Figure 2.3). The hyperplane is chosen to maximize the margin, which is the distance between the closest data points (called support vectors) from each class to the hyperplane. This focus on maximizing the margin makes SVM robust and effective, especially in cases where the data is not linearly separable. For such cases, SVM uses kernel functions like linear, polynomial and radial basis function to transform the data into a higher-dimensional space where it becomes separable. SVM is particularly powerful for small to medium-sized datasets and excels in handling high-dimensional data. However, it can be computationally intensive and less effective for very large datasets or noisy data. Overall, SVM is a reliable and versatile algorithm for classification problems like text categorization, image segmentation, and bioinformatics applications.

2.4.4 U-net

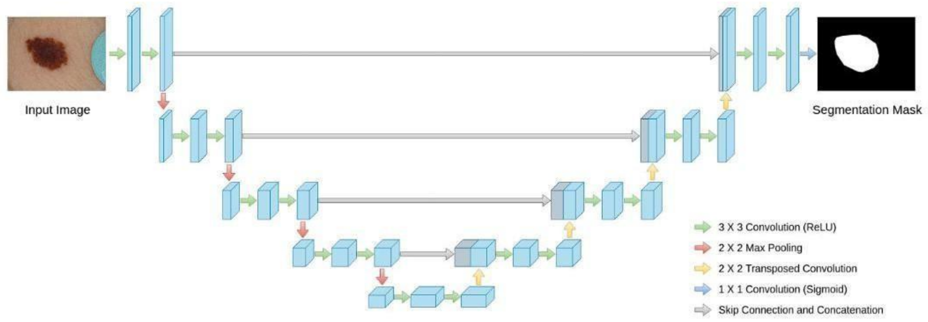


Figure 2.4 U-net (Klingler, 2024)

U-Net is a deep learning architecture specifically designed for image segmentation tasks, particularly in medical imaging (Siddique, 2021). It consists of two main parts: the contracting path (encoder) and the expansive path (decoder), which give it its U-shaped structure (refer Figure 2.4). The contracting path progressively down samples the input image, extracting high-level features, while the expansive path up samples the feature maps, restoring the original image size. A key feature of U-Net is the use of skip connections, which link corresponding layers in the contracting and expansive paths, allowing the model to preserve spatial details and improve segmentation accuracy. The final output is a pixel-wise classification, assigning labels (for example like tumor, tissue or background) to each pixel. U-Net is particularly effective in medical image segmentation, as it can work with relatively small datasets and produce high-quality results. In the context of cervical cancer detection, U-Net can be used to segment images from colposcopy or histopathology, helping to identify regions of interest such as tumors or abnormal tissues, which aids in early diagnosis and treatment.

2.5 Evaluation metrics

Evaluation metrics are essential for quantifying the performance of segmentation algorithms by comparing their output with the ground truth (Muller et al., 2022). It provides numerical values that objectively measure accuracy, overlap, and error, enabling the comparison of different methods in this study. These metrics not only help identify the strengths and weaknesses of each algorithm but also ensure that the chosen method generalizes well to unseen data, building confidence in the results and aligning with the project's goals. The following is a list of common evaluation metrics used to assess the performance of image segmentation algorithms in the medical field.

2.5.1 Accuracy

Accuracy is a measure of how often the model's predictions are correct. It is calculated as the ratio of correct predictions to the total number of predictions. In classification tasks, it represents the overall correctness of the model. The formula to find accuracy as shown in Equation 2.1:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Samples}}$$

Equation 2.1 Accuracy formula

where True Positives (TP) are instances where the model correctly predicts the positive class whereas True Negatives (TN) are instances where the model correctly predicts the negative class. The total number of Samples is the total number of instances or observations in the dataset, which includes all true positives, true negatives, false positives, and false negatives. The advantage of selecting accuracy as an evaluation metrics is that it is simple to understand

and interpret but can be misleading, especially in imbalanced datasets, as it may not reflect the model's performance on the minority class.

2.5.2 Precision

Precision (also known as positive predictive value) measures how many of the predicted positive instances are actually positive. It is the ratio of true positive predictions to the total predicted positives (true positives + false positives). The formula to find precision as shown in Equation 2.2:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Equation 2.2 Precision formula

where True Positives (TP) are instances where the model correctly predicts the positive class whereas False Positives (FP) is the number of instances where the model incorrectly (false) predicted the positive class. The perk of using precision in evaluating model performance is its useful in scenarios where false positives are costly or undesirable like diagnosing a disease where false positives could lead to unnecessary treatments. The downside of the metric is it does not account for false negatives, so it may be less useful in cases where the model needs to identify all positive cases.

2.5.3 Sensitivity (Recall or True Positive Rate)

Sensitivity, also known as recall or the true positive rate, measures how well the model identifies actual positive cases. It is the ratio of true positives to the total number of actual positives (true positives + false negatives). The formula to calculate Sensitivity of a model as shown in Equation 2.3:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Equation 2.3 Sensitivity formula

The Sensitivity evaluation metric is useful when identifying as many positive cases as possible is crucial, such as in medical diagnoses like cancer detection. However, high sensitivity can also result in an increase in false positives, making it less reliable in cases where false positives are costly.

2.5.4 Area Under the Curve (AUC)

Area Under the Curve (AUC) refers to the area under the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various classification thresholds. AUC provides an aggregate measure of performance across all possible classification thresholds. AUC ranges from 0 to 1, where 1 indicates perfect classification and 0.5 indicates a model that performs no better than random guessing.

The advantage of using AUC is its threshold-independent and provides a comprehensive measure of the model's ability to discriminate between positive and negative classes. Although AUC is a robust metric, it can be difficult to interpret in the context of specific thresholds, and may not reflect performance in highly imbalanced datasets if combined with other metrics.

2.5.5 Average Symmetric Surface Distance (ASSD)

Average Symmetric Surface Distance (ASSD) is a metric used in image segmentation tasks, particularly in medical imaging. It measures the average distance between the predicted segmentation surface and the true segmentation surface. In this context, the "surface" refers to the boundary of the segmented object in the image. ASSD is symmetric, meaning it calculates the distance from both the predicted surface to the true surface and the true surface to the predicted surface, and averages these distances. The formula to calculate ASSD as shown in Equation 2.4:

$$\mathrm{ASSD} = \frac{1}{N}\sum_{i=1}^{N}(\text{Distance from predicted to true surface} + \text{Distance from true to predicted surface})$$

Equation 2.4 Average Symmetric Surface Distance formula

The advantage of using ASSD is it provides a detailed evaluation of segmentation accuracy, particularly important in medical imaging where precise boundaries are essential. For instance, tumor boundaries. The downside of ASSD is it can be sensitive to small errors in the boundary of the segmentation, which might not always be significant for the clinical application.

2.5.6 Dice Coefficient or Dice Similarity Coefficient (DSC)

The Dice Coefficient is a widely used metric for evaluating the overlap between the predicted segmentation and the ground truth regions. It measures the similarity between two sets and ranges from 0 to 1, where 1 indicates perfect overlap. For dice coefficient, the higher values indicate better segmentation performance, emphasizing the overlap between the predicted and true regions. The formula for dice coefficient as shown in Equation 2.5:

$$\text{Dice Coefficient} = \frac{2 \times |P \cap G|}{|P| + |G|}$$

Equation 2.5 Dice coefficient formula

Where P represents the pixels in the predicted region, G represents the pixels in the ground truth region, whereas $|P \cap G|$ is the Intersection of predicted and ground truth pixels.

2.5.7 Intersection over Union (IoU)

Intersection over Union (IoU), also known as the Jaccard Index, calculates the ratio of the intersection to the union of the predicted and ground truth regions. The formula for Intersection over Union (IoU) as shown in Equation 2.6:

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|}$$

Equation 2.6 Intersection over union formula

Where $|P \cap G|$ is the intersection of predicted and ground truth pixels and $|P \cup G|$ is the union of predicted and ground truth pixels. The IoU values range from 0 to 1. The higher the IoU values, it indicates that the predicted region closely matches the ground truth.

2.5.8 F1 Score

The F1 Score is the harmonic mean of Precision and Recall, providing a balance between false positives and false negatives. For segmentation tasks, it relates to how accurately the algorithm identifies true positive pixels. The F1 Score ranges from 0 to 1, with higher values reflecting better segmentation. The formula for F1 Score as shown in Equation 2.7:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Equation 2.7 F1 score formula

Where precision is the division of true positives divided by the addition of true positives and false positives. Both the true positives (TP) and false positives (FP) values can be extracted from the confusion matrix table which are a table used to evaluate the performance of a classification model by comparing the actual (true) labels with the predicted labels from the model. The formula to find the precision of a model performance as shown in Equation 2.8:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

Equation 2.8 Precision formula

True Positive (TP) is the number of instances where the model correctly predicted the positive class. For instance, the model predicted a disease (positive), and the patient indeed has the disease (true). The False Positive (FP) is the number of instances where the model incorrectly (false) predicted the positive class. Example the model predicted a disease (positive), but the patient does not have the disease (also called a "Type I error"). Next, the recall formula is the division of true positives divided by the addition of true positives and false negatives. Equation 2.9 shows the formula to find the recall of a model:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

Equation 2.9 Recall formula

False Negative (FN) is the number of instances where the model incorrectly predicted the negative class such as the model predicted no disease (negative), but the patient actually has the disease (also called a "Type II error").

2.5.9 Pixel Accuracy

Pixel Accuracy measures the proportion of correctly classified pixels in the entire image. A higher Pixel Accuracy indicates that the algorithm successfully classifies a larger percentage of the pixels in the image. However, it might not reflect segmentation quality well if the dataset is imbalanced. For example, a small region of interest compared to the background. The formula to calculate pixel accuracy as shown in Equation 2.10:

$$\text{Pixel Accuracy} = \frac{\text{Number of Correctly Classified Pixels}}{\text{Total Number of Pixels}}$$

Equation 2.10 Pixel accuracy formula

In summary, this project will adopts six widely used evaluation metrics to assess the performance of segmentation models: Intersection over Union (IoU), Dice Coefficient, F1 Score, Accuracy, Precision, and Sensitivity (Recall). IoU and Dice Coefficient are crucial for measuring the spatial overlap between the predicted masks and the ground truth, making them highly suitable for medical image segmentation tasks. The F1 Score combines Precision and Recall into a single value, offering a balanced perspective when both false positives and false negatives are significant. Accuracy provides an overall measure of correctly classified pixels, while Precision focuses on the model's ability to minimise false positives. Sensitivity (Recall) reflects how effectively the model detects all relevant positive regions, which is crucial in medical imaging, where missed detections may have serious consequences.

2.6 Tools used

As shown in Table 2.2, these are the list of tools and specifications used to implement the project.

Table 2.2 List of tools used

Software Tools	**Dataset: Roboflow Colposcopy Images** The colposcopy image dataset used in this study was obtained from Roboflow, a platform offering diverse, high-quality datasets for machine learning research. The dataset includes high-resolution images of the colposcopy images, which were used for training, validating, and testing segmentation models. **Integrated Development Environment (IDE): Visual Studio Code** Visual Studio Code was chosen as the primary development environment for coding and managing the project. Its features, such as IntelliSense for code completion, debugging tools, and integration with Git, made it ideal for writing and debugging machine learning models. Additionally, extensions for Python and Jupyter Notebook enhanced workflow efficiency. **Programming Language: Python 3.8** Python was used due to its extensive library ecosystem and suitability for machine learning tasks. Libraries such as TensorFlow, Keras, NumPy, and Matplotlib were essential for implementing, training, and visualizing

	segmentation models. OpenCV and scikit-image were used for image preprocessing and augmentation.
Hardware Tools	**Laptop Specifications** The project was developed on a laptop with the following specifications: - Model: Nitro AN515-55 - Processor: Intel Core i5-10300H - RAM: 8GB - Storage: 512GB SSD - Graphics: NVIDIA GeForce GTX 1650 with 4GB GDDR6 - Operating System: Windows 11 Home Single Language
Integration of Tools	The colposcopy dataset from Roboflow was pre-processed and analyzed using Python in Visual Studio Code. Training and evaluation of the segmentation models were conducted on local laptop. The combination of tools ensured an efficient workflow, balancing local development with the computational demands of machine learning.

2.7 Summary

The Chapter 2 Literature Review provides a comprehensive review of existing research relevant to image segmentation in cervical cancer detection. The chapter begins by discussing the significance of accurate endocervical segmentation in colposcopy images for early detection and diagnosis of cervical cancer, while highlighting the clinical challenges and the need for automated approaches. The chapter then reviews various segmentation algorithms, such as K-means, Random Forest, U-net and other segmenting architectures, examining their applications and performance in medical imaging. Each technique's strengths, limitations, and suitability for endocervical segmentation are evaluated based on recent studies. Additionally, the chapter explores the role of different evaluation metrics, like Intersection over Union (IoU), Dice coefficient, F1 score and more to discusses their relevance in comparing model effectiveness for cervical imagery. Lastly, the chapter also included the integration of tools that will be used for this study.

CHAPTER 3 METHODOLOGY

3.1 Introduction

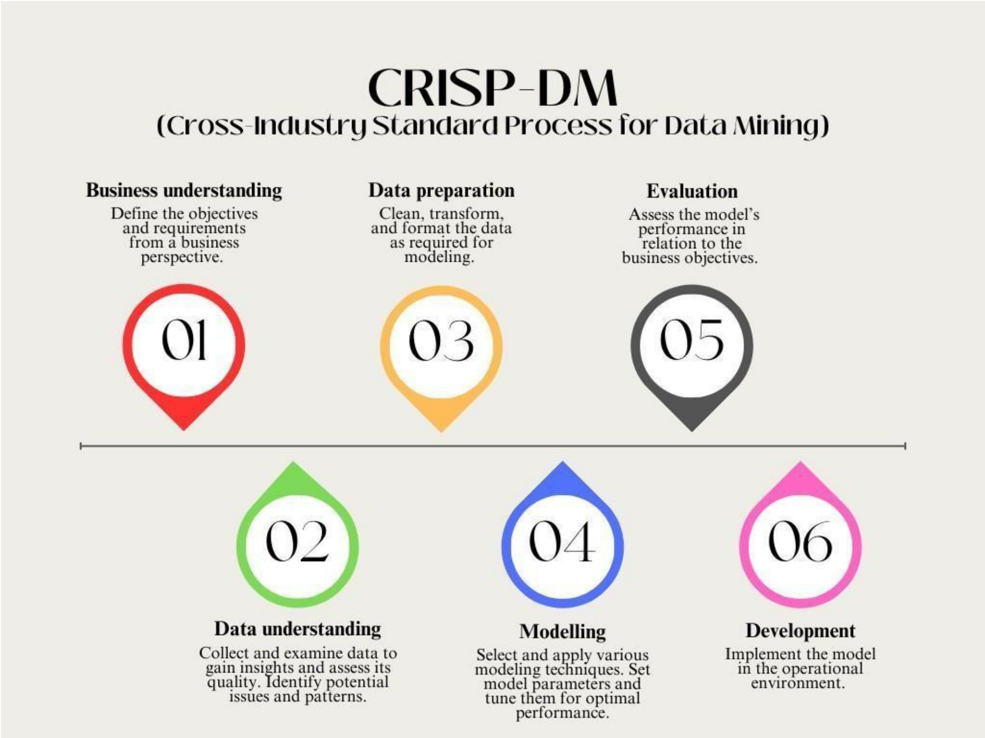


Figure 3.1 CRISP-DM methodology with brief explanation

This chapter presents the methodology (refer to Figure 3.1) adopted for this study, which focuses on comparing various segmentation techniques for segmenting the endocervix in colposcopy images. The research employs the CRISP-DM (Cross Industry Standard Process for Data Mining) framework to ensure a systematic and structured approach to the study. The methodology consists of six-phase process: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. Each phase has been carefully tailored to align with the objectives of the study and address the challenges associated with colposcopy image segmentation. By following this structured approach, the study aims to identify the most effective segmentation technique, providing a foundation for future advancements in cervical cancer diagnosis.

3.2 Business understanding

The business understanding phase focuses on understanding the project objectives and requirements from a business perspective. For this research, the primary objective of this study is to identify the most effective image segmentation technique for segmenting the endocervix in colposcopy images, aiming to support cervical cancer diagnosis. This research will compare several segmentation algorithms, such as K-means clustering, Random Forest, Support Vector Machine and U-net to evaluate their performance based on the evaluation metrics such as the Intersection over Union, Dice Coefficient, and F1 Score.

3.3 Data understanding

In this phase, data understanding involves exploring and familiarizing with the available data to assess its suitability for the project. The dataset for this study will be sourced from Roboflow, consisting colposcopy images of varying resolutions. This medium-sized dataset presents challenges such as anatomical variations and possible artifacts that may impact segmentation performance. During this phase, the quality, distribution, and potential challenges in the images, such as noise and irregular patterns, will be explored.

3.4 Data preprocessing

In data pre-processing phase, data preparation will involve several pre-processing steps to ensure the colposcopy images dataset are consistent and ready for segmentation in the modelling phase later on. Among the tasks include labelling, resizing, enhancing, and organizing images to ensure compatibility with various machine learning or deep learning models.

Firstly, the colposcopy image dataset will first be normalized for its intensity, then resized to a uniform resolution, and lastly enhanced for contrast, while mitigating artifacts to ensure consistency across the dataset. Next, if manually labelled regions (segmentation masks) showing the endocervical area (labelled ground truth masks) for the endocervical region are not available, manual annotation will be performed to generate the necessary segmentation labels. These labels are crucial for training the segmentation model because they provide a reference for what the model should learn to identify. Lastly, the dataset will be divided into two subsets: training (70%) and testing (30%) sets, to ensure that the models can be properly trained and evaluated.

3.5 Modelling

In the modelling phase, four different segmentation algorithms will be implemented for comparison. This study will implement and compare four segmentation algorithms: K-means clustering, Random Forest, Support Vector Machine (SVM), and U-Net. U-Net represents a deep learning approach, whereas the other models fall under traditional machine learning methods. The implementation will be carried out using Python libraries such as OpenCV and scikit-learn. Each algorithm's parameters will be tuned to optimize their performance, ensuring that each method is evaluated under optimal conditions. This phase is essential to test the efficacy of the different segmentation techniques.

3.6 Evaluation

In the evaluation phase, assessing the performance of each segmentation model is crucial to determine which is the most effective for segmenting the endocervix in colposcopy images. This study employs six quantitative evaluation metrics: Accuracy, Precision, Sensitivity (Recall), Intersection over Union (IoU), Dice Coefficient, and F1 Score. These metrics are calculated based on pixel-wise comparisons between the predicted segmentation masks and the corresponding ground truth binary masks. From these comparison, True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) were derived, enabling the computation of all six metrics. The evaluation process was applied consistently across all models to ensure a fair comparison. In addition to numerical results, visual comparisons between the original images, ground truth masks, and predicted segmentations were saved for qualitative inspection. The results of this evaluation allow for ranking the models and identifying the most suitable technique for endocervical segmentation.

3.7 Deployment

The final phase involves documenting and presenting the findings in the thesis, which will include both quantitative results and visual outputs. The best-performing segmentation model will be demonstrated using Hugging Face. Based on the evaluation results from the previous phase, the most effective model will also be recommended for use in colposcopy image analysis. In addition, this study will summarize the results, discuss the implications for cervical cancer detection, and suggest practical applications for clinicians.

3.8 Summary

Chapter 3 outlined the methodological approach adopted in this study, following the CRISP-DM framework. It began with a clear definition of the research objectives, emphasising the importance of accurate endocervical segmentation to support cervical cancer screening. The dataset, obtained from Roboflow, was introduced along with detailed data preparation steps including annotation, preprocessing, and resizing to optimise it for segmentation tasks. Four segmentation models — K-means clustering, Random Forest, Support Vector Machine (SVM), and U-Net with a ResNet34 backbone were selected, implemented, and trained on the prepared dataset. A consistent evaluation pipeline was applied across all models. Performance was assessed using six metrics: Accuracy, Precision, Sensitivity (Recall), IoU, Dice Coefficient, and F1 Score. This chapter also described the procedures for generating visual outputs and saving quantitative results for analysis. The structured methodology ensures reproducibility and provides a foundation for identifying the most effective segmentation approach, as further analysed in Chapter 4.

CHAPTER 4 IMPLEMENTATION AND TESTING

4.1 Introduction

This chapter outlines the implementation and testing processes carried out in this study to evaluate the effectiveness of various segmentation algorithms for endocervix segmentation in colposcopy images. The objective of this phase is to translate the design and methodology proposed in earlier chapters into a working solution and to assess its performance using appropriate evaluation metrics. The chapter begins by detailing the dataset preparation steps, including image preprocessing, mask creation, and dataset organization. Following this, the implementation of multiple segmentation techniques such as K-Means and other selected algorithms. Each technique is implemented using Python and tested using a consistent set of parameters to ensure fair comparison. Where applicable, additional steps such as model training, parameter tuning, and segmentation mask generation are explained. Subsequently, the chapter presents the testing procedures used to evaluate the segmentation performance of each algorithm. Evaluation metrics such as Dice Coefficient, Intersection over Union (IoU), and F1 score are employed to quantify segmentation quality. The results from each technique are then analyzed and compared to determine the most effective model for segmenting the endocervical region.

4.2 Data preprocessing

Data preprocessing is a crucial step in preparing the colposcopy image dataset for segmentation tasks. This stage involves organizing, cleaning, and transforming raw image and mask data into a structured format suitable for machine learning and deep learning models. Effective preprocessing enhances model consistency, improves segmentation accuracy, and minimizes the impact of noise and variability.

4.2.1 Dataset Description

This study uses a total of 1,000 colposcopy images obtained from the Roboflow VIA dataset, which focuses on cervical cancer screening (refer to Figure 4.1). Each image captures the cervix under magnification and reflects variations in lighting, contrast, and anatomical features, simulating the diversity encountered in clinical environments.

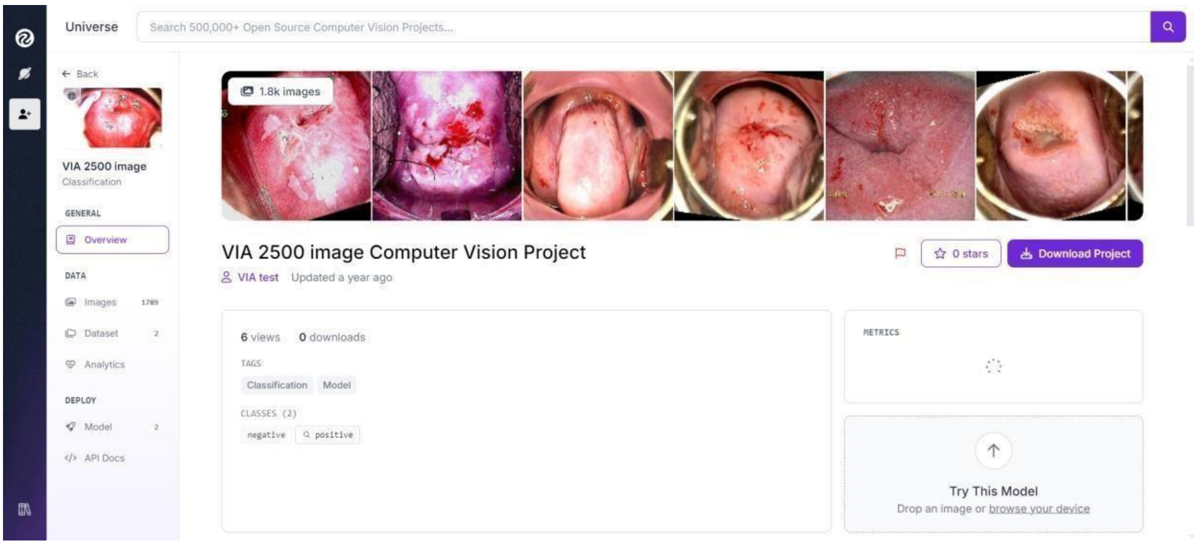


Figure 4.1 Roboflow VIA dataset

All images were standardized to a resolution of 640 × 640 pixels to maintain uniformity during model training. The dataset includes manually annotated segmentation masks that identify the endocervix region, which serves as the primary region of interest (ROI) for this research. A portion of the masks were verified by a medical expert, enhancing the reliability of the annotations used in supervised learning and evaluation.

4.2.2 Image Resizing

To ensure consistency in input dimensions and reduce computational load, all images and corresponding masks were resized to 640 × 640 pixels. This resizing was critical for compatibility with the input requirements of both machine learning and deep learning segmentation models and helped preserve important anatomical features.

4.2.3 Mask Creation and Verification

To establish a consistent definition of the target region during mask verification, anatomical references from relevant medical literature were consulted. The endocervix refers to the inner region of the cervix, which appears reddish and glandular, whereas the ectocervix is the outer region, typically exhibiting a pale pink appearance with stratified squamous epithelium as shown in Figure 4.2.

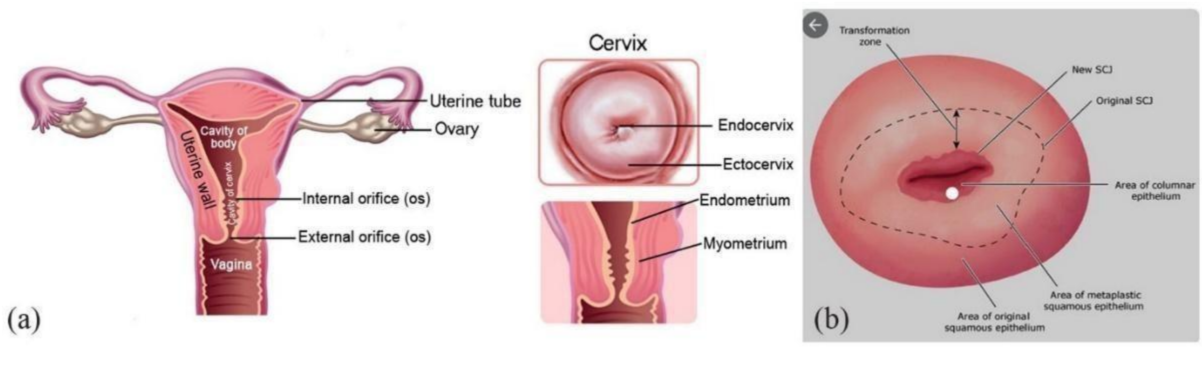


Figure 4.2 Anatomical view of cervix (a), cross-sectional view of cervix (b) (Dash et al., 2023)

Based on Figure 4.3, the squamocolumnar junction (SCJ) marks the transitional boundary between these two zones and can vary in shape, appearing either circular or irregular depending on individual anatomy. This anatomical understanding was used as a reference to validate the correctness of the annotated segmentation masks.

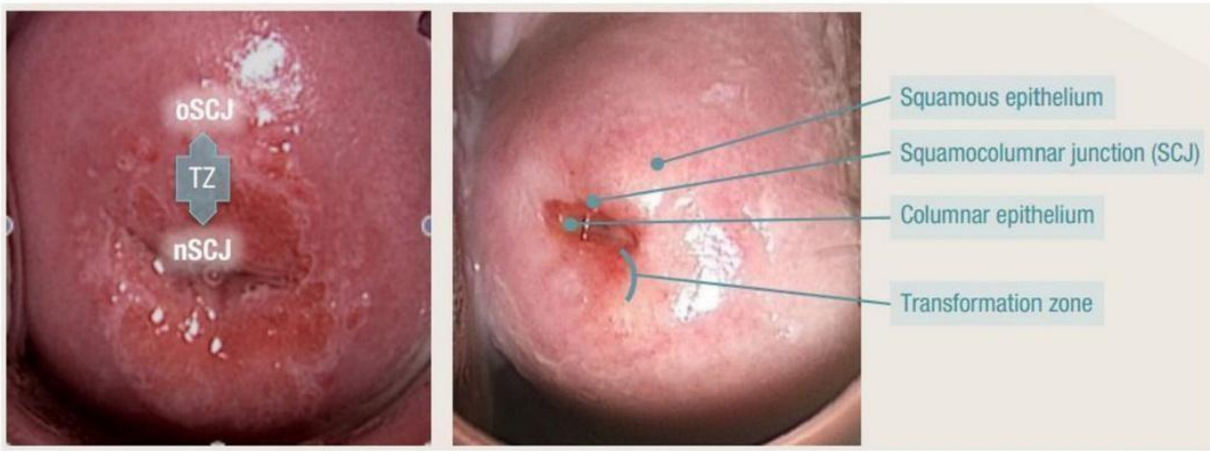


Figure 4.3 Illustration of transformation zone and squamocolumnar junction (Dash et al., 2023)

Each segmentation mask was carefully verified to ensure accurate alignment with its corresponding colposcopy image. The masks were encoded in binary format, where pixels labeled as 1 (white) indicate the endocervix region, and 0 (black) denote the background. All annotated masks and segmented region of the endocervix were stored locally for use in model training and evaluation. The comparison image of original image, ground truth mask and segmented region were also save (refer to Figure 4.4) to be use later on for comparing the segmentation model performance.

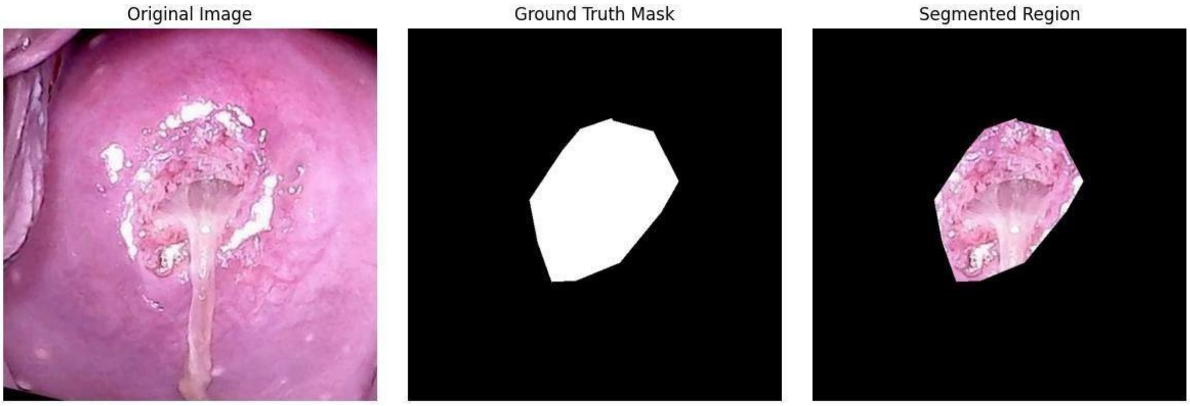


Figure 4.4 Original image, Ground Truth Mask and Segmented Region

4.2.4 Data Augmentation

To improve model generalization and reduce overfitting, data augmentation techniques were applied during the training phase. Specifically, horizontal and vertical flips, as well as random rotations, were introduced to simulate real-world variations in camera angles and patient positioning. These augmentations help enhance the robustness of the segmentation models by exposing them to a broader range of input scenarios. Importantly, each augmented image retained its corresponding segmentation mask to ensure the integrity and accuracy of the training labels were preserved.

4.2.5 Data Splitting

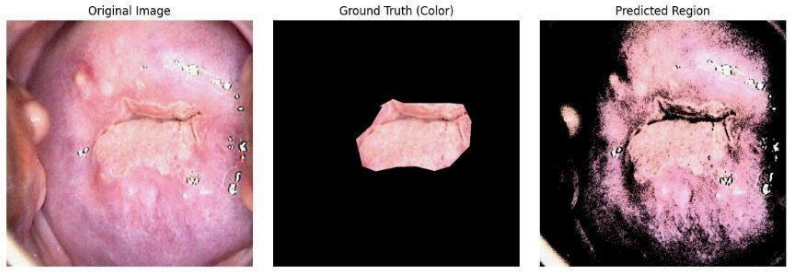
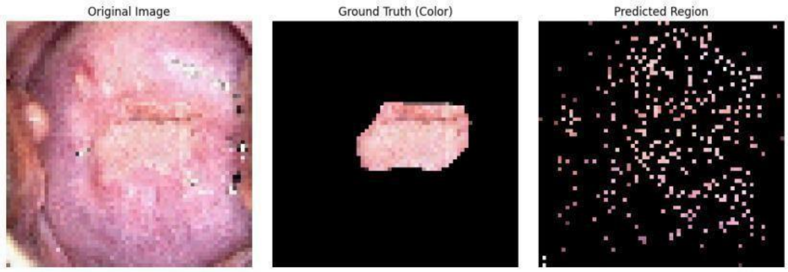
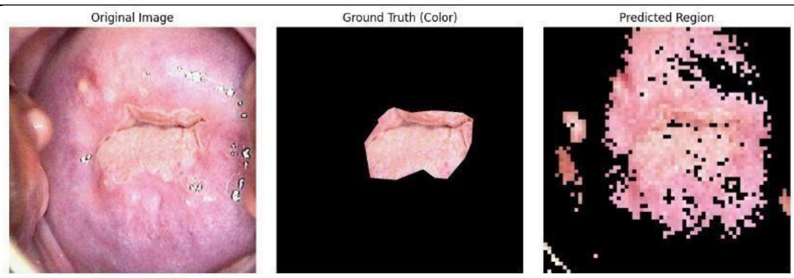
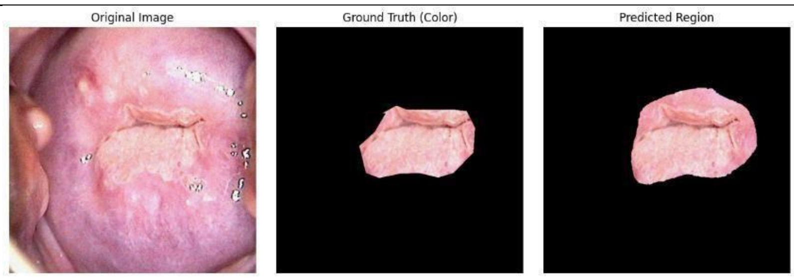
The dataset was divided into training and testing subsets using a 70:30 ratio. Seventy percent of the data was allocated for training, while the remaining thirty percent was used exclusively for testing model performance. The split was conducted randomly but reproducibly to ensure fair evaluation on unseen data. No separate validation set was used, as hyperparameter tuning was not the primary objective of this study.

4.3 Segmentation Models Implementation

In this study, four different segmentation algorithms were implemented to segment the endocervix region from colposcopy images. The models include K-means clustering, Random Forest, Support Vector Machine (SVM), and U-Net, representing a combination of traditional machine learning techniques and a deep learning-based approach.

K-means clustering is an unsupervised algorithm that partitions the image into clusters based on pixel intensity or color similarity. It is useful for identifying general patterns in the image without relying on labeled data. Random Forest, a supervised ensemble learning method, classifies image pixels using a collection of decision trees trained on image features such as color and texture. Support Vector Machine (SVM), another supervised algorithm, separates pixel classes using a hyperplane that maximizes the margin between classes in feature space. On the other hand, U-Net, a deep learning model based on convolutional neural networks (CNNs), was implemented using Python and TensorFlow. U-Net is specifically designed for biomedical image segmentation and features a symmetrical encoder-decoder structure that captures both contextual and spatial information. Given its strong performance in medical imaging tasks, U-Net was selected as the baseline model for this study.

All models were trained and evaluated using the same dataset and consistent input dimensions. Segmentation outputs from each model were stored and later analyzed both quantitatively using evaluation metrics (IoU, Dice coefficient, F1 Score) and qualitatively through visual comparison. Screenshots of the segmented prediction from the model (predicted region) and corresponding ground truth masks are presented in Figure 4.5 to 4.8 to support interpretation and comparison.

Segmentation Model	Output
K-means	 ***Figure 4.5 K-means Predicted mask***
Random Forest	 ***Figure 4.6 Random Forest Predicted mask***
Support Vector Machine	 ***Figure 4.7 Support Vector Machine Predicted mask***
U-net	 ***Figure 4.8 U-net Predicted mask***

4.4 Experimental Setup

To ensure the validity, fairness, and reproducibility of the segmentation model comparisons, all experiments were conducted under a controlled and consistent computational environment. The system used for experimentation was equipped with an Intel Core i7 processor, 16 GB of RAM, and an NVIDIA GeForce GTX 1660 GPU, providing sufficient processing power to handle both traditional machine learning algorithms and more computationally demanding deep learning models. The software environment was based on Python 3.9, which served as the primary programming language for implementation. Several key libraries were employed to facilitate model development and evaluation, including TensorFlow 2.8.0 for deep learning, scikit-learn for classical machine learning models, NumPy for efficient numerical computation, OpenCV for image processing, and the segmentation-models library for streamlined access to state-of-the-art deep learning architectures like U-Net.

The colposcopy image dataset was preprocessed to ensure uniformity and reliability. It was divided into 70% for training and 30% for testing, following a conventional split strategy to assess generalization performance. Prior to input into any model, all images and their corresponding masks were resized to a fixed dimension of 640×640 pixels. This resolution was selected to strike a balance between preserving anatomical detail and maintaining computational efficiency. The images were normalized to a \[0, 1] scale by dividing pixel values by 255, allowing for consistent input distributions across all models.

Each segmentation model was implemented according to its suitable configuration and operational characteristics. The K-Means clustering algorithm, an unsupervised method, was applied with $k = 2$ to distinguish between foreground (suspected lesion areas) and background based solely on pixel color similarity. It operated on flattened pixel arrays without incorporating spatial information. The Random Forest model was trained using 100 decision trees and utilized color-based pixel features extracted from labeled data, enabling it to make more informed predictions than K-Means. The Support Vector Machine (SVM) model employed a Radial Basis Function (RBF) kernel, and its performance was optimized through manual tuning of the C and γ parameters. This model leveraged a limited but informative set of features and aimed to construct an optimal hyperplane to separate foreground and background regions.

In contrast, the U-Net deep learning model offered an end-to-end segmentation pipeline capable of learning both local and global spatial features. The architecture adopted a ResNet34 backbone for its encoder, enabling it to capture hierarchical features effectively. The U-Net model was trained for 25 epochs with a batch size of 8, using the Adam optimizer, which is well-suited for training deep convolutional networks. A composite loss function combining binary cross-entropy and Dice loss was employed to penalize both pixel-wise classification errors and overlap mismatches between predicted and ground truth masks, thereby improving segmentation accuracy.

To ensure consistency in evaluation, all models were assessed using three widely recognized metrics: Dice Coefficient, Intersection over Union (IoU), and F1 Score. These metrics provided complementary perspectives on the quality of segmentation, capturing overlap accuracy, class balance, and pixel-level agreement. Furthermore, random seeds were set wherever applicable to reduce variability and enable reproducibility of experimental results.

This systematic and unified setup ensured that each model was evaluated fairly under comparable conditions, making the comparative results both reliable and meaningful.

4.5 Results and Analysis

This section presents both the quantitative and qualitative results of the segmentation models implemented in this study: K-means, Random Forest, Support Vector Machine (SVM), and U-Net. The performance of each model was evaluated using six standard segmentation metrics: Accuracy, Precision, Sensitivity (Recall), Intersection over Union (IoU), Dice Coefficient, and F1 Score. These metrics assess the similarity between predicted segmentation masks and the ground truth, providing a comprehensive understanding of each model's performance.

4.5.1 Quantitative Evaluation

Each model was tested on the same dataset split to ensure a fair comparison. The Table 4.1 below summarizes the average performance of each model across the test set:

Table 4.1 Performance Comparison Table

Segmentation Method	Mean IoU Score	Mean Dice Coefficient	Mean F1 Score	Mean Accuracy	Mean Precision	Mean Sensitivity (Recall)
K-means	0.25	0.38	0.38	0.63	0.30	0.64
Random Forest	0.20	0.32	0.32	0.76	0.39	0.34
SVM	0.29	0.42	0.43	0.60	0.34	0.80
U-net **(Highest)**	0.70	0.81	0.81	0.93	0.76	0.92

Among all the models evaluated, U-Net outperformed the others across all six metrics, achieving the highest values in IoU, Dice Coefficient, F1 Score, Accuracy, Precision, and Sensitivity. This indicates that U-Net is not only able to predict the region of interest more precisely, but also does so with fewer false positives and false negatives. In contrast, traditional machine learning methods like K-means, Random Forest, and SVM struggled to generalise well to the test data, producing lower segmentation accuracy.

4.5.2 Qualitative Comparison

To visually assess the segmentation quality, sample outputs from each model were compared with the corresponding ground truth masks. The segmentation results are illustrated in Figures 4.9 to 4.12 consisting the original image, ground truth, and predicted mask for each model.

Segmentation model	**Outputs**
K-means	 ***Figure 4.9 Visual output of K-means segmentation***

<table>
<tr>
<td>Random forest</td>
<td>

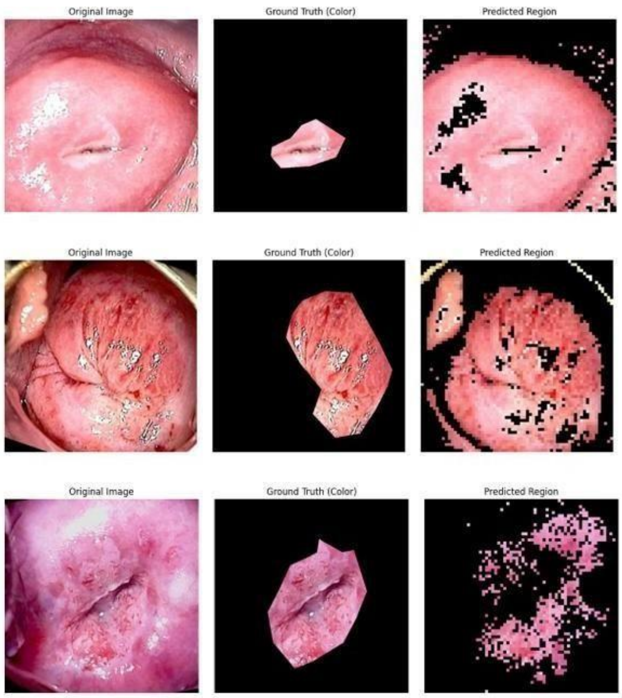
Figure 4.10 Visual output of Random Forest segmentation</td>
</tr>
<tr>
<td>Support vector
machine</td>
<td>

Figure 4.11 Visual output of SVM segmentation</td>
</tr>
</table>

U-net	 ***Figure 4.12 Visual output of U-Net segmentation***

From the visual analysis, it is evident that U-Net produced cleaner, more anatomically accurate segmentations that closely matched the annotated regions. In contrast, traditional ML models (K-means, Random Forest, SVM) occasionally misclassified background regions or missed finer boundary details.

4.6 Discussion

This chapter presented the implementation and evaluation of multiple image segmentation algorithms to identify the endocervix region in colposcopy images. The entire experimental process, from dataset preprocessing to mask alignment and model evaluation, was standardized to ensure fairness and consistency across all tested methods. The models explored included traditional machine learning approaches: K-Means clustering, Random Forest, and Support Vector Machine (SVM) as well as a deep learning model, U-Net with a ResNet34 backbone.

The evaluation based on Dice Coefficient, Intersection over Union (IoU), and F1 Score revealed clear performance differences among the models. U-Net consistently achieved the highest scores across all metrics, highlighting its robustness and effectiveness for medical image segmentation. This is largely due to U-Net's encoder-decoder architecture, which allows it to capture both spatial context and fine-grained details. The skip connections in U-Net help preserve important edge and shape information, which is critical when segmenting anatomical regions like the endocervix that may vary in texture, shape, and intensity.

In comparison, classical machine learning models performed moderately well but fell short in capturing the complex spatial relationships inherent in medical images. K-Means Clustering, as an unsupervised method, segmented images based purely on pixel color similarity. While straightforward and fast, this approach lacked awareness of spatial structure, resulting in noisy and fragmented segmentations, especially in cases where the endocervical region did not exhibit strong color contrast from surrounding tissue. Random Forest showed an improvement by learning from labeled data and leveraging the power of multiple decision trees to classify pixels based on extracted features. However, its reliance on handcrafted features and the absence of spatial context understanding limited its effectiveness, often producing masks that lacked smoothness and anatomical coherence. Support Vector Machine (SVM), on the other hand, outperformed both K-Means and Random Forest by establishing more refined decision boundaries in the feature space, leading to better localization of the target region. Nonetheless, SVM still struggled with global spatial awareness and was prone to pixel-level misclassifications in areas of ambiguity. These limitations highlight the fundamental gap between classical models and deep learning approaches in handling complex, structured visual data such as colposcopy images.

In essence, the difference in performance arises from how each algorithm processes information. While machine learning methods classify pixels independently, deep learning models like U-Net learn hierarchical features and spatial dependencies, enabling them to generalize better in visually complex tasks like endocervix segmentation. These results affirm the advantage of deep learning models in medical imaging applications. Despite their higher computational cost and longer training time, their ability to learn both visual semantics and spatial context make them superior for tasks requiring high precision and reliability.

4.7 Summary

Chapter 4 presented the implementation and evaluation of various image segmentation techniques for endocervix segmentation in colposcopy images. The dataset preparation process including image preprocessing, mask creation, and the organization of input-output pairs was carried out carefully to ensure the reliability and consistency of the experiments. Multiple segmentation models were developed and tested, including machine learning-based algorithms such as K-Means Clustering, Random Forest, and Support Vector Machine (SVM), as well as the deep learning-based U-Net model. Each model was evaluated using six standard performance metrics: Intersection over Union (IoU), Dice Coefficient, F1 Score, Accuracy, Precision, and Sensitivity (Recall). This multi-metric approach enabled a more comprehensive and balanced assessment of segmentation performance. The results demonstrated that U-Net consistently outperformed the traditional machine learning models across all metrics, highlighting its robustness and suitability for complex medical image segmentation tasks, particularly in endocervical localisation. This experimental phase successfully identified U-Net as the most effective and reliable segmentation technique within the scope of this study, setting a valuable benchmark for future research and improvements in automated cervical image analysis.

CHAPTER 5 CONCLUSION AND FUTURE WORKS

5.1 Summary of Unaddressed Areas

This research focused exclusively on image segmentation of the endocervix using colposcopy images. Although segmentation is a critical pre-processing step in cervical cancer analysis, this study did not extend to feature extraction or classification of cancer stages. These downstream tasks are essential for building a comprehensive computer-aided diagnosis (CAD) system but were deliberately excluded due to the project's defined scope and time limitations.

The dataset was limited to publicly available images sourced from the Roboflow VIA dataset. While no clinical validation was performed using private hospital patient data, a portion of the dataset includes expert-verified annotations of the endocervix region by a medical doctor. This contributes to the clinical reliability of the ground truth used for model training and evaluation.

5.2 Contributions

Despite the limited scope, this study has made meaningful contributions to the area of medical image segmentation. The project successfully implemented and compared multiple segmentation models, including traditional machine learning techniques, such as K-Means, Random Forest, and Support Vector Machine as well as a deep learning method, U-Net. These models were evaluated using quantitative performance metrics, Intersection over Union (IoU), Dice Coefficient, and F1 Score to identify the most accurate segmentation approach for the endocervical region.

The results demonstrated that U-Net outperformed the other techniques across most metrics. This finding provides valuable insight into which segmentation methods are more suitable for colposcopy images and could serve as a foundation for future cervical cancer diagnostic systems.

5.3 Limitations of the Study

While the research produced promising outcomes, several limitations were encountered. First, the study maintained a narrow focus solely on image segmentation and did not extend to subsequent stages such as classification or diagnosis, which are essential for a full computer-aided diagnosis pipeline. Second, the dataset used was limited in both size and diversity, potentially affecting the generalizability of the results to broader clinical applications. Additionally, the study lacked direct clinical input, as no feedback or validation was obtained from medical professionals. Lastly, the range of segmentation models tested was limited, more recent and advanced architectures, such as transformer-based models, were not explored due to time and resource constraints.

5.4 Future Work

Building upon the outcomes of this study, several promising directions can be pursued to further advance research in this area. A key progression would be to move beyond segmentation by integrating feature extraction and classification components, ultimately contributing to the development of a more comprehensive and functional cervical cancer diagnosis system. In parallel, investigating more advanced segmentation architectures such as DeepLabV3+, Attention U-Net, or transformer-based models may offer opportunities to enhance accuracy and overall performance.

To strengthen the reliability of the results, future studies should consider using larger and more diverse datasets, ideally including clinically annotated images from hospitals. This would help improve the robustness and generalizability of the models across different populations and imaging conditions. In addition, incorporating explainable AI (XAI) techniques and seeking input from medical professionals can make the models not only more transparent but also more aligned with real clinical needs. Finally, optimizing these models for real-time processing could support live clinical screenings and bring them one step closer to being used in practical, real-world healthcare settings.

5.5 Conclusion

This study demonstrated the effectiveness of both machine learning and deep learning techniques for segmenting the endocervix in colposcopy images. Among the models evaluated, U-net achieved the highest accuracy based on key evaluation metrics, highlighting its potential for use in medical image segmentation tasks. The findings reinforce the critical role of image preprocessing, particularly segmentation in the broader pipeline of automated cervical cancer detection systems.

Despite the limitations outlined, this research establishes a solid foundation for future development and refinement. It opens several pathways for extending the system into a comprehensive diagnostic tool. With continued enhancements and clinical validations, such AI-assisted approaches hold great promise for advancing cervical cancer screening, especially in low-resource healthcare environments where expert availability is limited.

REFERENCES

Ahishakiye, E., & Kanobe, F. (2024). Optimizing Cervical Cancer Classification through Transfer Learning and Kernel Methods: Analyzing the Performance of Deep Gaussian Processes and Support Vector Machines on Pap smear Image Data. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-4791585/v1

Arora, A., Tripathi, A., & Bhan, A. (2021, January 1). *Classification of Cervical Cancer Detection using Machine Learning Algorithms*. IEEE Xplore. https://doi.org/10.1109/ICICT50816.2021.9358570

Arum, A. W., Nurmaini, S., Rini, D. P., Agustiansyah, P., & Rachmatullah, M. N. (2021). Segmentation of Squamous Columnar Junction on VIA Images using U-Net Architecture. *Computer Engineering and Applications Journal*, *10*(3), 209–219. https://doi.org/10.18495/comengapp.v10i3.387

Babitz, K. (2018, July 5). *Introduction to k-Means Clustering with scikit-learn in Python*. Datacamp.com; DataCamp. https://www.datacamp.com/tutorial/k-means-clustering-python

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology*, *188*(12), 2222–2239. https://doi.org/10.1093/aje/kwz189

Burness, J. V., Schroeder, J. M., & Warren, J. B. (2020). Cervical Colposcopy: Indications and Risk Assessment. *American Family Physician*, *102*(1), 39–48. https://www.aafp.org/pubs/afp/issues/2020/0701/p39.html

Dasgupta, A., & Nath, A. (2016). Classification of machine learning algorithms. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, *3*(3), 6-11.

Desai, K. T., Befano, B., Xue, Z., Kelly, H., Campos, N. G., Egemen, D., Gage, J. C., Rodriguez, A., Sahasrabuddhe, V., Levitz, D., Pearlman, P., Jeronimo, J., Antani, S., Schiffman, M., & Sanjosé, S. (2021). The development of "automated visual evaluation" for cervical cancer screening: The promise and challenges in adapting deep-learning for clinical testing. *International Journal of Cancer*, *150*(5), 741–752. https://doi.org/10.1002/ijc.33879

Gunay, D. (2023, September 14). *Random Forest*. Medium. https://medium.com/@denizgunay/random-forest-af5bde5d7e1e

Hartmann, D., Müller, D., Soto-Rey, I., & Kramer, F. (2021). *Assessing the Role of Random Forests in Medical Image Segmentation*. https://arxiv.org/pdf/2103.16492

Hou, X., Shen, G., Zhou, L., Li, Y., Wang, T., & Ma, X. (2022). Artificial Intelligence in Cervical Cancer Screening and Diagnosis. *Frontiers in Oncology*, *12*. https://doi.org/10.3389/fonc.2022.851367

Jiménez Gaona, Y., Castillo Malla, D., Vega Crespo, B., Vicuña, M. J., Neira, V. A., Dávila, S., & Verhoeven, V. (2022). Radiomics Diagnostic Tool Based on Deep Learning for Colposcopy Image Classification. *Diagnostics*, *12*(7), 1694. https://doi.org/10.3390/diagnostics12071694

Klingler, N. (2024, April 23). *U-Net: A Comprehensive Guide to Its Architecture and Applications*. Viso.ai. https://viso.ai/deep-learning/u-net-a-comprehensive-guide-to-its-architecture-and-applications/

Li, Z., Feng, N., Pu, H., Dong, Q., Liu, Y., Liu, Y., & Xu, X. (2022). Pixel-Level Segmentation of Bladder Tumors on MR Images Using a Random Forest Classifier. *Technology in Cancer Research & Treatment*, *21*. https://doi.org/10.1177/15330338221086395

Mahesh, B. (2018). Machine learning algorithms - a review. *International Journal of Science and Research (IJSR) ResearchGate Impact Factor*, *9*(1). https://doi.org/10.21275/ART20203995

Mehta, A. (2023, March 11). *Support Vector Machine (SVM) Algorithm For Machine Learning*. Medium. https://ashish-mehta.medium.com/support-vector-machine-svm-algorithm-for-machine-learning-350fe9139a52

Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, *15*(1). https://doi.org/10.1186/s13104-022-06096-y

Nazir, N., Sarwar, A., Saini, B. S., & Shams, R. (2023). A Robust Deep Learning Approach for Accurate Segmentation of Cytoplasm and Nucleus in Noisy Pap Smear Images. *Computation*, *11*(10), 195. https://doi.org/10.3390/computation11100195

Riana, D., Rahayu, S., Hadianti, S., Frieyadie, Hasan, M., Karimah, I. N., & Pratama, R. (2022). Identifikasi Citra Pap Smear RepoMedUNM dengan Menggunakan K-Means Clustering dan GLCM. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(1), 1–8. https://doi.org/10.29207/resti.v6i1.3495

Saini, S., & Arora, K. (2014). A study analysis on the different image segmentation techniques. *International Journal of Information & Computation Technology, 4(14), 1445-1452.*

Sarica, A., Cerasa, A., & Quattrone, A. (2017). Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Frontiers in Aging Neuroscience*, *9*. https://doi.org/10.3389/fnagi.2017.00329

Shinohara, T., Murakami, K., & Matsumura, N. (2023). Diagnosis Assistance in Colposcopy by Segmenting Acetowhite Epithelium Using U-Net with Images before and after Acetic Acid Solution Application. *Diagnostics*, *13*(9), 1596–1596. https://doi.org/10.3390/diagnostics13091596

Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access*, *9*, 82031–82057. https://doi.org/10.1109/access.2021.3086020

Singh, D., Vignat, J., Lorenzoni, V., Eslahi, M., Ginsburg, O., Lauby-Secretan, B., Arbyn, M., Basu, P., Bray, F., & Vaccarella, S. (2022). Global Estimates of Incidence and Mortality of Cervical Cancer in 2020: a Baseline Analysis of the WHO Global Cervical Cancer Elimination Initiative. *The Lancet Global Health*, *11*(2). https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(22)00501-0/fulltext

Tian, X., Li, C., Hou, Y., Xie, J., Song, M., Liu, K., & Zhou, J. (2022). Artificial intelligence in brachytherapy for cervical cancer. *Journal of Cancer Research and Therapeutics*, *18*(5), 1241–1241. https://doi.org/10.4103/jcrt.jcrt_2322_21

Torode, J., Kithaka, B., Chowdhury, R., Simelela, N., Cruz, J. L., & Tsu, V. D. (2021). National action towards a world free of cervical cancer for all women. *Preventive Medicine*, *144*, 106313. https://doi.org/10.1016/j.ypmed.2020.106313

Vankayalapati, R., Ghutugade, K. B., Vannapuram, R., & Prasanna, B. P. S. (2021). K-Means Algorithm for Clustering of Learners Performance Levels Using Machine Learning Techniques. *Revue d'Intelligence Artificielle*, *35*(1), 99–104. https://doi.org/10.18280/ria.350112

Widodo, S., Brawijaya, H., & Samudi, S. (2021). Clustering Kanker Serviks Berdasarkan Perbandingan Euclidean dan Manhattan Menggunakan Metode K-Means. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(2), 687. https://doi.org/10.30865/mib.v5i2.2947

Wijaya, R. S. D., Adiwijaya, Suksmono, A. B., & Mengko. T. L. (2021). Segmentasi Citra Kanker Serviks Menggunakan Markov Random Field dan Algoritma K-Means. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(1), 139–147. https://doi.org/10.29207/resti.v5i1.2816

Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., & Liu, F. (2024). Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches. *Bioengineering, 11(10), 1034.* https://doi.org/10.3390/bioengineering11101034

Xue, P., Ng, M. T. A., & Qiao, Y. (2020). The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence. *BMC Medicine*, *18*(1). https://doi.org/10.1186/s12916-020-01613-x

Yang, C., Qin, L.-H., Xie, Y.-E., & Liao, J.-Y. (2022). Deep learning in CT image segmentation of cervical cancer: a systematic review and meta-analysis. *Radiation Oncology (London, England)*, *17*(1), 175. https://doi.org/10.1186/s13014-022-02148-6

Dash, S., Sethy, P. K., & Behera, S. K. (2023). Cervical Transformation Zone Segmentation and Classification based on Improved Inception-ResNet-V2 Using Colposcopy Images. Cancer Informatics, 22, 11769351231161477. https://doi.org/10.1177/11769351231161477