IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14650

A Survey on Prediction of Endometrial Cancer and its Grade using Image Preprocessing and Machine Learning

Prof. Dr. Vijayalaxmi Mekali[1], Neha V[2], Prakruthi G P[3], Preetha D'Souza[4], S Hyma[5]

Professor, Dept of CSE, KSIT, Karnataka, India[1]

Student, Dept of CSE, KSIT, Karnataka, India[2]

Student, Dept of CSE, KSIT, Karnataka, India[3]

Student, Dept of CSE, KSIT, Karnataka, India[4]

Student, Dept of CSE, KSIT, Karnataka, India[5]

Abstract: One of the most common gynaecological cancers affecting women globally is endometrial cancer, which develops from the lining of the uterus. The prognosis is improved by early diagnosis, but traditional techniques like biopsy and ultrasound are frequently intrusive, costly, or inaccessible. Recent developments in artificial intelligence, specifically in the areas of machine learning (ML) and image processing, present promising instruments for the automated, non-invasive, and precise detection and grading of endometrial cancer. This survey investigates how to improve the quality of histopathological images for analysis using image preprocessing methods like RGB to grayscale conversion, noise reduction, thresholding, segmentation, and feature extraction. Additionally, it assesses how well deep learning models—particularly Convolutional Neural Networks (CNNs) and transfer learning techniques—classify malignant tissues and forecast tumour stage. and

Keywords: Endometrial Cancer, Image Preprocessing, Machine Learning, CNN, Histopathology, Medical Imaging, Transfer Learning, Deep Learning.

I. INTRODUCTION

Endometrial cancer is the most prevalent cancer of the female reproductive system, occurring mostly in the lining of the uterus called the endometrium. It is usually diagnosed in postmenopausal women and has increasingly become prevalent, with the reasons being obesity, diabetes, hormone imbalances, and improved longevity. Early diagnosis is critical for successful treatment and enhanced survival. However, current diagnostic procedures such as transvaginal ultrasound, hysteroscopy, and biopsy are often invasive, time-consuming, and subject to human interpretation, which can lead to variability in diagnostic outcomes [2].

In recent years, the combination of artificial intelligence (AI), in the form of machine learning (ML) and image processing methods, has unlocked new opportunities in early detection and classification of endometrial cancer. Through digital histopathological images and the use of stringent preprocessing methods—like noise removal, thresholding, image sharpening, and segmentation—scientists can improve image quality and relevant feature extraction for proper classification [2].

Convolutional Neural Networks (CNNs), a form of deep learning model, have gained a lot of popularity because they can automatically learn and extract spatial information from medical images. With transfer learning, CNNs can enhance model performance even when available data are limited in size, which is a common drawback in medical imaging research. In addition, methods like TNM staging make tumour severity grading and grouping possible according to size, nodal status, and metastasis, facilitating the planning of clinician treatment.

This questionnaire is designed to give a thorough description of the existing methodologies and developments in image preprocessing and machine learning to predict endometrial cancer as well as its grading. It analyses the technological environment, summarizes key studies, assesses current challenges, and maps directions for the future to make such AI-based systems more accurate, explainable, and deployable in real-time clinical settings.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14650

II. LITERATURE SURVEY

1. Histopathological Image Analysis for Endometrial Cancer Detection—Emily Davis & Robert Brown
This review highlights classical image processing methods combined with machine learning techniques to detect endometrial cancer. The authors discuss challenges such as image quality, staining variations, and observer variability. They emphasize the need for standardized preprocessing methods to improve data consistency. Although comprehensive, the paper does not present experimental outcomes but serves as a valuable synthesis of existing literature. It underscores the potential of integrating segmentation and feature extraction with AI models for robust cancer detection[1].

2. Transfer Learning-Based Endometrial Cancer Detection—Sarah Lee & David Wilson
This study explores the effectiveness of transfer learning, where pre-trained models such as ResNet and VGG16 are fine-tuned using limited histopathological datasets. The approach accelerates model convergence and improves classification accuracy. The paper demonstrates that transfer learning can outperform traditional methods, especially when data availability is a limiting factor. However, the performance of these models heavily depends on the similarity between the source and target domains[2].

3. Challenges and Future Directions in Histopathological Image Analysis—Richard Anderson & Jessica Taylor
This paper addresses the ongoing challenges in automated histopathological analysis for endometrial cancer, including limited datasets, lack of interpretability, and integration issues with clinical workflows. It also suggests future research directions such as explainable AI (XAI), federated learning for data privacy, and cross-institutional collaboration to build comprehensive image repositories. While the paper provides strategic insight, it lacks implementation details and experimental validation[3].

4. Detection of Nuclei Cells in Histopathological Images—Dr. Shoba R. Patil
Dr. Patil's work focuses on the segmentation of cell nuclei using methods like morphological operations, watershed transformation, and k-means clustering. The study emphasizes colour deconvolution in haematoxylin and eosin (H&E)-stained images to separate nuclei, cytoplasm, and lumen regions. The accurate extraction of nuclear patterns is critical for grading cancer severity. The paper presents a semi-automated pipeline that improves segmentation accuracy but requires manual intervention at certain stages[4].

5. Cancer Detection Using Image Processing and Machine Learning—Dr. Anita Dixit
This paper combines machine learning techniques like K-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM) with medical image data such as MRIs and thermographs. It highlights the significance of feature engineering in improving classification results. The study uses training and testing datasets, extracting features based on colour, shape, and structure. Although not specific to endometrial cancer, the methodologies are applicable and relevant for adaptation[5].

6. Automatic Segmentation of Endometrial Cancer on Ultrasound Images—Lidiya Lilly Thampi
This study targets ultrasound image analysis and proposes the use of SRAD (Speckle Reducing Anisotropic Diffusion) filtering combined with Otsu thresholding for segmenting lesion regions. The use of Partial Differential Equation (PDE)-based techniques improves image clarity and edge detection. Although primarily focused on ultrasound rather than histopathological images, the segmentation techniques are applicable to noisy datasets and offer promising results in non-invasive diagnostics[6].

III. OBJECTIVES

- To evaluate image preprocessing techniques: such as grayscale conversion, noise removal, thresholding, and segmentation that enhance histopathological images for better analysis of endometrial cancer [1].
- To analyse machine learning and deep learning models: like CNNs for accurate classification, grading, and TNM staging of endometrial cancer using extracted image features [5].
- To identify current challenges and propose future directions: for integrating AI-based diagnostic systems into clinical workflows, aiming to improve early detection, reduce variability, and support personalized treatment strategies [6].

IV. METHODOLOGY

The methodology for predicting endometrial cancer and its grade using image preprocessing and machine learning is structured into several essential phases. Each stage contributes to building a robust diagnostic pipeline capable of delivering accurate and interpretable results.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14650

1. Data Collection

The foundation of this project lies in the acquisition of high-quality histopathological images of the endometrium. These images are typically obtained from public datasets such as those on Kaggle, academic research repositories, and digitized biopsy samples. The dataset should include diverse images of varying grades and cancer stages to ensure model robustness. Images must be in a consistent format (e.g., PNG or JPEG) and labelled accurately with the corresponding cancer grade and TNM staging, if available[1].

2. Image Preprocessing

Raw medical images often contain artifacts, noise, and inconsistent lighting that can hinder analysis. Preprocessing techniques are employed to enhance the image quality and extract relevant features. The following steps are applied:

- Grayscale Conversion: Converts RGB images into grayscale to simplify analysis. Methods include the luminosity and averaging techniques which weight pixel intensity based on RGB components[5].
- Noise Reduction: Filters like the Median filter and Gaussian filter are used to smoothen the image and eliminate salt-and-pepper noise while preserving edge details[5].
- Thresholding: Techniques such as global thresholding and adaptive thresholding are used to binarize the image by separating the foreground (suspected cancerous areas) from the background[1][5].
- Image Sharpening: High-pass filters and Laplacian filters are applied to enhance edges and minute details, making the morphological structure of cells clearer[5].
- Segmentation: The image is segmented to isolate regions of interest (ROI). This may include edge-based segmentation (e.g., Canny Edge Detection) or region-based methods (e.g., Watershed Algorithm)[5].

3. Feature Extraction

Once images are pre-processed, key features must be extracted to feed into the machine learning model. These features can be broadly classified into:

- Colour Features: Used when RGB-based information is preserved, helpful for distinguishing tissue types.
- Shape Features: Captures the morphology of the cells such as area, perimeter, and irregularities.
- Texture Features: Histogram of Oriented Gradients (HOG) is often used to describe gradient orientation and intensity, offering detailed structural information of cellular patterns[3].

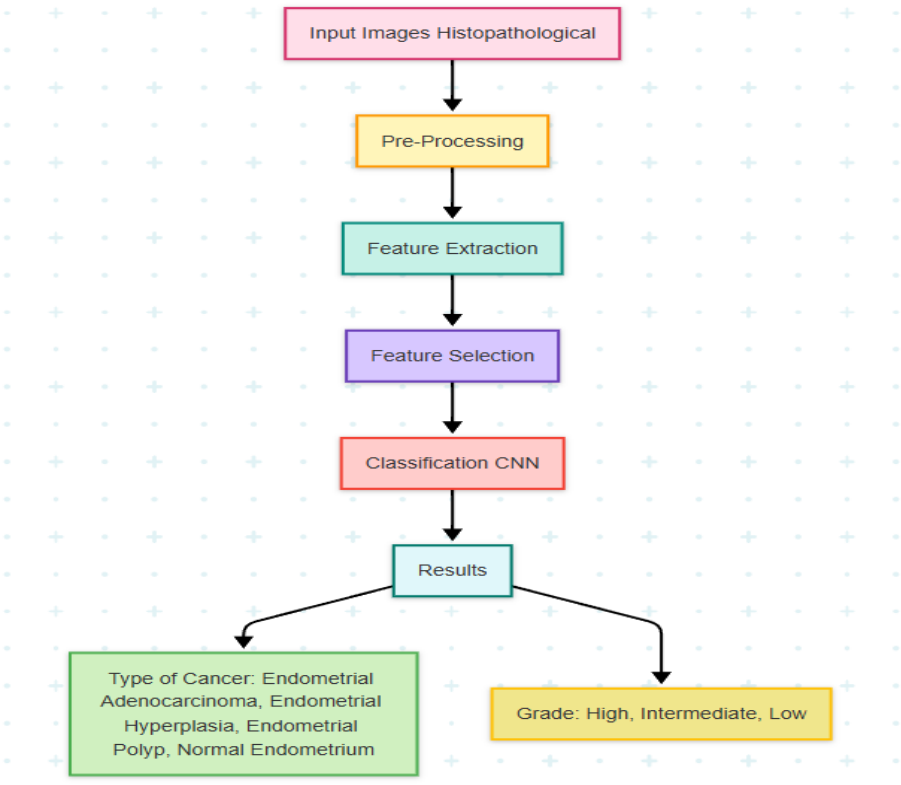

Fig.1 Architecture Design

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14650

4. Model Selection and Training

The classification phase involves selecting a suitable machine learning or deep learning model. Convolutional Neural Networks (CNNs) are the most widely used for image-based classification tasks due to their ability to automatically learn spatial hierarchies of features.

- Convolutional Layers: Extract low- to high-level features using kernel operations.
- Pooling Layers: Reduce dimensionality and help retain key features.
- Fully Connected Layer: Converts the 2D feature maps into a 1D feature vector for classification.
- Output Layer: Uses the SoftMax activation function to classify images into categories such as cancer grade (I, II, III) and Other models such as SSD (Single Shot Multibox Detector) and YOLO (You Only Look Once) may be explored for object detection and localization of tumours within histopathology slides[5].

5. Model Evaluation

Model performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and the confusion matrix. Validation is performed using a holdout dataset or through k-fold cross-validation. Receiver Operating Characteristic (ROC) curves and Area Under Curve (AUC) values may be used to assess the classifier's discriminative capability[2][5].

6. Grading

The last step involves grading the cancer based on histological differentiation and assigning stages by evaluating tumour size, nodal involvement, and presence of metastasis. The model outputs these classifications, which are then mapped to clinical interpretations[5].

7. Deployment and Interface

Make the system usable by clinicians, a graphical user interface (GUI) or web application is developed. Technologies like Tkinter (Python GUI), Flask (web server), or integration into hospital management systems may be used. This interface allows users to upload images, process them, and view predictions in real time[4].

V. APPLICATION REQUIREMENTS

Ensure the successful implementation and deployment of the endometrial cancer prediction system, the following **hardware and software requirements** must be met. These requirements are structured to support efficient image processing, deep learning model training, user interaction, and system scalability[5].

1. *Hardware Requirements*

- Processor: A multicore processor (e.g., Intel Core i5/i7 or AMD Ryzen 5/7) is essential to support real-time processing of histopathological images and machine learning tasks.
- RAM: A minimum of 8 GB is required for basic operations, while 16 GB or more is recommended for smooth execution of training, testing, and image enhancement algorithms.
- Storage: At least 512 GB of SSD storage is advised to store datasets, model checkpoints, logs, and user-uploaded images efficiently. SSDs enhance data read/write speed and overall application performance.
- GPU (Graphics Processing Unit): A CUDA-enabled NVIDIA GPU (e.g., GTX 1660, RTX 3060, or higher) is highly recommended for training deep learning models such as CNNs. GPU acceleration significantly reduces training time and improves inference speed.
- Display and Peripherals: A standard monitor with 1080p resolution, keyboard, and mouse for development and testing environments.

2. *Software Requirements*

➢ Operating System: The system is compatible with Windows 10/11, Ubuntu 20.04+, or macOS for development and deployment.

➢ Programming Language: Python 3.8+ is the core language used for developing preprocessing modules, CNN architecture, and GUI components.

➢ Development Tools:

- Visual Studio Code / PyCharm: IDEs for writing and debugging Python code.
- Jupiter Notebook: For experimenting with image preprocessing and ML models interactively.

➢ Libraries and Frameworks:

- OpenCV: For image loading, resizing, filtering, and visualization.
- NumPy / Pandas: For numerical computation and dataset handling.
- Matplotlib / Seaborn: For visualizing results, feature maps, and evaluation graphs.

DOI: 10.17148/IJARCCE.2025.14650

- TensorFlow / Keras or PyTorch: For developing and training CNN-based classification models.
- Scikit-learn: For feature selection and traditional ML algorithm support.

➢ Web/GUI Framework:

- Flask or Tkinter: To build a simple user interface for image uploading, result display, and model interaction.

➢ Cloud/Deployment (Optional):

- Google Colab: For free GPU access during model training.
- Firebase or AWS S3: To store and retrieve image data and model outputs securely.

VI. CONCLUSION

The classification and prediction of endometrial cancer based on image preprocessing and machine learning is an important contribution to computer-assisted diagnosis. This method seeks to overcome the shortcomings of conventional diagnostic techniques, which are invasive, time-consuming, and prone to human error. Through advanced preprocessing image methods like grayscale conversion, filtering out noises, thresholding, segmentation, and feature extraction, raw histopathology images can be converted into refined inputs for machine learning processing.[1]

Convolutional Neural Networks (CNNs) are most effective for classification of images with automated feature learning and high performance for identifying cancerous areas and tumour grade classification. With a good and diverse dataset, CNNs can differentiate among various endometrial conditions like adenocarcinoma, hyperplasia, and polyps, and can also tell us about the grade of the cancer—low, intermediate, or high.[2][3]

The combination of these methods in a single prediction pipeline provides several advantages. First, it minimizes the reliance on experienced pathologists for initial diagnosis, particularly in areas where medical resources are limited. Second, it increases diagnostic accuracy and reproducibility, which are essential for starting timely and adequate treatment. Finally, the ability to analyse in real-time using easy-to-use applications or cloud platforms allows for quicker decision-making in the clinical environment.[5]

Even with the positive outcomes, issues persist. The restricted availability of annotated histopathological datasets, the requirement for interpretability of models, and the challenge of integrating AI systems within their existing clinical workflows are hindrances that need to be overcome. However, with the ongoing advances in artificial intelligence, medical imaging, and cloud computing, these issues are likely to be countered in the course of time.[3]

In summary, this study demonstrates the revolutionizing power of integrating image preprocessing with machine learning for early and precise endometrial cancer prediction. Future research should aim to enhance dataset variety, model explainability, and interoperability with electronic health records. With further development, such systems would represent essential tools in contemporary gynaecological oncology, providing patients with accelerated diagnosis, personalized treatment plans, and better health outcomes.[1][5]

VII. ACKNOWLEDGEMENT

We would like to thank **Prof. Dr. Vijayalaxmi Mekali**. from the bottom of our hearts for the valuable and positive feedback provided under the project planning and development. We are extremely grateful for her donation of noble time. Additionally, we would like to thank the principal, professors of KSIT for their constant support and motivation.

REFERENCES

[1]. E. Davis and R. Brown, *"Histopathological Image Analysis for Endometrial Cancer Detection: A Review,"* in *International Journal of Biomedical Imaging*, vol. 2019, Article ID 1294571, 2019.

[2]. S. Lee and D. Wilson, *"Transfer Learning-Based Endometrial Cancer Detection in Histopathological Images,"* in *IEEE Access*, vol. 9, pp. 139405–139414, 2021.

[3]. R. Anderson and J. Taylor, *"Histopathological Image Analysis for Endometrial Cancer: Challenges and Future Directions,"* in *Computers in Biology and Medicine*, vol. 145, p. 105434, 2022.

[4]. S. R. Patil, *"Detection of Nuclei Cell in Histopathological Images of Uterine Cancer: Adenocarcinoma of Endometrium,"* in *Procedia Computer Science*, vol. 167, pp. 2391–2399, 2020.

[5]. A. Dixit, *"Cancer Detection using Image Processing and Machine Learning,"* in *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 5, pp. 2310–2315, June 2020.

[6]. L. L. Thampi, *"An Automatic Segmentation of Endometrial Cancer on Ultrasound Images,"* in *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 6, no. 2, pp. 456–462,21.