

Received 11 February 2025, accepted 6 April 2025, date of publication 15 April 2025, date of current version 30 April 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3561178

## RESEARCH ARTICLE

# Explainable Artificial Intelligence Driven Segmentation for Cervical Cancer Screening

NIRUTHIKKA SRITHARAN<sup>1</sup>, NISHAANTHINI GNANAVEL<sup>1</sup>, PRATHUSHAN INPARAJ<sup>1</sup>,  
DULANI MEEDENIYA<sup>1</sup>, (Senior Member, IEEE), AND  
PRATHEEPAN YOGARAJAH<sup>2</sup>, (Member, IEEE)

<sup>1</sup>Department of Computer Science and Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka

<sup>2</sup>School of Computing, Engineering and Intelligent Systems, Ulster University, BT48 7JL Londonderry, U.K.

Corresponding author: Pratheepan Yogarajah (p.yogarajah@ulster.ac.uk)

This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/Y002121/1.

**ABSTRACT** Cervical cancer remains an important global health challenge among women. Early and accurate identification of abnormal cervical cells is crucial for effective treatment and improved survival rates. This paper addresses the development of a novel weakly supervised segmentation framework that combines binary classification, Explainable Artificial Intelligence (XAI) techniques, and GraphCut to automate cervical cancer screening. Unlike traditional segmentation methods that rely on pixel-level annotations of medical images, which are costly, laborious, and require expertise in medical imaging, our approach leverages classification-driven insights to segment the nucleus, cytoplasm, and background regions. A key innovation of our framework is the use of XAI techniques such as Grad-CAM++ and LRP combined with GraphCut, to enable annotation-free segmentation using only classification-level labels. This represents a pioneering application of explainability techniques in the context of cervical cancer screening. Among the classification models explored, including fine-tuned variants of VGGNet and XceptionNet, VGG16-Adapted128 achieved the highest performance, marked by an accuracy of 0.94, precision of 0.94, recall of 0.94, and an F1 score of 0.94. This novel segmentation framework employed LRP and GradCAM++ as XAI techniques to gain insight into the decision-making process of classification models, with GradCAM++ demonstrating greater effectiveness. The performance of these XAI methods was assessed through both visual inspection and quantitative metrics, including entropy and pixel flipping. This innovative approach to segmentation is formally introduced through two algorithms detailed in this paper. The weakly supervised segmentation framework achieved a Dice Similarity Coefficient (DSC) of 62.05% and an Intersection over Union (IoU) of 61.89%. In addition, it has received high satisfaction ratings from expert evaluations and has been seamlessly integrated into a user-friendly Web application, offering clinicians a transparent and reliable tool to improve the precision of decision-making in the detection of cervical cancer. Although this work represents an early step, it lays a strong foundation for advancing XAI-driven, weakly supervised segmentation techniques in medical imaging, particularly in resource-constrained cervical cancer screening contexts.

**INDEX TERMS** Cervical cancer, explainable AI, image segmentation, weakly supervised learning.

## I. INTRODUCTION

Cervical cancer remains one of the most common cancers among women globally [1]. Despite its high mortality rate, cervical cancer is highly treatable when

detected early. Screening methods such as the Pap Smear Test and Liquid-Based Cytology Test have significantly reduced incidence rates over recent decades, underscoring the critical role of early detection in treatment success. However, traditional screening methods are time-consuming, [2], with a notable risk of false negatives. This highlights the urgent need for automated accurate

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Zuo<sup>1</sup>.

screening solutions to improve efficiency and diagnostic precision [3].

The detection of cervical cancer traditionally relies heavily on the cytological analysis of cell samples, where features such as the nucleus area and nucleus-to-cytoplasm ratio are key indicators of cancerous or precancerous cells. Cancerous cells typically exhibit a larger nucleus and a higher nucleus-to-cytoplasm ratio. However, manual analysis of Pap smear images is not only time-consuming but also susceptible to human error, emphasizing the necessity for automation in this field. Although automated segmentation of cervical cells promises to enhance accuracy and efficiency [2], it presents its own challenges, including the scarcity of annotated training data and the complexity of interpreting automated methods. These methods also struggle with overlapping cells, variable staining quality, and inconsistent image characteristics. Furthermore, they require extensive computational resources and often lack the robustness needed for diverse clinical samples, limiting their practical implementation in resource-constrained settings. Techniques like U-net and CNN-based methods show potential, but limitations such as redundancy in multiscale information and extended training times remain significant hurdles [4], [5].

The key contribution of this study lies in its pioneering use of XAI techniques for cervical cell segmentation, presenting a paradigm shift from traditional methods that heavily depend on labor-intensive, pixel-level precise annotated segmentation masks and manual analysis [6]. Unlike existing approaches predominantly based on U-Net and Mask R-CNN architectures, which require extensive annotated datasets, this study is the first to explore the combination of binary classification, XAI and GraphCut for this task. This novel approach eliminates the need for exhaustive manual annotations while providing interpretable insights into model decisions - a critical advantage in medical applications where understanding diagnostic reasoning is essential. By leveraging classification-driven insights to guide segmentation, our framework significantly reduces the annotation burden on medical experts while maintaining high segmentation accuracy. Existing segmentation techniques face critical limitations: supervised methods require prohibitively expensive pixel-level annotations that demand specialized medical expertise; unsupervised approaches often yield imprecise results requiring domain-specific refinements and suffer from parameter-sensitive initialization processes; and both approaches typically lack transparency in their decision-making, undermining clinical trust and adoption. Additionally, current methods struggle with cell boundary ambiguity, variable staining quality, and computational inefficiency when processing large cytological datasets.

This approach not only alleviates the burden of manual annotation but also provides transparency in model decision-making, addressing a critical gap in existing segmentation techniques. Transparency and interpretability are crucial attributes in the medical domain, where understanding the rationale behind diagnostic decisions is essential for

clinicians [8], [9], [10]. By providing insights into model's reasoning, XAI methods also enable error detection and model improvement, fostering trust and acceptance among healthcare practitioners [7]. This enhanced transparency not only supports improved diagnostic accuracy but also enhances the reliability and credibility of the system in clinical settings.

The research is guided by the following critical Research Questions (RQs) that highlight its novelty and practical relevance:

- **RQ1:** How can a supervised binary classification model be developed to accurately classify cervical cell images into malignant and benign categories with performance metrics that meet clinical standards for diagnostic support?
- **RQ2:** In what ways can XAI be applied to facilitate pixel-wise segmentation of malignant cells, ensuring transparency and interpretability that aligns with clinicians' diagnostic workflows?
- **RQ3:** How does the proposed weakly supervised approach, which combines classification, XAI and GraphCut, compare with traditional segmentation methods in terms of accuracy, efficiency, and reliability in real-world clinical scenarios?
- **RQ4:** How can the developed system be transformed into a practical diagnostic support tool for healthcare practitioners, enhancing diagnostic workflows and improving patient outcomes?

This study addresses the above RQs, laying the groundwork for a new line of research exploring the integration of binary classification, XAI, and GraphCut into a weakly supervised segmentation framework. By tackling critical challenges such as the scarcity of manually annotated ground truth segmentation data, the need for interpretability, and the demand for efficiency, this research marks a significant step forward in automated cervical cancer screening and AI-driven diagnostic systems.

## II. BACKGROUND AND RELATED STUDIES

### A. OVERVIEW OF CERVICAL CANCER

Cervical cancer is highly preventable and treatable when detected early, largely due to the extended precancerous phase that allows for timely intervention. During this phase, significant nuclear changes can be observed, such as nuclear enlargement, an increased nuclear-to-cytoplasmic (N/C) ratio, and irregularities in the nuclear membrane [11]. These alterations are crucial indicators of abnormal cell conditions, as illustrated in Fig. 1, which compares cervical cells in normal and abnormal states.

### B. CERVICAL CELL SEGMENTATION

Segmentation of nuclear and cytoplasmic regions in cervical cell images is essential for automated screening, as nuclear abnormalities are early indicators of infection. Accurate segmentation enables the calculation of metrics like cell

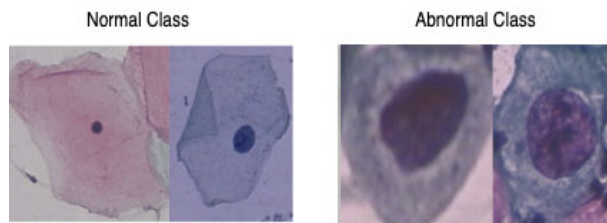


FIGURE 1. Cervical cell images.

diameter and nucleus-to-cytoplasm ratio, enhancing classification systems, especially in cases of uneven staining and overlapping cells [12].

Many studies employ direct segmentation with annotated masks, often using modified U-Net architectures. The vanilla U-Net is generally avoided due to issues like redundant information and lack of contextual awareness. Li et al. [12] improved U-Net by embedding global dependency, spatial, and channel attention, achieving over 75% in Zijbendos Similarity Index (ZSI), precision, and recall. Zhao et al. [13] proposed a multi-stage U-Net model, achieving a ZSI of 92.5%, precision of 90.1%, and recall of 96.8%. Chowdary and Yogarajah [14] embedded Residual and, Squeeze and Excitation modules in U-Net, resulting in a precision of 97.24%, recall of 96.20%, and a ZSI of 97.0%.

Mask R-CNN is another popular segmentation model, with variants achieving high precision and recall in both single and overlapping cells. For example, Mask R-CNN with a ResNet10-based Feature Pyramid Network achieved 92% precision, 91% recall, and 91% ZSI [15], while Rettenberger et al. [16] demonstrated its superiority over U-Net for instance segmentation of overlapping cells, even those that intersect.

Supervised pixel-wise segmentation methods demand extensive manual labeling, which is both time-consuming and prone to errors due to inconsistent annotations and unclear object boundaries. To overcome these issues, some studies have explored unsupervised segmentation approaches that eliminate the need for pixel-level labels.

Ragothaman et al. [17] used a Gaussian Mixture Model (GMM) for rough segmentation, achieving over 80% Dice Similarity Coefficient (DSC) with parameters estimated via the Expectation Maximization (EM) algorithm. Jeffree et al. [18] employed active contouring for unsupervised segmentation, while Gautam et al. [19] combined superpixel estimation with adaptive mean-shift and Simple Linear Iterative Clustering, yielding 85% precision, 88% recall, and 86% F1 score. Bandyopadhyay and Nasipuri [20] used k-means clustering for nucleus segmentation.

These unsupervised methods often require further refinement. Techniques like the Hough transform [17], intensity-weighted adaptive thresholding [19], and morphological operations are used to enhance segmentation accuracy. These refinement techniques tend to rely on a priori information and parameter-sensitive initialization processes, which can limit the effectiveness of these unsupervised approaches.

Given the limitations of both unsupervised and fully supervised segmentation methods [25], this study introduces a novel weakly supervised segmentation approach. Traditional methods either demand extensive pixel-level annotations or rely on purely unsupervised clustering techniques, which often require additional domain-specific refinements. In contrast, our approach integrates supervised binary classification of cervical cell images into normal and abnormal categories, leverages XAI techniques for model interpretation, and incorporates GraphCut within a weakly supervised segmentation framework. To the best of our knowledge, this is the first study to formally present two algorithms for implementing this framework. The remainder of this section II delves into existing cervical cell classification methods and explores applications of XAI in this context.

### C. CERVICAL CELL CLASSIFICATION

Certain studies have explored classification of cervical cell images using various feature extraction methods. For instance, Arya et al. [21] extracted multiple texture-based features using first-order histogram, Gray-Level Co-occurrence Matrix (GLCM), LBP, Laws, and Discrete Wavelet Transform (DWT). They trained Artificial Neural Networks (ANN) and Support Vector Machines (SVM) with these texture-based features, achieving impressive results: 99.50% accuracy, 99.00% sensitivity, and 99.00% specificity. In contrast, Bandyopadhyay and Nasipuri [20] focused on shape-based features, such as area, perimeter, eccentricity, circularity, and compactness derived from the nucleus contours. Using these features, they trained a Random Forest classifier, which attained a classification accuracy of 92.15%.

Several studies bypass the feature extraction step, relying on Deep Learning (DL) models to inherently extract features from cervical cell images, thus streamlining the classification. For instance, Sritharan et al. [22] enhanced the pretrained XceptionNet by adding additional layers and training the new layers, achieving results of 89% accuracy, 90% precision, 89% recall and 88% F1 score. Similarly, Gnanavel et al. [23] adopted a comparable approach by enhancing XceptionNet and implementing pretrained, fine-tuned models such as VGGNet and EfficientNet. All these models achieved accuracy, precision, recall, and F1 scores above 85%. Tan et al. [24] evaluated 13 pre-trained deep convolutional neural networks (CNN), and DenseNet-201 showed the highest accuracy of 87.02% for multi-class classification of cervical cell images. Fang et al. [26] combined a CNN for local feature extraction with a transformer for capturing global features. They used a feature fusion branch with Differential Feature Fusion (DIFF) blocks to enhance the performance. Moreover, Alquran et al. [27] passed cervical cell images through their novel Cervical Net and pretrained Shuffle Net models. They obtained feature vectors from the global average pooling layers, combined these features using Canonical Correlation Analysis, and then trained an SVM. They showed a 5-class classification accuracy of 99.10%. Furthermore, Rahaman et al. [4]

extracted features from the last layer before the softmax layer from 4 fine-tuned DL models: VGG16, VGG19, XceptionNet and ResNet50. These extracted feature vectors were then fed into a sequential Neural Network to perform 5-class classification that achieved 99.2% precision, 99.0% recall, 99.0% F1 score and 99.14% accuracy.

Additionally, some studies have employed a fusion of handcrafted features and abstract features generated in the hidden layers of CNNs. For instance, Jia et al. [28] extracted texture, morphology, and chroma-based features using GLCM, Fourier transformation, Gabor transformation, and Markov random field. They have fused these handcrafted features with those from the hidden layers of a trained LeNet-5 CNN model. The combined features were input into an SVM, achieving an accuracy of 99.3%, sensitivity of 98.9%, and specificity of 99.4%. Similarly, Chowdary and Yogarajah [14] generated 2 handcrafted descriptors using bag-of-features and linear-binary-patterns. These features were then fused with feature vectors extracted by VGG19, VGG-F and CaffeNet. The combined features were passed to a multi-layer perceptron model, which performed classification with 99.16% accuracy, 99.15% recall and 99.75% specificity. In this manner, by integrating these approaches, researchers have harnessed the strengths of both traditional feature extraction methods and advanced deep learning techniques to enhance the classification of cervical cell images.

#### D. EXPLAINABLE ARTIFICIAL INTELLIGENCE

There is significant motivation to apply Explainable Artificial Intelligence (XAI) to automated systems, particularly in healthcare settings such as cervical cancer diagnosis. XAI enhances the transparency and interpretability of opaque models, thereby fostering greater trust and acceptance among clinicians. Additionally, it aids in error detection, debugging, and overall model improvement.

Consequently, several studies have integrated XAI into automated classification systems in healthcare settings to address these needs. Among them, Pitroda et al. [29] compared the explanations produced for chest X-ray images' classifier by LRP (Layer-wise Relevance Propagation), LIME (Local Interpretable Model Agnostic Explanation), DTD (Deep Taylor Decomposition) and GB (Guided Backpropagation) quantitatively using image entropy and pixel-flipping metric. Patel et al. [30] applied various XAI techniques, including occlusion, saliency, integrated gradients, deconvolution, LRP, LIME, and Class Activation Mapping (CAM), to a lung X-ray image classifier. They quantitatively evaluated these techniques by obtaining Mean Opinion Scores from expert radiologists. Bhandari et al. [31] applied SHAP (Shapley additive explanation) and LIME to a lightweight CNN designed for CT image classification. Ong et al. [32] applied LIME and SHAP to elucidate and interpret their SqueezeNet model for lung X-ray image classification. They compared the explanations generated by both XAI techniques qualitatively through visual inspection.

However, studies specifically applying XAI in the cervical cancer domain remain limited. Civit-Masot et al. [33] utilized GradCAM to generate heat maps, highlighting the specific areas where the cervical cancer classifier focused. Gnanavel et al. [23] utilized GradCAM (Gradient Weighted Class Activation Mapping), GradCAM++, and LRP to investigate the interpretability of a VGG16-based classifier. They assessed the generated explanations both quantitatively, using mean image entropy and the pixel-flipping metric, and qualitatively. Sritharan et al. [22] generated an ensemble explanation using the median values of class activation maps from the existing CAM methods: GradCAM, GradCAM++, Score-CAM, LayerCAM and Eigen-CAM, to highlight the salient regions of the image pertaining to the classification model's decision. Their results demonstrated that the proposed EnsembleCAM outperformed each of the five individual CAM methods, particularly in terms of the pixel-flipping metric.

#### E. COMBINING CLASSIFICATION AND XAI

Segmentation scenarios that integrate both classification and XAI are rare [2], [34], and even in those few instances, the utilization of this integration differs from our proposed framework.

Kaur et al. [34] introduced a model called 'GradXcepUNet,' which combines the Xception classification network, the U-Net segmentation model, and the Grad-CAM XAI technique. The process begins with feeding original 2D images and ground truth segmentation masks, labeled as "original" and "liver\_only," into the Xception classification model. Grad-CAM is then used to highlight crucial regions within the liver\_only class images based on the classification model's outputs. These Grad-CAM-generated saliency maps, along with the original images, are subsequently fed into the U-Net model for liver segmentation. It is important to note that though GradXcepUNet demonstrates an integration of classification and XAI, it still relies on manually annotated ground truth segmentation masks during the classification phase.

However, Seibold et al. [2] advanced a method that integrates classification with XAI to perform segmentation without using ground truth segmentation masks. Instead of traditional segmentation masks, they utilized images of magnetic tiles and sewer pipe surfaces with damages and without any damages to segment the damages areas in these surfaces. They used VGG for binary classification, paired with Layer-wise Relevance Propagation (LRP). Additionally, they employed Simple, Gaussian, and Beta Mixture Models to generate segmentation maps from relevance distributions provided by LRP.

### III. METHODOLOGY

#### A. PROCESS OVERVIEW

The cervical cell images used as input were sourced from the Herlev Dataset [35]. We then applied data preprocessing and augmentation techniques to improve the dataset.



A supervised binary classification model was developed to differentiate between “normal” and “abnormal” cervical cells. Various deep learning architectures, including VGGnet and XceptionNet, were explored, and the models’ effectiveness was evaluated using standard metrics such as accuracy, precision, recall, and F1 score on separate validation and test datasets. Following model evaluation, XAI techniques like GradCAM++ and LRP were used to identify the pixels contributing to the classification decision. The performance of these XAI methods was assessed using image entropy and the pixel flipping performance metric. Relevance grouping techniques, such as Simple Thresholding, K-means clustering, Gaussian Mixture Model (GMM), and the Inter-means Algorithm, were used to create segmentation masks. These masks were further refined with the GraphCut algorithm to produce more precise pixel-wise segmentation masks. The generated segmentation masks were evaluated using metrics like Intersection over Union (IoU) score and Dice Similarity Coefficient (DSC), comparing them against the ground truth masks from the Herlev Dataset. An overview of our proposed methodology is shown in Fig. 2.

## B. DATA MATERIALS AND PREPROCESSING

This research utilizes the cervical cell images from the Herlev dataset. It originally consists of seven distinct classes. However, considering the scope of this research, we reclassified the dataset into 2 classes: “normal” (images without cancer cells) and “abnormal” (images with cancer cells). Our dataset consists of 917 cell images, with 242 images belonging to the “normal” class and 675 images belonging to the “abnormal” class.

At first, all images were resized to a uniform size of  $224 \times 224$  pixels. Following this, noise reduction and image enhancements were implemented due to the inherent noise and low contrast in the Pap smear images. Noise reduction was achieved through the application of a median filter, while histogram equalization and normalization were employed to enhance contrast. The increased contrast facilitated the extraction of important information from the images and Fig. 3 shows our data preprocessing process.

Eight distinct sets of data augmentation procedures were developed for this study. For the abnormal training images, one augmentation was randomly selected from each set and applied, resulting in an eightfold increase in their number. Similarly, for the normal training images, three augmentations from each set were randomly chosen and applied, leading to a twenty-fourfold increase. This approach ensured a balanced and diverse augmented dataset for training.

The first set of augmentations included translation, scaling, and cropping a random portion of the image, which was then resized to the original image’s dimensions. The second set focused on rotation transformations. Subsequent sets incorporated various techniques, such as distortions, Contrast Limited Adaptive Histogram Equalization (CLAHE) to manage noise and contrast, gamma adjustments for brightness,

addition of Gaussian noise or blur, color adjustments, and image sharpening. Examples of these augmentation techniques are illustrated in Fig. 4.

## C. CLASSIFICATION MODELS

### 1) XCEPTIONNET

For the cervical cell classification model, XceptionNet was selected due to its depthwise separable convolutions, which significantly reduce the number of parameters compared to traditional convolutional layers, leading to more efficient computations and enhanced performance on image classification tasks. XceptionNet was used as a feature extractor with pre-trained weights from the ImageNet dataset. The model included a Global Average Pooling layer to reduce spatial dimensions, followed by Batch Normalization to improve training stability. A densely connected layer with ReLU activation and regularization was then introduced to prevent overfitting. Two versions of the XceptionNet model were developed: one with 256 units in the Dense layer and another with 1024 units. A dropout layer with a rate of 0.45 was employed for further regularization. The final layer was a dense softmax layer with the number of output classes. The base layers of XceptionNet were frozen to retain the pre-trained knowledge during fine-tuning. The model was compiled using the Adam optimizer and categorical cross entropy as the loss function. Key training hyperparameters included a batch size of 16 and 20 training epochs. Fig. 5 shows the model architectures with 1024 dense layer units.

The key reasons for choosing XceptionNet is its utilization of Depthwise separable convolutions and its fewer parameters compared to traditional architectures, allowing for quicker training and inference. Additionally, leveraging a pre-trained, well-established architecture like XceptionNet ensures robustness even with relatively limited cervical image datasets.

### 2) VGGNET

After reviewing existing research and testing various configurations, two fine-tuned versions of the VGG16 model were developed for cervical cell classification. Empirical testing revealed that an optimal input size for images was  $128 \times 128$ . As an image progresses through the network, the feature dimensions before the average pooling layer are  $4 \times 4 \times 512$ . To maintain these features without distortion by the average pooling function, the average pooling layer was adjusted to use a  $4 \times 4$  filter. Consequently, the first fully connected layer was configured to have 8192 neurons. These modifications constitute the first version of the enhanced VGGNet, referred to as “VGG16-Adapted128”.

The second version, named “VGG16-Adapted128-Simplified,” involved a further simplification of the model by removing the last convolutional block. This change resulted in feature dimensions of  $8 \times 8 \times 512$  at the average pooling layer. To accommodate these dimensions without feature distortion, the average pooling layer was adjusted to an

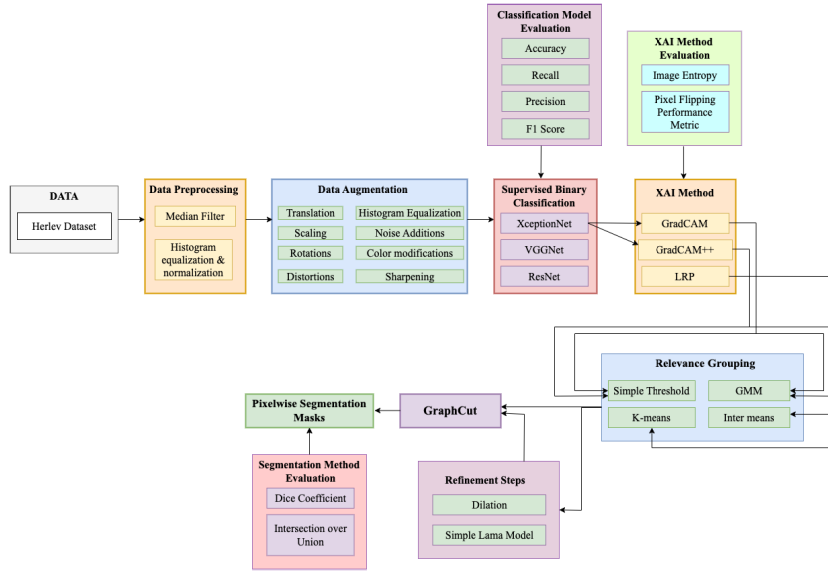


FIGURE 2. Process overview diagram.

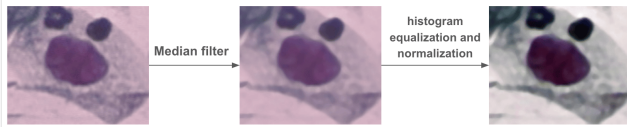


FIGURE 3. Data preprocessing.

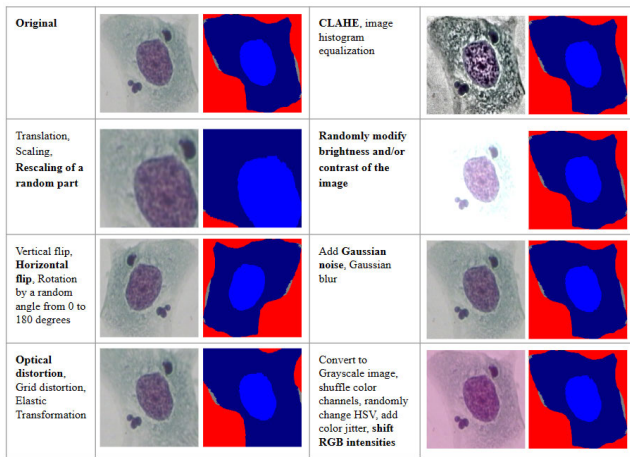


FIGURE 4. Data augmentation.

$8 \times 8$  filter. The dense layer was then configured with 32768 neurons. These streamlined architectures are shown in Fig. 6.

Both versions of the model were trained using cross-entropy loss as the loss criterion and optimized with Stochastic Gradient Descent (SGD). The learning rate was initially set at 0.001 and was reduced by a factor of 0.1 every 7 epochs. Training was scheduled for 50 epochs, but early stopping was triggered based on validation performance to prevent overfitting.

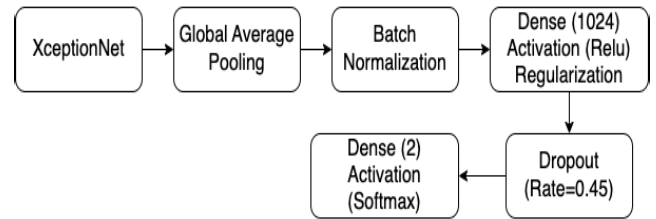


FIGURE 5. XceptionNet model with 1024 dense layer units.

Utilizing a VGG16-based architecture offers several advantages for our study. A significant benefit is the ease with which LRP can be applied, thanks to VGG16's simple, linear structure and the absence of skip connections. This straightforward architecture facilitates a more intuitive analysis of relevance scores throughout the network, aligning well with the objectives of our study.

#### D. EXPLAINABLE ARTIFICIAL INTELLIGENCE METHODS

##### 1) GRADCAM++

GradCAM++ enhances the GradCAM method by incorporating higher-order derivatives, allowing it to localize multiple relevant regions within an image with greater precision. The theoretical justification lies in the use of first-order and second-order derivatives of the target class score with respect to the activations of the final convolutional layer. The first derivatives was used to weigh the importance of activation maps, whereas the second derivatives improved this weighting. By combining these gradients, GradCAM++ provides a better understanding of the model's reliance on various features, ensuring that the explanations align with the learned representations of the XceptionNet model.

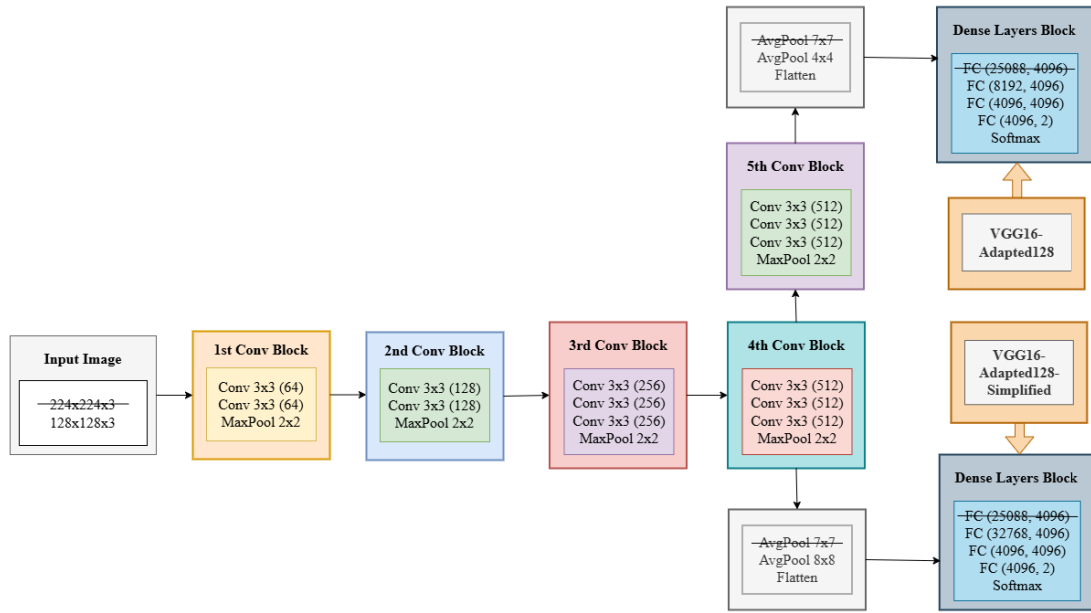


FIGURE 6. VGGNet architectures.

The above methodology can be mathematically described as follows, let  $\frac{\partial L}{\partial A_k}$  be the first-order gradients of the target class score  $L$  with respect to the activation maps  $A_k$  of the final convolutional layer and  $\frac{\partial^2 L}{\partial A_k^2}$  be the second-order gradients representing the curvature of the score with respect to  $A_k$ . First, to aggregate the spatial contributions of the gradients, the global sum is calculated as:

$$G_k = \sum_{i,j} \frac{\partial L}{\partial A_k^{ij}}. \quad (1)$$

GradCAM++ introduces the alpha coefficients to weigh spatial locations within the activation maps. These coefficients are defined as:

$$\alpha_k^{ij} = \frac{\frac{\partial^2 L}{\partial A_k^2} \cdot G_k}{2 \cdot \frac{\partial^2 L}{\partial A_k^2} + G_k \cdot \frac{\partial L}{\partial A_k} + \epsilon}, \quad (2)$$

where  $\epsilon$  is a small constant to prevent division by zero. Then the alpha coefficients are normalized for each feature map as:

$$\alpha_k^{ij} \leftarrow \frac{\alpha_k^{ij}}{\sum_{i,j} \alpha_k^{ij} + \epsilon}. \quad (3)$$

The deep linearization weights are computed by combining the above obtained normalized alphas and the ReLU activated gradients:

$$w_k = \sum_{i,j} \alpha_k^{ij} \cdot \text{ReLU}\left(\frac{\partial L}{\partial A_k^{ij}}\right). \quad (4)$$

Finally, The final GradCAM++ map is obtained by taking a weighted sum of the activation maps as:

$$\text{CAM}^{ij} = \sum_k w_k \cdot A_k^{ij}. \quad (5)$$

The map is also passed through a ReLU function to retain positive attributions and normalized for visualization:

$$\text{CAM} = \frac{\text{ReLU}(\text{CAM})}{\max(\text{CAM})}. \quad (6)$$

The heatmaps generated by GradCAM++ are constructed by combining the activation maps with weights and alphas derived from normalized gradients, visually highlighting the areas most influential in the prediction of the model. These heatmaps are theoretically grounded because they represent the model's feature importance and decision making processes in an interpretable way.

GradCAM++ was specifically chosen for this study due to its ability to generate fine-grained and class-discriminative explanation maps, which is particularly critical in medical imaging tasks such as cervical cancer screening. Unlike traditional saliency-based methods, GradCAM++ is more effective in highlighting multiple relevant features within complex biological structures such as the nucleus and cytoplasm of cervical cells.

## 2) LAYER-WISE RELEVANCE PROPAGATION

Layer-wise Relevance Propagation (LRP) [36] operates by tracing the relevance score of the output back through the network layers to the input pixels, thereby attributing importance to each pixel based on its contribution to the final decision. It calculates relevance scores for each neuron in every layer by evaluating the network's weights and

**TABLE 1.** Formulae for LRP rules [37].

LRP Rule	Formula
LRP-0	$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} R_k$
LRP- $\epsilon$	$R_j = \sum_k \frac{a_j w_{jk}}{\epsilon + \sum_{0,j} a_j w_{jk}} R_k$
LRP- $\gamma$	$R_j = \sum_k \frac{a_j (w_{jk} + \gamma w_{jk}^+)}{\sum_{0,j} a_j (w_{jk} + \gamma w_{jk}^+)} R_k$
$z^B$ -rule	$R_i = \sum_j \frac{x_i w_{ij} - l_i w_{ij}^+ - h_i w_{ij}^-}{\sum_i x_i w_{ij} - l_i w_{ij}^+ - h_i w_{ij}^-} R_j$

activations. This reverse propagation is governed by a set of specific propagation rules. The equations of some of these rules employed in our study are demonstrated in Table 1.

In the simplest LRP-0 rule,

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} R_k \quad (7)$$

$j$  and  $k$  are two neurons of any consecutive layers.  $a$  denotes the activation of the respective neuron, and  $w$  is the weight between the two neurons. The numerator of the fraction is the amount to which the neuron  $j$  influences the neuron  $k$ , and this is divided by the sum of contributions of all neurons of the lower layer. The outer sum over  $k$  means that the relevance of neuron  $j$  is determined by the sum of its influence on all neurons  $k$  of the following layer, times the relevance of these neurons.

In LRP- $\gamma$  and  $z^B$ -rule,  $w^+$  and  $w^-$  indicate non-negative and non-positive weights respectively. The first layer receives pixels, rather than ReLU activations, as input. As a result, in  $z^B$ -rule, which is suitable for pixels,  $l_i$  and  $h_i$  define box constraints of the input domain [37].

Using a single LRP rule to propagate relevance values across all layers of a neural network can have limitations. For example, LRP-0 often results in overly complex explanations with numerous local artifacts. On the other hand, while LRP- $\epsilon$  can provide a faithful explanation with a limited number of features, its outputs are frequently too sparse. To address these issues, we developed a composite rule set, “LRP Ruleset1,” by combining and adapting various LRP propagation rules as suggested by Montavon et al. [37]. This approach aims to balance the detail and clarity of the explanations.

Additionally, we utilized another rule set, “LRP Ruleset2,” which primarily employs the LRP- $\gamma$  rule. This choice was made because Uniform LRP- $\gamma$  tends to be more interpretable, offering a clearer and more coherent highlight of relevant features [37]. The specific propagation rules and the layers to which they were applied are detailed in Table 2.

LRP was selected for this study due to its notable advantages. It is computationally efficient, requiring only a single backward pass, and is supported by solid theoretical foundations and mathematical principles. Additionally, LRP’s longstanding tradition and widespread use underscore

**TABLE 2.** LRP rules for VGG.

Modified VGG16	Changed layers	LRP Rule-set1	LRP Rule-set2
conv1_1 conv1_2 maxpool	avgpool	$z^B$ rule LRP $\gamma$ LRP $\gamma$	$z^B$ rule LRP $\gamma$ LRP $\gamma$
conv2_1 conv2_2 maxpool	avgpool	LRP $\gamma$ LRP $\gamma$ LRP $\gamma$	LRP $\gamma$ LRP $\gamma$ LRP $\gamma$
conv3_1 conv3_2 conv3_3 maxpool	avgpool	LRP $\gamma$ LRP $\gamma$ LRP $\gamma$ LRP $\gamma$	LRP $\gamma$ LRP $\gamma$ LRP $\gamma$ LRP $\gamma$
conv4_1 conv4_2 conv4_3 maxpool	avgpool	LRP $\epsilon$ LRP $\epsilon$ LRP $\epsilon$ LRP $\epsilon$	LRP $\gamma$ LRP $\gamma$ LRP $\gamma$ LRP $\gamma$
conv5_1 conv5_2 conv5_3 maxpool	avgpool	LRP $\epsilon$ LRP $\epsilon$ LRP $\epsilon$ LRP $\epsilon$	LRP $\gamma$ LRP $\gamma$ LRP $\gamma$ LRP $\gamma$
avgpool 4x4 flatten	Not considered for LRP		
FC(8192, 4096) FC(4096, 4096) FC(4096, 2)	Conv 4x4 Conv 1x1 Conv 1x1	LRP 0 LRP 0 LRP 0	LRP $\gamma$ LRP $\gamma$ LRP $\gamma$

its reliability [38]. Its modular nature also allows for the application of various rules tailored to different layers, further enhancing its versatility.

## E. RELEVANCE GROUPING METHODS

The application of XAI on neural networks, as demonstrated in the methodology thus far, is relatively common. However, the real innovation lies in the next step. How can the insights from a model that classifies cervical cell images as cancerous or non-cancerous be used to perform cervical cell segmentation? To the best of our knowledge, this study is the first to explore this novel approach. Our findings suggest that the relevance of each pixel in the input image to the model’s decision can potentially indicate the region (nucleus, cytoplasm, or background) to which the pixel belongs. This stems from the fact that the nucleus and cytoplasm are key indicators of whether a cell is malignant. Consequently, we required a mechanism to group pixels based on their relevance scores obtained through XAI. To achieve this, we experimented with four different methods, which are detailed below.

- **Simple Thresholding:** This is a method that assigns pixel values to specific areas based on set thresholds [2]. This method was applied to heatmaps generated from GradCAM++ to group relevance values in the activation maps and create a segmentation map. Pixels that exceeded a high threshold were marked in blue as part of the nucleus, while those above a lower threshold but below the high threshold were identified as cytoplasm and shown in dark blue. Pixels below the low threshold were classified as background and shown in red.
- **Gaussian Mixture Model (GMM):** This is another technique used to cluster relevance values from XAI



methods [17]. GMM estimates the probability of each pixel belonging to different components—such as nucleus, cytoplasm, and background—and assigns each pixel to the component with the highest likelihood. This process produces a pixel-wise segmentation map, where each pixel is distinctly categorized according to its most likely cluster.

- **K-means:** Next we explored the effectiveness of the K-Means clustering method [20]. The relevance scores from the XAI methods were input into the K-Means algorithm, with  $k = 3$  to identify and segment three distinct regions: cytoplasm, nucleus, and background.
- **Inter-means:** Our final approach to image segmentation involved the Inter-means segmentation method, a threshold-based technique designed to categorize relevance scores into the same 3 distinct groups [2]. The process was iterative and revolved around determining two thresholds that effectively segment the relevance scores. To initiate this process, we selected the 1<sup>st</sup> and 3<sup>rd</sup> quartiles as the initial threshold values. Subsequently, these thresholds were employed to classify relevance values into the three aforementioned groups. The mean relevance score for each group was then computed, leading to determine two new thresholds as the averages between pairs of consecutive mean values, specifically  $(\text{mean1} + \text{mean2})/2$  and  $(\text{mean2} + \text{mean3})/2$ . This iterative procedure continued until the thresholds in successive iterations exhibited minimal changes, ensuring stability in the segmentation outcomes.

#### F. GRAPHCUT ALGORITHM

However, grouping the relevance values alone did not yield clear segmentation maps. Further refinement was necessary, which led us to integrate GraphCut into our novel framework. Renowned for its iterative refinement capabilities, GraphCut enabled the progressive enhancement of segmentation results. The process begins with approximate segmentation masks generated through the relevance grouping methods discussed above in III-E, which are then iteratively refined to achieve more precise and accurate segmentation.


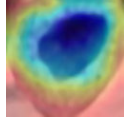


Specifically, we used GrabCut from the OpenCV-Python library for this refinement. GrabCut is particularly suited for complex biomedical images, such as cervical cell images, as it can dynamically adapt to the content of the image. Its iterative approach ensures continuous improvement in delineating cell boundaries, thereby enhancing segmentation accuracy and making it a robust choice for our application.

#### G. GENERATING PIXEL WISE SEGMENTATION MASKS

##### 1) CAM-BASED ALGORITHM

As the heatmaps from GradCAM++ effectively highlighted areas within both the nucleus and cytoplasm, it was able to create pixel-wise segmentation maps that allowed for the separate segmentation of the nucleus, cytoplasm, and background using relevance grouping techniques. Two different relevance grouping techniques were employed for

**TABLE 3.** Clustering GradCAM++ Heatmap values into 3 groups.

Original Image	CAM Heatmap	Simple Thresholding	GMM
			

the heatmaps from GradCAM++: the Simple Thresholding method and the Gaussian Mixture Model, as shown in Table 3. Among these two techniques, the Gaussian Mixture Model produced better segmentation maps than simple thresholding as the threshold setting in Simple Thresholding method is a manual process, it did not perform well across all cervical cell images and often failed to accurately distinguish the nucleus and cytoplasm boundaries. In contrast, the Gaussian Mixture Model was more effective in identifying the nucleus and cytoplasm boundaries.

Next, the GraphCut algorithm was applied to refine and enhance the accuracy of the segmentation maps. During the implementation of the GraphCut algorithm, it was applied separately to the nucleus and cytoplasm masks, resulting in refined nuclear and cytoplasmic segmentation maps. These refined maps were then combined to produce a comprehensive pixel-wise segmentation map with improved clarity. The full CAM based algorithm is detailed in Algorithm 1 and Fig. 7 illustrates the pipeline for our newly proposed CAM-based algorithm.

#### Algorithm 1 CAM-Based Segmentation Algorithm

**Require:** A set of images to be segmented

**Ensure:** Segmented images with masks for nucleus and cytoplasm

```

1: segmentation_masks ← []
2: for all image in images do
3:   cam_heatmap ←
   gradcamplusplus(fin tuned_XceptionNet, image)
4:   nucleus_mask ←
   separateNucleus(approx_segmentation_mask)
5:   refined_nucleus_mask ← graphCut(image,
   nucleus_mask)
6:   cytoplasm_mask ←
   separateCytoplasm(approx_segmentation_mask)
7:   refined_cytoplasm_mask ← graphCut(image,
   cytoplasm_mask)
8:   segmentation_mask ←
   combine(refined_nucleus_mask,
   refined_cytoplasm_mask)
9:   segmentation_masks ← segmentation_masks +
   [segmentation_mask]
10: end for
11: return segmentation_masks

```

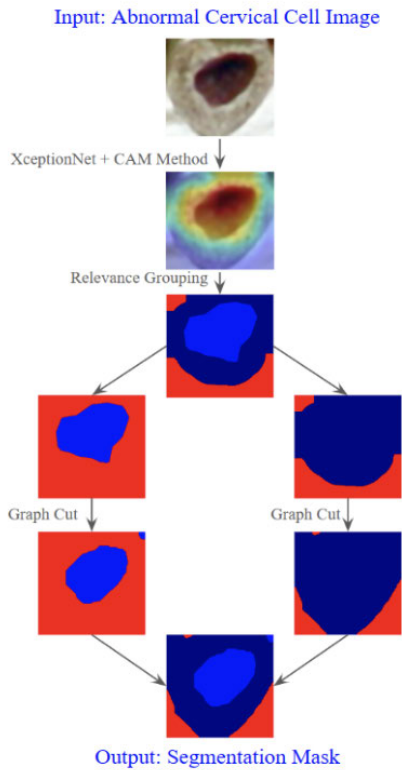


FIGURE 7. Pipeline for our newly proposed CAM based algorithm.

2) LRP-BASED ALGORITHM

CAM methods are capable of highlighting the entire areas of the nucleus and cytoplasm. In contrast, LRP methods typically provide only an approximate boundary around the salient areas and often emphasize only the nucleus region in cervical cell images. Consequently, when attempting to cluster LRP relevance values into the three categories: nucleus, cytoplasm, and background, the results were suboptimal, as shown in Table 4. The cytoplasm boundary was not detected, and the nucleus region was erroneously divided into two separate clusters.

TABLE 4. Clustering LRP relevance values into 3 groups.

Original Image	LRP Heatmap	Simple Thresholding	K-Means Clustering	GMM	Inter-means Segmentation

However, when the relevance values were logarithmically scaled, clustering into three distinct groups became feasible. This improved clustering is illustrated in Table 5.

TABLE 5. Clustering logarithmically scaled LRP relevance values into 3 groups.

Original Image	LRP Heatmap	Logarithmically Scaled Relevance Grouping

The logarithmic scaling was applied using the formula provided in Formula 8, with a scale factor of 1000.

$$\frac{\log_e(relevance * scalefactor + 1)}{\log_e(scalefactor)} \tag{8}$$

Initially, the relevance values for nucleus pixels were so high compared to those for all other pixels that the clustering algorithm struggled to differentiate between cytoplasm and background pixels. However, by applying logarithmic scaling, the larger relevance values (representing the nucleus) were compressed more significantly than the smaller values. This scaling reduced the large gap between nucleus pixel values and the rest, bringing the nucleus values closer to those of the other pixels. As a result, the clustering algorithm could more effectively distinguish between the nucleus, cytoplasm, and background regions.

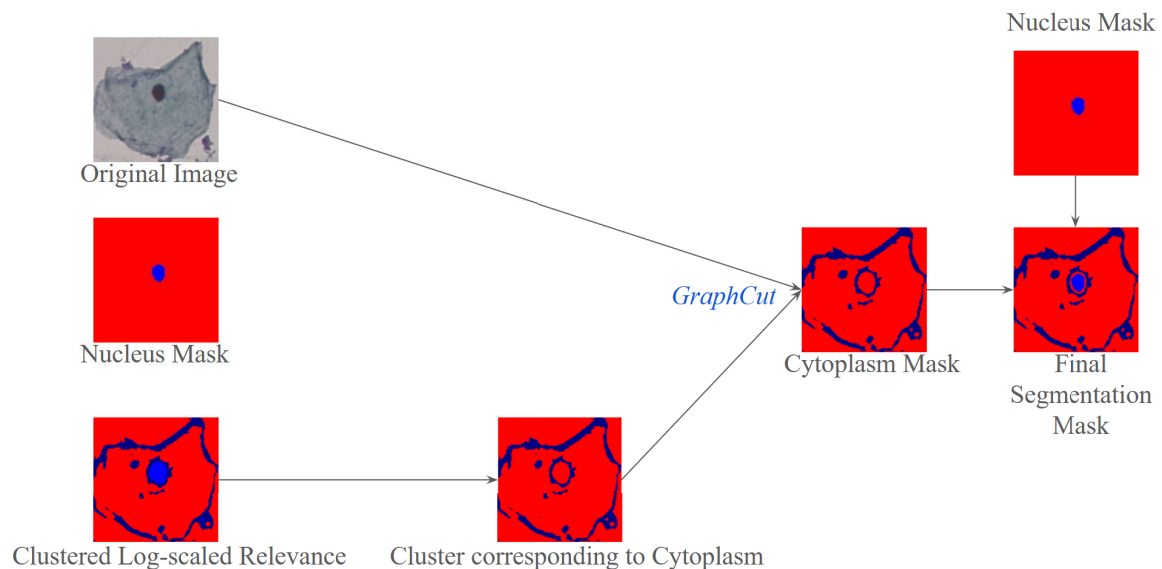
Next, we fed the clustered relevance values into the GraphCut algorithm, along with the original image, to assess its ability to detect the cytoplasm. Unfortunately, the performance was poor, as shown in Fig. 8.

To address this issue, we first removed the nucleus from both the original image and the 3-class clustering using dilation and the Simple LAMA Model. We then applied these modified images to the Graph Cut algorithm, which produced significantly improved results, as shown in Fig. 9. Subsequently, we combined the nucleus mask with the newly obtained cytoplasm mask, resulting in much more accurate segmentation.

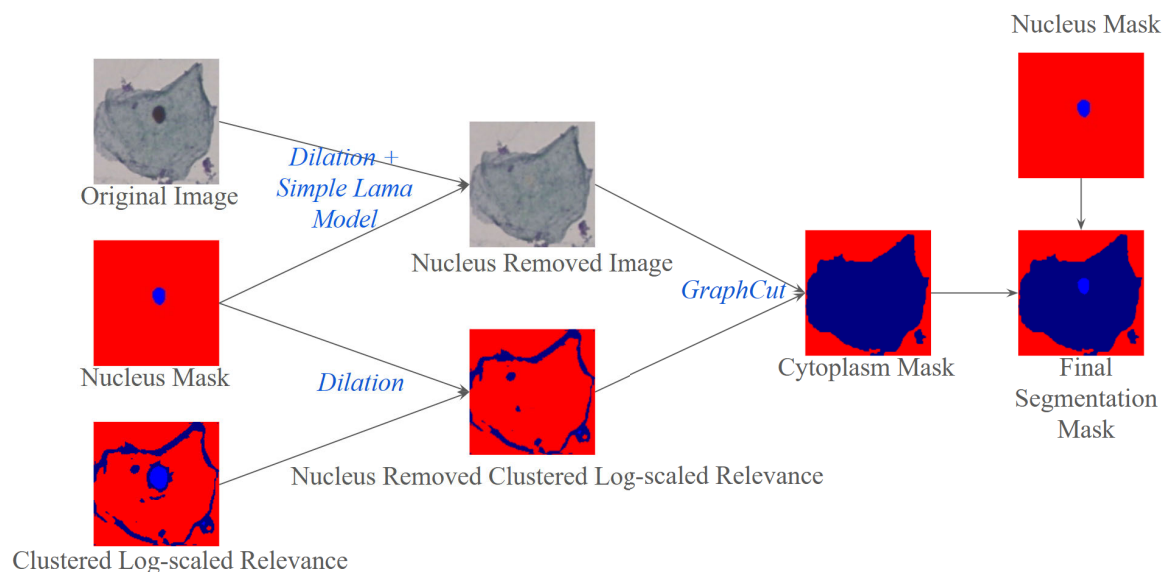
Our newly proposed LRP-based algorithm is detailed in Algorithm 2 and visually represented in Fig. 10.

H. SEGXPERS: WEB APPLICATION

To make our two novel segmentation algorithms accessible to medical professionals, we developed a web application called SegXperts [39]. By incorporating these advanced algorithms, SegXperts is designed to deliver precise and detailed segmentation masks, ultimately enhancing the analytical capabilities available to end users. The system processes input images to create pixel-wise masks that classify regions into nucleus, cytoplasm, and background. Key features include the generation of XAI heatmaps to assist in interpreting classification decisions, and a tabulated summary of cell descriptors such as nucleus and cytoplasm areas and their ratio. The user-friendly interface offers various pages including a Landing Page for introduction and feature



**FIGURE 8.** Workflow of deriving segmentation mask using LRP.



**FIGURE 9.** Improved workflow of deriving segmentation mask using LRP.

highlights, a User Manual Page with detailed instructions, a Configurations Page for setting parameters, a Results Page for viewing and downloading results, and an About Us Page for team and research information.

### I. COMPUTATIONAL EFFICIENCY

The deep learning classification models underlying our two novel algorithms in this study were trained on Google Colab using frameworks such as PyTorch, TensorFlow, and Keras. These models were fully trained on Google Colab's free-tier T4 GPU within a single session. The trained VGGNet model has a size of 512 MB, while the XceptionNet model

is 104 MB. With the T4 GPU, the entire CAM-based or LRP-based segmentation pipelines process an input cervical cell image in just a few seconds to generate the corresponding segmentation mask.

This efficiency enabled the deployment of both algorithms in our novel weakly supervised segmentation framework on a web application, leveraging free-tier resources. The frontend of the application is built using ReactJS to ensure an interactive and user-friendly interface. The backend services are hosted on HuggingFace's CPU Basic free-tier server, while the frontend is deployed via Netlify. This lightweight, cost-effective setup demonstrates the accessibility and

Algorithm 2 LRP-Based Segmentation Algorithm

**Require:** A set of images to be segmented  
**Ensure:** Segmented images with nucleus and cytoplasm masks

```
1: segmentation_masks ← []
2: for all image in images do
3:   lrp_heatmap ← lrp(finetuned_VGG16, image)
4:   nucleus_boundary ←
     twoClassClustering(lrp_heatmap)
5:   nucleus_mask ← graphCut(nucleus_boundary,
     image)
6:   log_scaled_heatmap ← logScaling(lrp_heatmap)
7:   approx_segmentation_mask ←
     threeClassClustering(log_scaled_heatmap)
8:   cytoplasm_boundary ←
     removeNucleus(approx_segmentation_mask)
9:   cytoplasm_boundary ←
     dilation(cytoplasm_boundary, dilated_nucleus_mask)
10:  nucleus_removed_image ←
     simpleLamaModel(image, dilated_nucleus_mask)
11:  cytoplasm_mask ← graphCut(cytoplasm_boundary,
     nucleus_removed_image)
12:  segmentation_mask ← combine(nucleus_mask,
     cytoplasm_mask)
13:  segmentation_masks ← segmentation_masks +
     [segmentation_mask]
14: end for
15: return segmentation_masks
```

practicality of “SegXperts” for clinical settings, even with minimal computational resources. Fig. 11 shows the system architecture of our “SegXperts” application.

IV. RESULTS ANALYSIS

A. PERFORMANCE OF CLASSIFICATION MODELS

The performance metrics for all classification models are summarized in Table 6, and the confusion matrices are depicted in Fig. 12. Among the two XceptionNet models, the one with 1024 dense layer units slightly outperformed the other, achieving perfect recall for abnormal cells, which indicates that it successfully identified all instances of abnormalities. However, both VGGNet versions demonstrated superior overall performance. Although VGGNet-Adapted128 showed marginally better performance compared to VGGNet-Adapted128-Simplified, the latter achieved a perfect recall of 1.00 for the abnormal class. Given the high stakes of misclassifying abnormal cells as benign, this perfect recall is highly valued. Consequently, the CAM-based algorithm was implemented with XceptionNet (1024 units), while the LRP-based algorithm was applied using VGGNet-Adapted128-Simplified.

At first glance, these metrics might appear inferior to some of the results presented in the related studies section II-C. However, it is important to note that the studies reporting

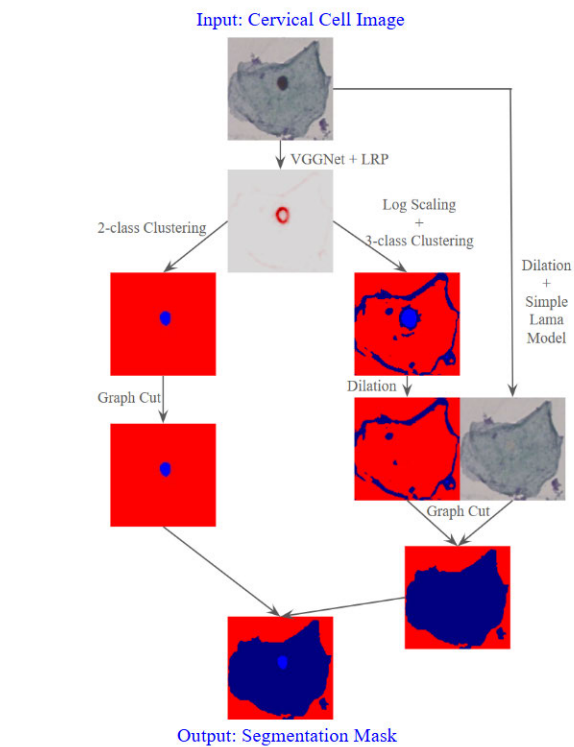


FIGURE 10. LRP based algorithm.

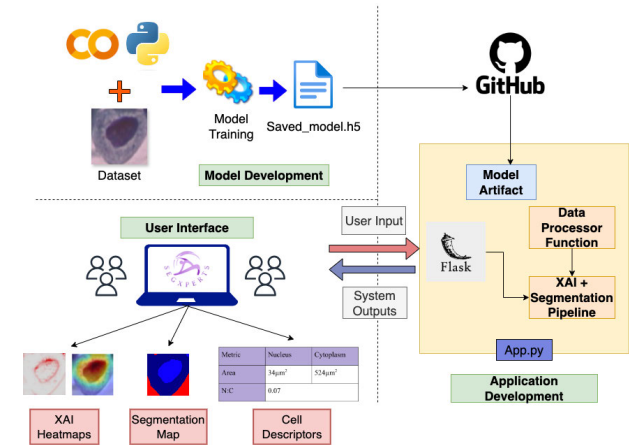


FIGURE 11. System architecture of SegXperts applicaton.

TABLE 6. Summary of performance of classification models.

Model	Acc.	Preci.	Recall	F1
XceptionNet (256 units)	0.81	0.81	0.81	0.81
XceptionNet (1024 units)	0.88	0.90	0.89	0.88
VGG16-Adapted128	0.94	0.94	0.94	0.94
VGG16-Adapted128-Simplified	0.92	0.93	0.92	0.91

higher performance typically rely on complex ensemble model implementations or intricate feature extraction



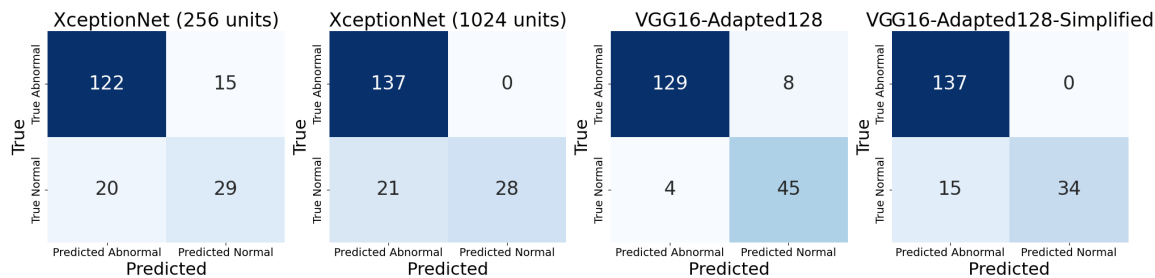


FIGURE 12. Results of confusion matrix.

techniques, which inherently increase computational complexity and make the application of XAI more challenging. In contrast, our study, being the first to explore this novel approach, deliberately focuses on optimizing a single classification model without incorporating ensemble structures. This facilitates the effective application and comparison of a broader range of XAI techniques. Despite this simplicity, our classification results are remarkably close to those achieved by more complex ensemble models, highlighting the effectiveness and practicality of our framework.

B. PERFORMANCE OF XAI TECHNIQUES

XAI techniques were assessed through both visual inspection and quantitative metrics. Table 7 showcases heatmaps generated by GradCAM++ and LRP. Compared to LRP Ruleset1, which primarily outlined the nucleus, LRP Ruleset2 provided a slightly more detailed view by highlighting certain internal regions of the nucleus more distinctly and even faintly outlining the cytoplasm. In contrast, GradCAM++ covered the entire nucleus and cytoplasm regions, with activation maps showing the highest values within the nucleus and slightly lower values within the cytoplasm. However, when images are distorted or contain significant noise, the classification models tend to make predictions with low confidence,

TABLE 7. XAI heatmap examples.

Original Image	Class	Grad CAM++	LRP Ruleset1	LRP Ruleset2
	normal			
	abnormal			
	normal			
	abnormal			

which in turn reduces the effectiveness of the XAI-generated heatmaps.

Table 8 presents the mean entropy values for the XAI methods. LRP, with the lowest mean entropy value, indicates that it focused on highlighting only the regions most influential to the classifier’s decisions. In contrast, CAM methods exhibited higher mean entropy values, reflecting their broader feature visualizations that covered both the nucleus and cytoplasm regions of the cell images.

TABLE 8. Mean image entropy for the XAI methods.

XAI method	Abnormal Class Mean Entropy	Normal Class Mean Entropy	Overall Mean Entropy
GradCAM++	5.37	5.38	5.37
LRP Ruleset1	4.10	2.65	3.37
LRP Ruleset2	4.74	3.12	3.93

Fig. 13 illustrates the assessment of XAI methods using the pixel flipping metric. The GradCAM++ curve shows the steepest decline compared to other XAI methods, indicating that GradCAM++ effectively pinpointed the critical regions influencing the model’s decisions. This steep drop suggests that significant distortions in these key areas led to substantial shifts in the model’s predictions and confidence scores.

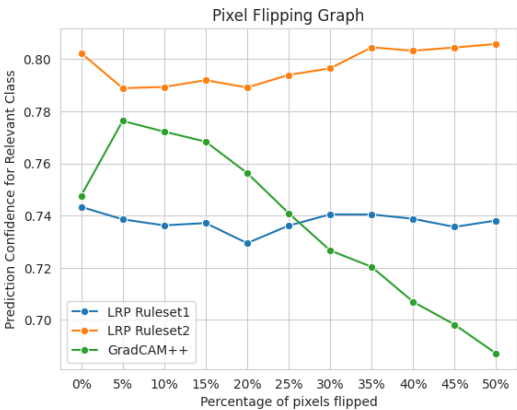


FIGURE 13. Pixel flipping graph.

C. PERFORMANCE OF SEGMENTATION METHODS

Table 9 summarizes the performance of our two algorithms, CAM based algorithm using GradCAM++ and LRP based algorithm using Layer-wise Relevance Propagation (LRP), in generating segmentation masks for cervical cell images. The segmentation masks obtained from these 2 algorithms were evaluated against the ground truth masks using the evaluation metrics Dice Similarity Coefficient (DSC) and Intersection over Union (IoU). The results indicate that GradCAM++ is marginally more effective than LRP in generating segmentation masks for cervical cells.



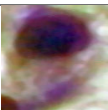

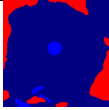

TABLE 9. Segmentation performance.

XAI	DSC	IoU
GradCAM++	62.05%	61.89%
LRP	61.57%	61.24%

This work represents a pioneering effort in leveraging the combination of binary classification, XAI, and GraphCut for weakly supervised segmentation of cervical cells, without relying on ground truth segmentation masks. Achieving this level of performance on a novel framework demonstrates the potential of this approach and lays a strong foundation for further refinement and development.

Table 10 shows some examples of segmentation masks for cervical cell images obtained using our weakly supervised segmentation framework.

TABLE 10. Examples of generated segmentation masks.

Original Image			
Segmentation Mask			

D. SYSTEM VALIDATION AND USABILITY

In order to ensure the validity of the proposed solution, we conducted an expert evaluation involving 37 medical professionals, including 28 medical students, 4 consultants, 3 medical officers, and 2 registrars. These experts assessed the accuracy and quality of our nucleus and cytoplasm segmentations. For the evaluation, experts were given a survey of 10 images consisting of a mix of normal and abnormal samples, with varying image quality. They were asked to rate the segmentations of nuclei and cytoplasm generated by the system. The results were highly favorable, with over 70% expressing satisfaction with the nucleus segmentations and more than 65% with the cytoplasm segmentations.

The System Usability Study (SUS) was used to assess the usability of the SegXperts application. The SUS survey

consists of 10 questions, rated on a 5-point Likert scale, which address both positive and negative usability aspects. The final SUS score is calculated by averaging individual SUS scores across all participants, with scores below 51 classified as “Awful,” 51 to 66 as “Poor,” 67 as “Okay,” 68 to 80.2 as “Good,” and above 80.3 as “Excellent.”

The study involved 20 participants, including 12 medical students, 3 intern house officers, 3 medical officers, and 2 consultants. Their experience in the cervical cancer domain varied, with 14 having less than one year of experience and 6 having 1 to 10 years of experience. The feedback of the participants on the positive and negative statements is illustrated in the percentage breakdowns in Figure 14. As demonstrated here, the SUS results indicated a strong positive reception, with 75% or more agreement on all positive statements, reflecting the user-friendliness and integration of the application. However, 50% of the participants felt that they needed technical assistance, a concern addressed by the comprehensive user manual included in the application. The average SUS score was 77.5, categorizing the application as “Good.”

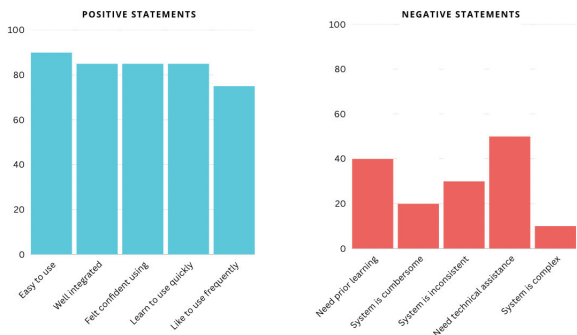


FIGURE 14. System usability study results.

V. DISCUSSION

The performance of the proposed solution is compared with other existing studies that perform segmentation in a supervised, unsupervised, or weakly supervised manner in Table 11. Here, DSC denotes the Dice Similarity Coefficient and IoU stands for Intersection over Union. Given that this is the first attempt to explore weakly supervised segmentation of cervical cells through the novel synergy of binary classification, XAI, and GraphCut, this performance is a significant step forward and demonstrates the feasibility of the method.

One of the primary challenges in this research area is the limited number of studies focused on converting explanations generated by XAI tools into segmentation maps. Additionally, there is a need to develop a cervical cancer classification model that is both less complex and requires fewer computational resources. Another significant challenge lies in the difficulty of evaluating the explanations provided by different XAI techniques, which is crucial for ensuring their reliability and effectiveness.

**TABLE 11.** Comparison of the proposed solution with the previous studies.

Study	Dataset	Type	Technique	DSC	IoU
Chen et al. [41]	ISBI 2014 [42] ISBI 2015 [43]	Supervised	Mask R-CNN	0.920 0.920	
Bandyopadhyay et al. [20]	Herlev [35]	Unsupervised	K-Means Clustering		0.718
Kaur et al. [34]	3D-IRCADb-01 [44]	Supervised	XceptionNet, GradCAM, U-Net	0.977	0.788
Seibold et al. [2]	Sewer Pipe and Magnetic Tile Surface	Weakly Supervised	VGGNet, LRP, Mixture Models	0.402	
Proposed study	Herlev [35]	Weakly Supervised	XceptionNet, VGGNet, Grad-CAM++, LPR, K-Means, GMM, Inter-means, GraphCut	0.621	0.619

Our research effectively addresses all four research questions mentioned in I. For RQ1 (developing an accurate classification model), our VGG16-Adapted128 model that achieved excellent performance metrics (accuracy: 0.94, precision: 0.94, recall: 0.94, and F1 score: 0.94) for distinguishing between malignant and benign cervical cell images, meeting clinical standards for diagnostic support. Regarding RQ2 (applying XAI for transparent segmentation), our implementation of GradCAM++ and Layer-wise Relevance Propagation to provide transparent pixel-wise segmentation of malignant cells, making our model's decisions interpretable to clinicians. For RQ3 (comparing our approach with traditional methods), our novel weakly supervised approach combining classification, XAI, and GraphCut achieved a Dice Similarity Coefficient of 62.05% and Intersection over Union of 61.89% without requiring labor-intensive pixel-level annotations, demonstrating competitive performance compared to traditional methods. Finally, addressing RQ4 (creating a practical diagnostic tool), we transformed our algorithms into "SegXperts," a user-friendly web application that received high satisfaction ratings from medical professionals and a "Good" System Usability Score of 77.5, demonstrating its value for enhancing diagnostic workflows in real clinical settings.

## VI. FUTURE EXTENSIONS AND RESEARCH DIRECTION

This research opens several avenues for enhancing the proposed weakly supervised segmentation framework. Notably, improvements in both the classification models and the application of explainable AI (XAI) methods should be explored simultaneously, as these components are deeply interdependent.

In this study, Layer-wise Relevance Propagation (LRP) was applied exclusively to VGGNet due to its straightforward linear structure. Future research could extend LRP to more sophisticated architectures such as XceptionNet, ResNet, and EfficientNet, which leverage complex skip connections and dense connectivity patterns. Applying LRP to these advanced architectures could offer deeper insights into their decision-making processes and enhance the assignment of pixel significance to the model's decisions.

The use of ensemble models for classifying cervical cell images into malignant and benign categories is frequently observed in related studies II-C. Therefore, a promising direction for future work involves multi-model XAI integration, where interpretations from multiple classification models are combined into a unified visualization framework. Such an approach could aggregate insights from diverse model architectures, enabling researchers to study the consistency and variability of explanations across models. This ensemble XAI strategy could pave the way for more robust and reliable heatmaps, ultimately improving segmentation outcomes.

Further refinement could also be achieved by integrating more advanced and recent CAM methods, such as Score-CAM, LayerCAM, and EigenCAM. This could facilitate more accurate attribution of pixel relevance to model decisions.

Finally, future work should focus on improving the transformation of XAI-generated heatmaps into segmentation masks. Developing more sophisticated clustering, optimization, or refinement algorithms could bridge the gap between heatmap relevance and precise segmentation, contributing to a more reliable and effective weakly supervised segmentation framework.

## VII. CONCLUSION

In conclusion, cervical cancer remains a critical global health challenge, with traditional screening methods such as Pap smears being time-intensive and prone to errors. This study addresses the need for automation by developing a system for automated segmentation of cervical cell images, focusing on the nucleus, cytoplasm, and background regions. Two algorithms were formally introduced to implement a novel weakly supervised segmentation framework that integrates binary classification, XAI, and GraphCut.

The VGGNet-Adapted128-Simplified and XceptionNet (1024 units) models demonstrated impressive classification performance comparable to the existing literature, without relying on ensemble structures. XAI techniques such as GradCAM++ and LRP were employed to extract pixel-level relevance, enabling the identification of regions corresponding to the nucleus, cytoplasm, and background. These heat maps were effectively converted into segmentation masks using clustering methods combined with GraphCut for refinement.

This novel framework delivers a dual benefit: it eliminates the burden of manual annotation while enhancing

transparency in model decision-making. Expert evaluation of the segmentation results revealed high satisfaction rates, underscoring the practical utility of the framework. The system significantly reduces screening time compared to manual methods, allowing clinicians to process more samples and reduce diagnostic backlogs. Additionally, the framework's consistent performance across diverse sample qualities helps standardize diagnostic procedures across different clinical settings, potentially reducing regional disparities in screening quality. The integration capabilities with existing laboratory information systems further streamline clinical workflows, minimizing disruption during implementation. Furthermore, the CAM-based and LRP-based algorithms were seamlessly integrated into a user-friendly web application, providing practitioners with a transparent and reliable tool for informed decision-making in cervical cancer screening. This study represents a significant first step in this innovative direction, establishing a strong foundation to advance weakly supervised segmentation techniques in cancer diagnostics and beyond.

## REFERENCES

- [1] Y. Fan, Z. Tao, J. Lin, and H. Chen, "An encoder-decoder network for automatic clinical target volume target segmentation of cervical cancer in CT images," *Int. J. Crowd Sci.*, vol. 6, no. 3, pp. 111–116, Aug. 2022, doi: [10.26599/IJCS.2022.9100014](https://doi.org/10.26599/IJCS.2022.9100014).
- [2] C. Seibold, J. Künzel, A. Hilsman, and P. Eisert, "From explanations to segmentation: Using explainable AI for image segmentation," 2022, *arXiv:2202.00315*.
- [3] M. Alsallat, H. Alquran, W. A. Mustafa, Y. M. Yacob, and A. A. Alayed, "Analysis of cytology pap smear images based on ensemble deep learning approach," *Diagnostics*, vol. 12, no. 11, p. 2756, Nov. 2022, doi: [10.3390/diagnostics12112756](https://doi.org/10.3390/diagnostics12112756).
- [4] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, X. Wu, X. Li, and Q. Wang, "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104649, doi: [10.1016/j.compbiomed.2021.104649](https://doi.org/10.1016/j.compbiomed.2021.104649).
- [5] D. N. Diniz, M. T. Rezende, A. G. C. Bianchi, C. M. Carneiro, E. J. S. Luz, G. J. P. Moreira, D. M. Ushizima, F. N. S. de Medeiros, and M. J. F. Souza, "A deep learning ensemble method to assist cytopathologists in pap test image classification," *J. Imag.*, vol. 7, no. 7, p. 111, Jul. 2021, doi: [10.3390/jimaging7070111](https://doi.org/10.3390/jimaging7070111).
- [6] S. Wickramanayake, S. Rasnayaka, M. Gamage, D. Meedeniya, and I. Perera, "Explainable artificial intelligence for enhanced living environments: A study on user perspective," in *Internet of Things: Architectures for Enhanced Living Environments*, vol. 133, Cambridge, MA, USA: Academic Press, ch. 1, pp. 1–32, doi: [10.1016/bs.adcom.2023.10.002](https://doi.org/10.1016/bs.adcom.2023.10.002).
- [7] S. Dasanayaka, V. Shantha, S. Silva, D. Meedeniya, and T. Ambegoda, "Interpretable machine learning for brain tumour analysis using MRI and whole slide images," *Softw. Impacts*, vol. 13, Aug. 2022, Art. no. 100340, doi: [10.1016/j.simpa.2022.100340](https://doi.org/10.1016/j.simpa.2022.100340).
- [8] L. Gamage, U. Isuranga, D. Meedeniya, S. De Silva, and P. Yogarajah, "Melanoma skin cancer identification with explainability utilizing mask guided technique," *Electronics*, vol. 13, no. 4, p. 680, Feb. 2024, doi: [10.3390/electronics13040680](https://doi.org/10.3390/electronics13040680).
- [9] T. Shyamalee, D. Meedeniya, G. Lim, and M. Karunarathne, "Automated tool support for glaucoma identification with explainability using fundus images," *IEEE Access*, vol. 12, pp. 17290–17307, 2024, doi: [10.1109/ACCESS.2024.3359698](https://doi.org/10.1109/ACCESS.2024.3359698).
- [10] S. Dasanayaka, S. Silva, V. Shantha, D. Meedeniya, and T. Ambegoda, "Interpretable machine learning for brain tumor analysis using MRI," in *Proc. 2nd Int. Conf. Adv. Res. Comput. (ICARC)*, Feb. 2022, pp. 212–217, doi: [10.1109/ICARC54489.2022.9754131](https://doi.org/10.1109/ICARC54489.2022.9754131).
- [11] P. Jiang, X. Li, H. Shen, Y. Chen, L. Wang, H. Chen, J. Feng, and J. Liu, "A systematic review of deep learning-based cervical cytology screening: From cell identification to whole slide image analysis," *Artif. Intell. Rev.*, vol. 56, no. S2, pp. 2687–2758, Nov. 2023, doi: [10.1007/s10462-023-10588-z](https://doi.org/10.1007/s10462-023-10588-z).
- [12] G. Li, C. Sun, C. Xu, Y. Zheng, and K. Wang, "Cervical cell segmentation method based on global dependency and local attention," *Appl. Sci.*, vol. 12, no. 15, p. 7742, Aug. 2022, doi: [10.3390/app12157742](https://doi.org/10.3390/app12157742).
- [13] J. Zhao, L. Dai, M. Zhang, F. Yu, M. Li, H. Li, W. Wang, and L. Zhang, "PGU-net+: Progressive growing of U-net+ for automated cervical nuclei segmentation," 2019, *arXiv:1911.01062*.
- [14] G. J. Chowdary, G. Suganya, M. Premalatha, and P. Yogarajah, "Nucleus segmentation and classification using residual SE-UNet and feature concatenation approach incervical cytopathology cell images," *Technol. Cancer Res. Treatment*, vol. 22, pp. 1533033822113483–3, Jan. 2023, doi: [10.1177/15330338221134833](https://doi.org/10.1177/15330338221134833).
- [15] K. Allehaibi, L. Nugroho, L. Lazuardi, A. Prabuwo, and T. Mantoro, "Others segmentation and classification of cervical cells using deep learning," *IEEE Access*, vol. 7, pp. 116925–116941, 2019.
- [16] L. Rettenberger, F. R. Münke, R. Bruch, and M. Reischl, "Mask R-CNN outperforms U-net in instance segmentation for overlapping cells," *Current Directions Biomed. Eng.*, vol. 9, no. 1, pp. 335–338, Sep. 2023, doi: [10.1515/cdbme-2023-1084](https://doi.org/10.1515/cdbme-2023-1084).
- [17] S. Ragothaman, S. Narasimhan, M. G. Basavaraj, and R. Dewar, "Unsupervised segmentation of cervical cell images using Gaussian mixture model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 1374–1379, doi: [10.1109/CVPRW.2016.173](https://doi.org/10.1109/CVPRW.2016.173).
- [18] A. Jeffree, C. Pahl, H. N. Abduljabbar, I. Ramli, N. Aziz, Y. M. Myint, and E. Supriyanto, "Cervical segmentation in ultrasound images using level-set algorithm," in *Proc. WSEAS Int. Conf. Biomed. Health Eng.*, 2013, pp. 25–30.
- [19] S. Gautam, K. Gupta, A. Bhavsar, and A. K. Sao, "Unsupervised segmentation of cervical cell nuclei via adaptive clustering," in *Proc. 21st Annu. Conf. Med. Image Understand. Anal. (MIUA)*, Edinburgh, U.K., Jan. 2017, pp. 815–826, doi: [10.1007/978-3-319-60964-5\\_71](https://doi.org/10.1007/978-3-319-60964-5_71).
- [20] H. Bandyopadhyay and M. Nasipuri, "Segmentation of pap smear images for cervical cancer detection," in *Proc. IEEE Calcutta Conf. (CALCON)*, Feb. 2020, pp. 30–33, doi: [10.1109/CALCON49167.2020.9106484](https://doi.org/10.1109/CALCON49167.2020.9106484).
- [21] M. Arya, N. Mittal, and G. Singh, "Texture-based feature extraction of smear images for the detection of cervical cancer," *IET Comput. Vis.*, vol. 12, no. 8, pp. 1049–1059, Dec. 2018, doi: [10.1049/iet-cvi.2018.5349](https://doi.org/10.1049/iet-cvi.2018.5349).
- [22] N. Sritharan, N. Gnanavel, P. Inparaj, D. Meedeniya, and P. Yogarajah, "EnsembleCAM: Unified visualization for explainable cervical cancer identification," in *Proc. Int. Res. Conf. Smart Comput. Syst. Eng. (SCSE)*, Apr. 2024, pp. 1–6, doi: [10.1109/scse61872.2024.10550859](https://doi.org/10.1109/scse61872.2024.10550859).
- [23] N. Gnanavel, P. Inparaj, N. Sritharan, D. Meedeniya, and P. Yogarajah, "Interpretable cervical cell classification: A comparative analysis," in *Proc. 4th Int. Conf. Adv. Res. Comput. (ICARC)*, Feb. 2024, pp. 7–12, doi: [10.1109/icarc61713.2024.10499737](https://doi.org/10.1109/icarc61713.2024.10499737).
- [24] S. L. Tan, G. Selvachandran, W. Ding, R. Paramesran, and K. Kotecha, "Cervical cancer classification from pap smear images using deep convolutional neural network models," *Interdiscipl. Sci., Comput. Life Sci.*, vol. 16, no. 1, pp. 16–38, Mar. 2024, doi: [10.1007/s12539-023-00589-5](https://doi.org/10.1007/s12539-023-00589-5).
- [25] D. Meedeniya, *Deep Learning: A Beginners' Guide*. Boca Raton, FL, USA: CRC Press, 2023.
- [26] M. Fang, M. Fu, B. Liao, X. Lei, and F.-X. Wu, "Deep integrated fusion of local and global features for cervical cell classification," *Comput. Biol. Med.*, vol. 171, Mar. 2024, Art. no. 108153, doi: [10.1016/j.compbiomed.2024.108153](https://doi.org/10.1016/j.compbiomed.2024.108153).
- [27] H. Alquran, M. Alsallat, W. Mustafa, R. Abdi, and A. Ismail, "Cervical net: A novel cervical cancer classification using feature fusion," *Bioengineering*, vol. 9, p. 578, Oct. 2022, doi: [10.3390/bioengineering9100578](https://doi.org/10.3390/bioengineering9100578).
- [28] A. D. Jia, B. Zhengyi Li, and C. C. Zhang, "Detection of cervical cancer cells based on strong feature CNN-SVM network," *Neurocomputing*, vol. 411, pp. 112–127, Oct. 2020, doi: [10.1016/j.neucom.2020.06.006](https://doi.org/10.1016/j.neucom.2020.06.006).
- [29] V. Pitroda, M. M. Fouda, and Z. M. Fadlullah, "An explainable AI model for interpretable lung disease classification," in *Proc. IEEE Int. Conf. Internet Things Intell. Syst. (IoTIS)*, Nov. 2021, pp. 98–103, doi: [10.1109/IOTIS53735.2021.9628573](https://doi.org/10.1109/IOTIS53735.2021.9628573).
- [30] N. Patel, S. Parmar, P. Singh, and M. Mohanty, "XAIForCOVID-19: A comparative analysis of various explainable AI techniques for COVID-19 diagnosis using chest X-ray images," in *Proc. Int. Conf. Comput. Vis. Image Process.*, Jan. 2023, pp. 503–517, doi: [10.1007/978-3-031-31417-9\\_38](https://doi.org/10.1007/978-3-031-31417-9_38).



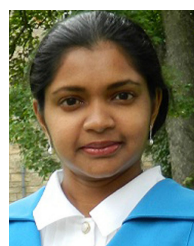
- [31] M. Bhandari, P. Yogarajah, M. S. Kavitha, and J. Condell, "Exploring the capabilities of a lightweight CNN model in accurately identifying renal abnormalities: Cysts, stones, and tumors, using LIME and SHAP," *Appl. Sci.*, vol. 13, no. 5, p. 3125, Feb. 2023, doi: [10.3390/app13053125](https://doi.org/10.3390/app13053125).
- [32] J. H. Ong, K. M. Goh, and L. L. Lim, "Comparative analysis of explainable artificial intelligence for COVID-19 diagnosis on CXR image," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2021, pp. 185–190, doi: [10.1109/ICSIPA52582.2021.9576766](https://doi.org/10.1109/ICSIPA52582.2021.9576766).
- [33] J. Civit-Masot, F. Luna-Perejon, L. Muñoz-Saavedra, M. Domínguez-Morales, and A. Civit, "A lightweight xAI approach to cervical cancer classification," *Med. Biol. Eng. Comput.*, vol. 62, no. 8, pp. 2281–2304, Aug. 2024, doi: [10.1007/s11517-024-03063-6](https://doi.org/10.1007/s11517-024-03063-6).
- [34] A. Kaur, G. Dong, and A. Basu, "GradXcepUNet: Explainable AI based medical image segmentation," in *Proc. Int. Conf. Smart Multimedia*, Jan. 2022, pp. 174–188, doi: [10.1007/978-3-031-22061-6\\_13](https://doi.org/10.1007/978-3-031-22061-6_13).
- [35] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," in *Proc. Nature Inspired Smart Inf. Syst. (NiSIS)*, Jan. 2005, pp. 1–9.
- [36] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLoS ONE*, vol. 10, no. 7, Jul. 2015, Art. no. e0130140, doi: [10.1371/journal.pone.0130140](https://doi.org/10.1371/journal.pone.0130140).
- [37] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K. Müller, "Layer-wise relevance propagation: An overview," in *Explainable AI: Interpreting, Explaining Visualizing Deep Learn.* Cham, Switzerland: Springer, Jan. 2019, pp. 193–209, doi: [10.1007/978-3-030-28954-6\\_10](https://doi.org/10.1007/978-3-030-28954-6_10).
- [38] A. Holzinger, A. Saranti, C. Molnar, P. Biecek, and W. Samek, "Explainable AI methods—A brief overview," in *Proc. Int. Workshop Extending Explainable AI Beyond Deep Models Classifiers*, Vienna, Austria, Jul. 2022 pp. 13–38, doi: [10.1007/978-3-031-04083-2\\_2](https://doi.org/10.1007/978-3-031-04083-2_2).
- [39] *SegXperts—Segxperts.live*. Accessed: Sep. 15, 2024. [Online]. Available: <https://segxperts.live>
- [40] J. Kauffmann, K.-R. Müller, and G. Montavon, "Towards explaining anomalies: A deep Taylor decomposition of one-class models," *Pattern Recognit.*, vol. 101, May 2020, Art. no. 107198, doi: [10.1016/j.patcog.2020.107198](https://doi.org/10.1016/j.patcog.2020.107198).
- [41] J. Chen and B. Zhang, "Segmentation of overlapping cervical cells with mask region convolutional neural network," *Comput. Math. Methods Med.*, vol. 2021, pp. 1–10, Oct. 2021, doi: [10.1155/2021/3890988](https://doi.org/10.1155/2021/3890988).
- [42] G. Carneiro and A. Bradley. *ISBI 2014 Cervical Cancer Segmentation Challenge Dataset*. Accessed: Aug. 19, 2024. [Online]. Available: [https://cs.adelaide.edu.au/carneiro/isbi14\\_challenge/dataset.html](https://cs.adelaide.edu.au/carneiro/isbi14_challenge/dataset.html)
- [43] G. Carneiro and A. Bradley. *ISBI 2015 Cervical Cancer Segmentation Challenge Dataset*. Accessed: Aug. 19, 2024. [Online]. Available: [https://cs.adelaide.edu.au/carneiro/isbi15\\_challenge/dataset.html](https://cs.adelaide.edu.au/carneiro/isbi15_challenge/dataset.html)
- [44] *3D IRCADb-01 Liver Segmentation*. Accessed: Aug. 19, 2024. [Online]. Available: <https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/>



**NISHAANTHINI GNANAVEL** received the B.Sc. degree in computer science and engineering, specializing in data science, from the University of Moratuwa. She has authored research publications in the fields of deep learning and explainable AI (XAI). Her main research interests include deep learning and large language models.



**PRATHUSHAN INPARAJ** received the B.Sc. degree in computer science and engineering with a specialization in data science, from the University of Moratuwa. He has authored publications on explainable artificial intelligence, particularly in the area of cervical cancer segmentation. His research interests include machine learning, deep learning, and large language models.



**DULANI MEEDENIYA** (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of St Andrews, U.K. She is currently a Professor of computer science and engineering with the University of Moratuwa, Sri Lanka. She is also the Director of the Bio-Health Informatics Group at her department and engages in many collaborative research. She is the co-author of more than 100 publications in indexed journals, peer-reviewed conferences, and international book chapters. Her main research interests include software modeling and design, bio-health informatics, deep learning, and technology-enhanced learning. She is a fellow of HEA (U.K.) and a member of IET and ACM. She has received several awards and grants for her contribution to research. She serves as a reviewer, a program committee member, and an editorial team member for many international conferences and journals. She is a Chartered Engineer registered at EC (U.K.).



**PRATHEEPAN YOGARAJAH** (Member, IEEE) received the degree (Hons.) in computer science from the University of Jaffna, Sri Lanka, in 2001, the M.Phil. degree in computer vision from Oxford Brookes University, U.K., in 2006, and the Ph.D. degree from Ulster University, U.K., in 2015. He has been a Lecturer in computing science with Ulster University, since January 2016. His research interests include biometrics, computer vision, image processing, steganography and digital watermarking, and machine learning. He is a member of the British Computer Society (BCS). He was a recipient of Oxford Brookes University HMGCC Scholarship Award, in 2005. He was also a co-recipient of the Proof of Principle (PoP) Award from Ulster University, in 2012, and the Proof of Concept (PoC) from Invest Northern Ireland (Invest NI), in 2013.



**NIRUTHIKKA SRITHARAN** received the B.Sc. degree in data science from the Computer Science and Engineering Department, University of Moratuwa. She has authored publications in the areas of deep learning and explainable artificial intelligence (XAI). Her primary research interests include deep learning, large language models (LLMs), and XAI.