International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Automated Endometrial Cancer Diagnosis with Machine Learning

Suhani A. Bodkhe[1], Saksham R. Dhumale[2], Nishant S. Deshmukh[3], Vaishnavi M. Kulkarni[4], Yash G. Iskape[5], Dr. N. M. Yawale[6]

[1,2,3,4,5]*Department of CSE, Prof. Ram Meghe Institute Of Technology And Research, Amravati, Maharashtra, India*
[6]*Assistant Professor, Dept. of CSE, Prof. Ram Meghe Institute of Technology & Research, Amravati, India*

***Abstract*— Endometrial cancer is one of the most common types of gynecological cancer and detecting it early can make a significant difference in patient outcomes. Currently, the most reliable way to diagnose it is through the microscopic examination of endometrial tissue. While effective, this manual process can be slow and may vary depending on who's interpreting the results.**

With the rise of digital pathology and the rapid advancement of machine learning, there's now an opportunity to improve how we detect cancer. This project explores a modern approach to identifying endometrial cancer by analyzing histopathological images using machine learning.

By training algorithms on a wide-ranging dataset of digitized endometrial tissue samples, the system learns to recognize patterns and features associated with cancer. The goal is to provide fast, accurate analysis — potentially in real time — which could help doctors detect cancer earlier and act more quickly. This technology holds promise as a powerful support tool in clinical settings, reducing diagnostic delays and improving care.

***Keywords*-- Endometrial Cancer, Histopathological Images, Machine Learning, Digital Pathology, Cancer Detection, Convolutional Neural Network (CNN), Medical Image Analysis**

I. INTRODUCTION

Endometrial cancer (EC), also known as uterine corpus cancer, is the sixth most commonly diagnosed cancer among women worldwide. In 2020 alone, there were an estimated 417,000 new cases and 97,000 deaths globally [1]. According to the American Cancer Society, in the United States, approximately 65,950 new cases and 12,550 deaths were expected in 2022 [2].Endometrial cancer is classified into four stages based on the 2009 staging criteria from the International Federation of Gynecology and Obstetrics (FIGO). In the early stages (FIGO Stage I or II), the tumor remains confined to the uterus, and the 5-year survival rate ranges from 74% to 91%. However, when the cancer spreads beyond the uterus, such as in FIGO Stage IV, the survival rate drops drastically to between 20% and 26%. This stark contrast highlights the importance of accurate and timely diagnosis.

Histopathological imaging plays a key role in diagnosis and treatment planning. It allows clinicians to closely examine tissue samples and make informed decisions. The American College of Radiology emphasizes the use of histopathology as the preferred method for evaluating the extent of disease, due to its ability to provide detailed insights. n traditional image analysis, techniques such as the Otsu method are often used to segment the uterine cavity and tumor regions. However, these methods typically require extensive domain knowledge and manual tuning of parameters. Although histopathological image classification has reached accuracy levels of around 85%, there remains a significant gap in research focused specifically on detecting lesion regions within these images. In traditional image analysis, techniques such as the Otsu method are often used to segment the uterine cavity and tumor regions. However, these methods typically require extensive domain knowledge and manual tuning of parameters.

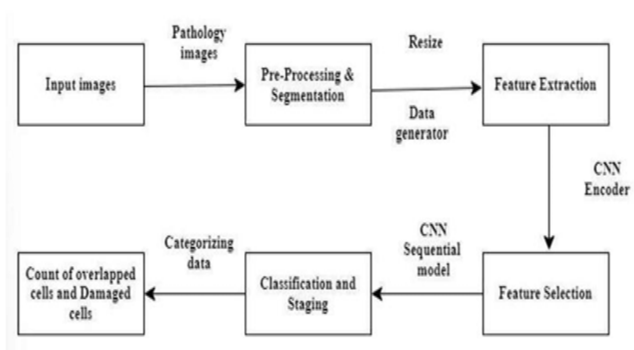

Fig. 1: Architectural Diagram of the Proposed System using CNN.

II. LITERATURE SURVEY

Haoming Lin and Jianhua Yao (2019) [1] developed a morphology-based fine-grained classification model using convolutional neural networks. Their model emphasizes the significance of cellular morphological features for accurate cancer staging, achieving noteworthy results in distinguishing endometrial cell types.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Md Mamunur Rahaman et al. (2020) [2] provided a comprehensive survey on deep learning techniques applied to endometrial histopathology. They evaluated a wide range of methods and highlighted the importance of preprocessing, segmentation, and model depth in achieving reliable classification outcomes.

Khaled Mabrouk Amer Adweb and colleagues (2021) [3] applied very deep neural networks with various activation functions to improve classification robustness. Their work explored how deeper architectures can enhance feature representation and perform better on complex histopathological datasets.

Zhiyun Xue et al. (2021) [4] introduced a deep metric learning approach for classifying endometrial images. By learning similarity metrics instead of direct classification, their model achieved high precision in distinguishing between subtle variations in tissue structures, which is critical in histopathological diagnostics.

Qazi Mudassar Ilyas and Muneer Ahmad (2021)[5] proposed an ensemble learning approach that integrates multiple classifiers for improved accuracy in endometrial cancer diagnosis using machine intelligence.

These studies support the use of machine learning—especially deep learning—as a reliable and scalable tool for automating endometrial cancer detection from histopathological images.

III. METHODOLOGY

This study proposes an automated system for diagnosing endometrial cancer using a hybrid approach of classical computer vision and deep learning. The methodology is executed in four sequential stages to accurately process and classify histopathological slide images:

1. *Image Preprocessing:* To prepare the uploaded histopathological images for analysis, they undergo a four-step computer vision pipeline using OpenCV. The images are first converted from RGB to grayscale to isolate intensity data, followed by Canny edge detection (with hysteresis thresholds of 250 and 254) to sharply map cellular boundaries. Finally, a global binary thresholding (value of 128) segments the image, distinctly separating the foreground cellular structures from the background.

2. *Cellular Feature Extraction:* Before classification, the system quantifies the cellular composition by applying a Gaussian blur (5x5 kernel) and contour detection to count the total number of cells. It identifies abnormal or damaged cells by filtering out contours with an area of less than 100 pixels. Additionally, the system detects overlapping cells—a common indicator of hypercellularity—by calculating spatial image moments and ensuring the Euclidean distance between cell centroids is less than 20 pixels.

3. *Deep Learning Classification:* The system's core diagnostic engine is a Convolutional Neural Network (CNN) built with TensorFlow and Keras. The preprocessed image is resized to a standardized 128x128 pixel tensor and normalized to a [0,1] range. This tensor is fed into the CNN, which outputs a Softmax probability distribution to classify the tissue into one of four categories: Endometrial Adenocarcinoma (Stage 1), Endometrial Hyperplasia (Stage 2), Endometrial Polyp (Stage 3), or Normal Endometrium. The class with the highest confidence score determines the final diagnosis.

4. *Application Architecture:* The entire diagnostic pipeline is deployed via a user-friendly Flask-based web application, utilizing an SQLite database for secure credential management and historical tracking. The backend dynamically generates comparative accuracy bar charts using Matplotlib, which are rendered on the HTML/CSS frontend alongside real-time cellular counts and sequential visualizations of the preprocessing stages.

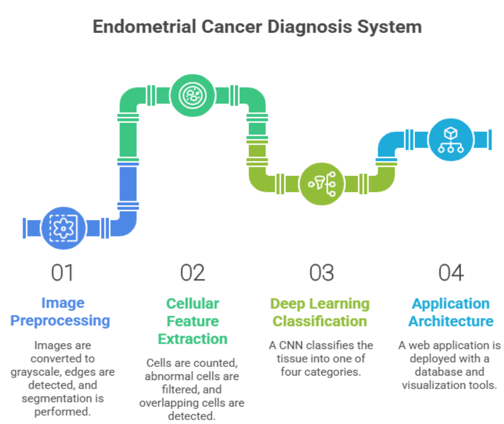

Fig. 2: Workflow of the Proposed System

IV. RESULT & ANALYSIS

The proposed system for endometrial cancer detection was implemented using Python and Convolutional Neural Networks (CNN), and evaluated using histopathological image datasets. The system processes input images through multiple stages including preprocessing, segmentation, feature extraction, and classification to produce accurate diagnostic results.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

In the preprocessing stage, input images are converted from RGB to grayscale to reduce computational complexity. Noise removal is performed using median filtering, followed by thresholding and high-pass filtering for image sharpening. These steps enhance the quality of images and improve the visibility of important cellular structures. Edge detection and segmentation techniques are then applied to isolate regions of interest, enabling more precise analysis of affected areas.

Feature extraction is carried out automatically by the CNN model, which learns hierarchical patterns from the processed images. The classification stage categorizes the images into cancerous and non-cancerous classes, and also assists in identifying the stage of cancer. The system additionally performs cell analysis such as counting total cells, detecting damaged cells based on area thresholds, and identifying overlapping cells using centroid distance calculations.

The results of the system are presented through various visual outputs including grayscale images, edge-detected images, segmented regions, and classification outputs. Graphical representations such as accuracy and loss curves are used to evaluate the model's training performance. The accuracy graph shows a steady improvement over epochs, indicating effective learning, while the loss graph demonstrates a decreasing trend, confirming proper convergence of the model.

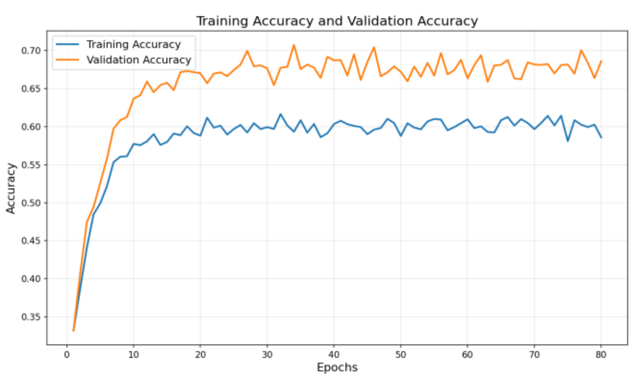

Fig. 3: Accuracy Graph of CNN Model

Fig. 4: Loss Graph during Training

Further evaluation using a confusion matrix shows that the majority of predictions are correctly classified, with only a small number of false positives and false negatives. Performance metrics such as precision and F1-score indicate that the model maintains a good balance between sensitivity and specificity, ensuring reliable detection of cancerous cells.

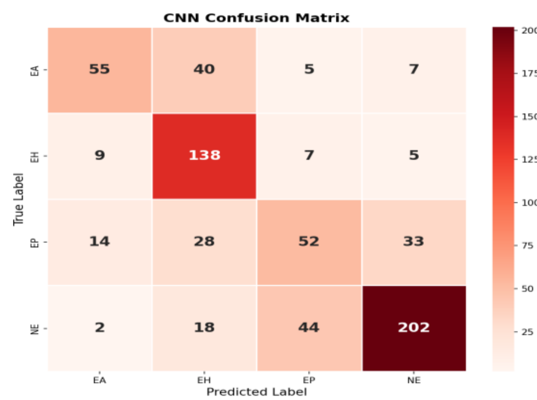

Fig. 5: Confusion Matrix of Classification Results

Overall, the experimental results demonstrate that the proposed system is capable of efficiently analyzing histopathological images and providing accurate predictions. The integration of image processing techniques with deep learning improves diagnostic performance and reduces dependency on manual analysis. This system can serve as a supportive tool for medical professionals in the early detection and diagnosis of endometrial cancer.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

V. Conclusion

This project presents an effective and automated system for the detection of endometrial cancer using image processing techniques and Convolutional Neural Networks (CNN). The proposed system was designed to analyze histopathological images through a structured pipeline consisting of preprocessing, segmentation, feature extraction, and classification. This systematic approach ensures that the input data is enhanced and refined before being used for final prediction, thereby improving the overall accuracy and reliability of the system.

The preprocessing stage plays a vital role in improving image quality by converting RGB images into grayscale, removing noise using median filtering, and applying thresholding and high-pass filtering techniques. These steps help in highlighting important cellular features and removing irrelevant information. The segmentation process further isolates the regions of interest, enabling focused analysis of potentially cancerous areas. The CNN model then automatically extracts meaningful features and classifies the images effectively, eliminating the need for manual feature engineering.

In addition to classification, the system incorporates advanced analytical capabilities such as cell counting, detection of damaged cells based on contour area, and identification of overlapping cells using centroid distance calculations. These additional features enhance the diagnostic strength of the system and provide more comprehensive insights into the condition of the tissue samples.

The performance evaluation of the system indicates that it achieves reliable accuracy, with consistent results observed across different evaluation metrics such as precision, recall, and F1-score. The use of graphical analysis, including accuracy and loss curves, confirms that the model is well-trained and capable of generalizing effectively. The confusion matrix further validates the model's ability to correctly classify most of the input samples with minimal errors.

Overall, the proposed system demonstrates significant potential in assisting medical professionals by providing a fast, accurate, and consistent method for detecting endometrial cancer. It reduces dependency on manual analysis, minimizes human error, and speeds up the diagnostic process. This makes it especially useful in clinical environments where timely and accurate diagnosis is critical.

For future work, the system can be enhanced by incorporating larger and more diverse datasets to improve generalization. Advanced deep learning architectures and optimization techniques can also be explored to further boost performance. Additionally, integrating the system into real-time clinical applications or web-based platforms can make it more accessible and practical for widespread use. With continuous improvements, this approach can contribute significantly to the advancement of computer-aided diagnosis in the medical field.

REFERENCES

[1] H. Lin and J. Yao, "Fine-Grained Classification of Endometrial Cells Using Morphological Appearance Based Convolutional Neural Networks," IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 5, pp. 1995–2004, May 2019.

[2] M. M. Rahaman, C. Li, and X. Wu, "A Survey for Endometrial Histopathology Image Analysis Using Deep Learning," Computerized Medical Imaging and Graphics, vol. 84, pp. 101772, Mar. 2020.

[3] K. M. A. Adweb, N. Cavus, and B. Sekeroglu, "Endometrial Cancer Diagnosis Using Very Deep Networks Over Different Activation Functions," IEEE Access, vol. 9, pp. 37664–37673, Mar. 2021.

[4] Z. Xue, B. Befano, and A. C. Rodriguez, "Deep Metric Learning for Endometrial Image Classification," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 2176–2183, Jun. 2021.

[5] Q. M. Ilyas and M. Ahmad, "An Enhanced Ensemble Diagnosis of Endometrial Cancer: A Pursuit of Machine Intelligence Towards Sustainable Health," Journal of King Saud University - Computer and Information Sciences, Nov. 2021.