Automatic Segmentation of Nuclei in Single-cell and Multi-cell images using Deep Learning Approaches

Hridya Nair Suresh
3096874

ABSTRACT

Cervical cancer is a type of cancer that occurs in the cells of the cervix, the lower part of the uterus that connects to the vagina. Various strains of the human papillomavirus (HPV) play a role in causing most cervical cancer. A Papanicolaou test (pap smear test) is the most reliable screening method for identifying cancerous cells. Pathologists visually examine the slide after performing the test to look for malignancy. But this is a time-consuming and challenging task, often limited by the availability of certified professionals. Additionally, the diagnostic efficacy of visual screening is compromised by the increased workload of pathologists, leading to decreased diagnostic accuracy. Thus, a computer-aided diagnosis system is essential in order to enhance the quality of the Pap test results and enable early diagnosis. Nuclei segmentation is a crucial step in automated cervical cancer detection, as the nucleus carries essential information about the alterations occurring during the disease's progression. Thus, automated segmentation of nuclei can significantly influence the diagnostic results. Recent surveys found that neural networks outperformed various traditional segmentation methods. This research is in two phases, the first phase involves the segmentation of nuclei in a single-cell slide(Herlev Dataset) image and the second phase includes the segmentation of nuclei in a multicell slide image(CCEDD). The base deep learning model used here is Unet with EfficientNetb0 encoder.

KEYWORDS

Unet, Efficientnet, Semi-supervised learning, Papsmear dataset, Cervical Cancer, Nuclei segmentation

1 INTRODUCTION

Cervical cancer is one of the most common causes of death among females. It is the uncontrolled growth of cancer cells in or on the cervix. Since this cancer is slow-growing, it provides better chances for early detection and effective treatment. Cervical cancer is associated with high-risk strains of the human papillomavirus (HPV). Precancerous lesions that eventually advance to cervical cancer can be identified by cytologic screening of cervical cells. The Papanicolaou test (pap smear test), a cervical cancer screening test in which cells from the outside and inside of the cervix are used as samples for testing, locates HPV-induced changes that have the potential to progress to cancer as well as detects cervical cancer cells. Pap smear slides are obtained after a pap smear examination. Pap smear slides contain cell images that consist of nuclei, cytoplasm, etc. The shape, colour, and texture of the nuclei are abnormal when an infected cervical cell is examined. The cancer cells show nucleus enlargement and irregularity. A pathologist visually examines the slide, which is a laborious and time-consuming process due to the limited number of qualified pathologists and the high workload they face. Therefore, automating the screening process would yield faster and more accurate results, leading to a decrease in cervical cancer mortality rates. Nuclei segmentation is a crucial step when implementing an automated system, as the nucleus carries essential information about the alterations occurring during the disease's progression. Hence, by automating the segmentation of nuclei, we can enhance the performance of the classification model, resulting in a faster and more efficient diagnosis of cervical cancer. Medical image segmentation plays a significant role in computer-aided diagnosis of diseases as it gives a clear picture of the images and hence increases the accuracy and efficiency of CAD. There are several techniques and models that have been developed to perform this task of segmentation. Among them, the early approaches were based on edge detection, machine learning, template matching techniques, etc. But using these traditional methods, the segmentation task was still a challenging one due to the difficulty in representing the features of images. The development of deep learning techniques led to the achievement of hierarchical feature representation of images and also provided excellent segmentation results. Some popular medical image segmentation tasks include cell nuclei segmentation, colorectal cancer segmentation, liver, and liver tumour segmentation, and lung tumour segmentation, etc.

Compared to single cell nuclei segmentation multi cell nuclei segmentation is complex and through this research, the effectiveness of the model trained in single-cell nuclei on multi-cell nuclei segmentation is also examined.

2 RELATED STUDIES

2.1 Nuclei segmentation using Supervised Deep learning models

Deep learning models have shown good performance compared to traditional methods in medical image segmentation. If labelled data is available for the task of segmentation supervised deep learning models are suitable to use.

Unet is a common architecture used in segmentation. To improve the efficiency and performance, the encoder of the Unet is often swapped with other models like Resnet, EfficientNet, etc. [6] proposes a multi-task network based on U-net with ResNet-34 encoder for the segmentation process. An attention learning module is also introduced behind each skip connection in the network. This helps in learning focused attention.

Another approach [7] proposes an automated cervical nuclei segmentation method using a deformable multipath ensemble model(D-MEM). The approach adopts a U-shaped convolutional network as a backbone network, in which dense blocks are used to transfer feature information more effectively. To increase the flexibility of the model, deformable convolution was used which deals with different nuclei irregular shapes and sizes. As the abnormal nuclei exhibit different shapes, the traditional Unet model cannot handle varying

shapes. Thus, deformable convolution is introduced to enhance the transformation modelling capability.

In the same manner [2] proposes a combination of Nested Unet and EfficientNet models for the nuclei segmentation of the cell images in a Cryonuseg dataset. This model performed better when compared to other Unet models for this dataset. The Cryonuseg dataset has images from 10 human organs and is used mainly for nuclei segmentation of cryosectioned H&E-Stained Histological images. The building block of EfficientNet is MBConv (Mobile Bottleneck Convolution) which has squeeze and excitation layers for optimisation. It is the most compact and high-performing model and thus can extract the feature map efficiently. As this model performs well with a variety of cell images, in this research the Unet with EfficientNet model is used for the nuclei segmentation task.

2.2 Nuclei segmentation using Semi-Supervised Deep learning models

Deep learning models require a large amount of data, but the availability of labelled data can be sometimes a challenge, as it is a time-consuming manual task. Semi-supervised learning(SSL) aids in such situations. Many researchers in the field of nucleus segmentation have used this approach in different ways. Pseudolabeling a method where the model trained on labelled data is used to predict the label of unlabelled data(pseudo labels) is one technique used in SSL. Another method used for semi-supervised learning is the generative adhesive network(GAN) based method where a generator and a discriminator are trained in an adversarial manner to improve the quality of generated samples and enhance the performance of the model on both labelled and unlabeled data.

Another method explored is consistency-based semi-supervised learning [1]. In this approach, the model is trained with both unlabelled and labelled data to learn the features based on the consistency enforced by the varying contexts and features. A variation of this method is semi-supervised learning using dual-task consistency[5] where the model is trained with a combination of labelled and unlabelled data and for the training criteria, an additional loss function called dual-task consistency loss along with the traditional loss function like dice loss. In this research, this technique is used to segment the nuclei in multi-cell images, as the available dataset is unlabeled. The approach was first experimented with an unsupervised method with only dual task consistency loss as the training factor and later tried with a semi-supervised approach where a part of the multicell dataset was labelled with the masks of the single-celled data.

3 DATASET AND PREPROCESSING

3.1 Single cell nuclei segmentation - Phase 1

3.1.1 ***Dataset :*** The dataset used for this phase is the Herlev dataset. It contains pap smear images of single cervical cells with their corresponding masks. These images were distributed among 7 files. Each file represents the 7 classes of cervical nuclei which are carcinoma in situ (Figure 1), light dysplasia (Figure 2), moderate dysplasia (Figure 3), normal columnar (Figure 4), normal intermediate (Figure 5), normal superficial (Figure 6) and severe dysplasia (Figure 7). The files were merged into a single file, containing a total of 917 images. Then divided the dataset into training and test set. After splitting there were 317 images on the training sets and 600 images on the test sets. Data augmentation was done to the training set to make our model learn the precise details.

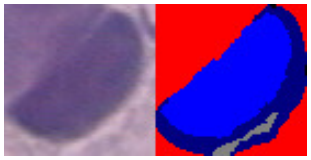


Figure 1: Squamous cell carcinoma in situ

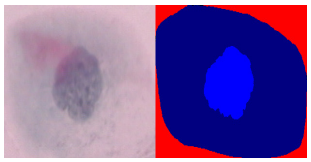


Figure 2: Light dysplastic

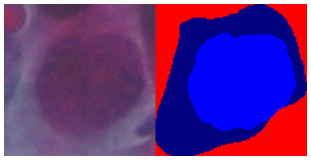


Figure 3: Moderate dysplastic

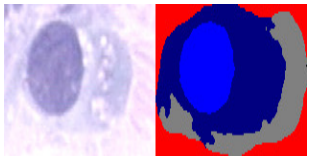


Figure 4: Normal columnar

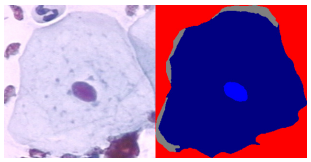


Figure 5: Normal intermediate

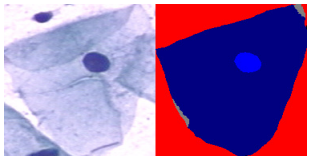
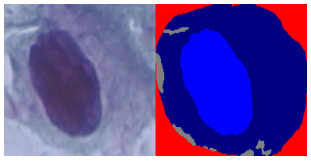


Figure 6: Normal superficial



3.1.2 ***Data Augmentation:*** Data augmentation is the process of artificially increasing the size of a dataset by generating different versions of the existing data. After train test splitting we had 340 images for training. A deep learning model generally works well when it has a huge amount of data. In general, the more data we have better will be the performance of the model. The problem with the lack of a good amount of data is that the deep learning model might not learn the patterns or the functions from the data and hence it might not perform well. So the training set was augmented using different data augmentation techniques. For semantic segmentation tasks, we need to augment both the input image and output masks. The augmentation techniques used to enhance the dataset are horizontal flip, vertical flip, rotation, sharpness, grayscale, equalize, affine transform, posterize, brightness, hue, contrast, and saturation. Torchvision.transforms module contains many transform techniques to manipulate images. After augmentation 6357 images were obtained for training. The torchvision.transforms module from the PyTorch library provides all the functions for performing different data augmentation techniques.

HorizontalFlip: It involves producing a mirrored version of an original image by reflecting it across a vertical axis. The RandomHorizontalFlip() function is used to perform HorizontalFlip in the PyTorch library.

VerticalFlip: It involves creating a mirror image of an original image by flipping it vertically along the horizontal axis. It is performed by using the RandomVerticalFlip() function provided by the PyTorch library.

Rotation: Images are rotated by a predetermined angle. The RandomRotation() function provided by the PyTorch library is used to perform this technique

AdjustSharpness: This technique modifies the sharpness of an image by adding or removing some level of sharpness. It is adjusted by applying a blur or sharpen filter. The level of adjustment is determined by a factor between a specified range that controls the intensity of the blur or sharpen filter. RandomAdjustSharpness() is the function used for this.

Grayscale: It involves converting a colour image to grayscale by removing the colour channels (red, green, and blue) and keeping only the brightness channel. The Grayscale() function of the PyTorch library performs this technique. The Number of input channels is a function parameter.

ColorJitter: It is an image transformation that randomly changes the brightness, contrast, saturation, and hue of an image. ColorJitter() function performs this augmentation technique. Brightness, contrast, saturation, and hue are taken as input parameters of this function.

Equalize: It enhances the contrast of an image by balancing the histogram of the image. RandomEqualize() function performs this augmentation technique.

Affine: Affine transformations can be used to perform a variety of geometric transformations on images, such as rotation, scaling, and translation. RandomAffine() function of the torchvision.transforms module performs this function.

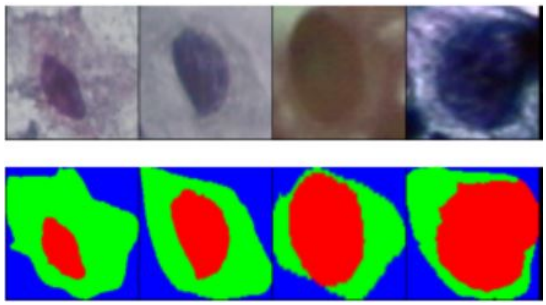


Figure 9: Preprocessed masks

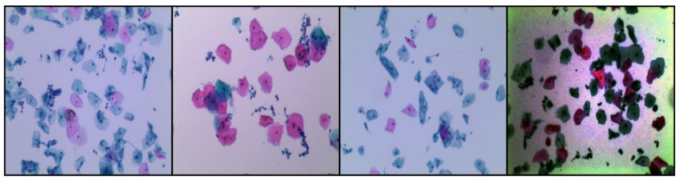


Figure 10: Multicell slide images of cervical cells

Posterizarion: It involves reducing the number of colours in an image. In the end, the image has fewer colour shades and looks more like a poster. RandomPosterize() function of the torchvision.transforms module performs this function. Bits is an input parameter for this function that represents the number of bits for each colour channel to keep.

3.1.3 Preprocessing of Masks. The original mask of the pap smear dataset comprised four colours. Light blue indicates the nucleus, while dark blue represented the cytoplasm. The red color indicated the background and an unknown grey region was present in some masks. To preprocess the masks they were first converted into RGB format and used color thresholding to separate the different regions. To remove the grey region, separate masks were created for the nucleus, cytoplasm, and background. Finally, these masks were merged into a single array to get the new preprocessed masks where red indicates the nucleus, green indicates the cytoplasm and blue represents the background. The preprocessed masks are shown in Figure 9

3.2 Multicell nuclei segmentation- Phase 2

3.2.1 Dataset. The dataset used in the second phase of the research is from the CCEDD (cervical cell edge detection dataset) [see figure 10]. The Dataset consist of 686 slide images of cervical cells(multicells). Masks were not available with the dataset. But edge-detected images of the cells were available.

3.2.2 Data Augmenation. The training set was augmented using the following techniques horizontal flip, vertical flip, rotation, sharpness, grayscale, equalize, affine transform, posterize, brightness, hue, contrast, and saturation.

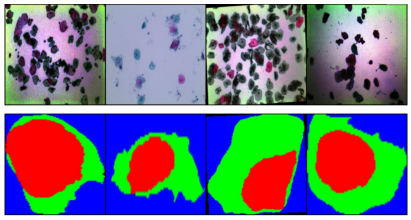


Figure 11: Multicell image with masks

3.2.3 Combining the masks and the images. In the next step of the multicell nuclei segmentation(semi-supervised learning), part of the dataset required labels. So half of the multicell dataset was combined with the masks of the single-cell nuclei of the pap smear as they both represented the same type of cells [see figure 11].

4 PROPOSED SYSTEM

4.1 Single cell nuclei segmentation - Phase 1

4.1.1 . **Unet**

Unet is a U-shaped Convolutional network that has a specific scheme for its encoder and decoder side as shown in Figure 12. It consists of a contracting path and an expansive path. The encoder part, which is the contracting path reduces the spatial dimensions and increases the number of channels where as the decoder, which is the expansive path increases the spatial dimensions and reduces the number of channels[3]. The architecture includes skip connections that cross from the same sized part in the downsampling (contracting) path to the upsampling (expansive) path. With the input mapped almost exactly to the output, this causes all of the input images' fine details to appear at the top of the U-Net. In the original architecture of Unet, the encoder and decoder part is nearly symmetric, but in this proposed system, EfficientNet is used as the encoder instead.

4.1.2 . **EfficientNet**

EffiecientNet is a family of convolutional networks which has versions from B0 to B7 based on the number of parameters. Scaling of convolutional networks is done to achieve better accuracy. These architectures can be scaled by width, depth or resolution. Depth scaling involves adding more layers to the network, to make it deeper. Width scaling increases the number of filters in each layer, making the network wider and resolution scaling involves increasing the size of the input image, which increases the number of pixels.

In this study the model was experimented with these different versions for the encoder and found that all of them achieved comparable performance, however, the B0 version was chosen as it has less number of parameters.

EffievientNetb0 has 5.3 million parameters and consists of mainly 7 blocks while starting with a 3x3 convolution layer[4]. Each block contains Mobile Inverted Residual Bottleneck layers(MBconv layers) to which squeeze and excitation layers are also added. MBconv layers have three main components.

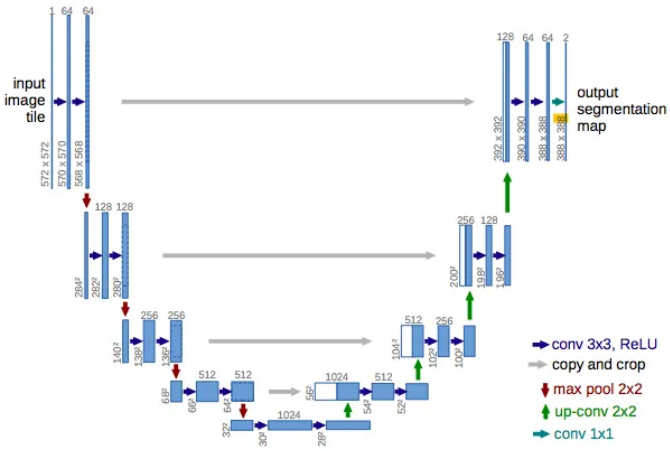


Figure 12: Unet architecture from [3]

Depthwise convolution: This type of convolutional layer applies a single filter to each input channel which results in the reduction of the number of parameters and makes the feature extraction computationally effective.

Pointwise Convolution: These are the convolutional layers with a 1x1 filter to perform a linear transformation of the input tensor. In efficientnetb0 this layer of MBconv is to increase the number of channels after depthwise convolution.

Skip Connection: A skip connection is established from the output layer of the block to the input layer to add information. It helps to improve the flow of gradients through the network. The Architecture of the EfficientNetb0 is given in Figure 13

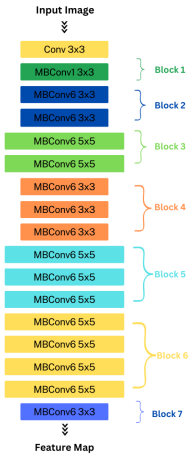


Figure 13: EfficientNet b0 Architecture

The proposed model was obtained by combining the EfficientNet encoder with the Unet architecture, as illustrated in Figure 14. The EfficientNet b0 encoder extracts the high-level features from the input image and the decoder of the Unet transforms these features back into the pixel space to produce the output segmentation map, here the map with the segmented nucleus.

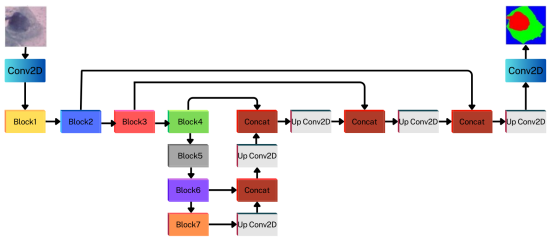


Figure 14: U-Net with EfficientNet b0 Architecture

Further details regarding the training process will be discussed in the next sections.

4.2 Multicell nuclei segmentation

The model trained on the single-cell images was used as the base model, by using the transfer learning approach. The dataset didn't have masks, so non-supervised learning was adapted. For this purpose, more layers were added to the model for learning. The architecture of the proposed model consists of two main components: the **segmentation head** and the **regression head**.

Segmentation head: The segmentation head of the Dual U-Net model utilizes the pre-trained weights from the base model trained on single-cell images. This transfer learning approach allows the model to leverage the learned features from the single-cell images and apply them to the multicell nuclei segmentation task.

Regression head: It consists of several fully connected layers that perform regression to estimate the level sets of the nuclei boundaries. This regression head provides supplementary information about the shapes and boundaries of the nuclei.

In non-supervised learning with dual-task consistency, the model aims to learn meaningful representations from the unlabeled data. The segmentation head predicts the pixel-wise segmentation of the nuclei (masks), while the regression head estimates the level sets. By enforcing consistency between the predictions of these two tasks, the model learns to capture relevant patterns and representations in the absence of labelled masks. This dual-task consistency helps the model to improve its understanding of the nuclei structures and enhances its segmentation performance without the use of labelled data.

As mentioned before, this experiment was done in two ways - unsupervised(without masks) and semi-supervised (partially labelled images).

Unsupervised Learning

In this approach all images in the dataset were unlabelled. As mentioned before the architecture used has 2 heads. The segmentation head provides a segmentation pixel-wise map of the input

image and the regression head provides a level-set function output of the input image.

The level set function from [5] is used and is defined as follows

$$F(x) = \begin{cases} -inf_{y \in \partial S} \|x - y\|_2 & \text{if } x \in S_{\text{in}} \\ 0 & \text{if } x \in \partial S \\ +\inf_{y \in \partial S} \|x - y\|_2 & \text{if } x \in S_{\text{out}} \end{cases}$$

where x and y are two different pixels in a segmentation mask. ∂S represents the contour or boundary of the target object, which is nuclei in this case. S_{in} and S_{out} denote the inside region and outside region of the target object.

The learning process depends on the dual task consistency loss (L_{DTC}) which is calculated between the predicted segmentation map and the converted pixel-wise map of the level set output.

To convert the level set output to a pixel-wise map the function F1 (see Eq 1) was used.

$$F1(z) = \frac{1}{1 + e^{-k \cdot z}} \tag{1}$$

where z means the level set value at pixel x

The function F1 is the inverse of the level set function F(x). But is not practical to find the exact inverse of this function due to the non-differentiability. Thus a sigmoid function with the input multiplied by a factor k, which is selected as large as possible to approximate the inverse transform of F(x) is utilised whereS_{in} are assigned to 1 while those of S_{out} are assigned to 0 in the transformed prediction map,

To calculate the Dual Task Consistency loss(L_{DTC}) between the segmentation map $f_1(x)$ and the converted level set function $F1(f_2(x_i))$, the following function is used:

$$L_{\text{DTC}}(x) = \sum_{x_i \in D} ||f_1(x_i) - F1(f_2(x_i))||^2 \tag{2}$$

where D is the dataset.

The above-proposed approach is described in figure 15

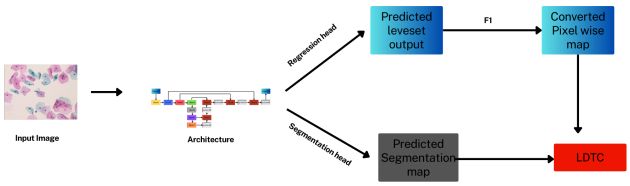


Figure 15: Overview of the proposed unsupervised model

Semi-Supervised Learning

To improve the learning process, half of the dataset images were appended with masks from the pap smear dataset. Since there is an additional learning factor guiding the process, this can be considered semi-supervised learning. In addition to L_{DTC}, the Dice Loss function L_{dice} and the loss function for the Level set function L_{LSF} are used for the learning process.

Dice Loss: This is the supervised loss function calculated between the predicted segmentation map and the masks of the labelled images. The loss function is the common dice loss function as defined below:

$$\text{Dice Loss} = 1 - \frac{2 \sum_{(x_i, y_i) \in D_l} (f_1(x_i) \cdot y_i)}{\sum_{x_i \in D_l} (f_1(x_i)) + \sum_{y_i \in D_l} (y_i)} \tag{3}$$

whereD_{l} is the labelled dataset.

Loss function for level set function L_{LSF}: L_{LSF} is also a supervised loss function calculated between the predicted level set function output and the level set function of the masks (using the equation in 4.2) of the labelled images.

This loss function from [5] is defined as follows:

$$L_{LSF} = \sum_{(x_i, y_i) \in D_l} ||f_2(x_i) - F(y_i)||^2 \tag{4}$$

where $f_2(x_i)$ is the predicted level set head output and $F(y_i)$ is the level set function of the masks of images in the labelled dataset D_l

So the total loss function can be defined as:

$$\text{Total loss} = L_{DTC} + L_{Dice} + L_{LSF} \tag{5}$$

An overview of this system is given in figure 16

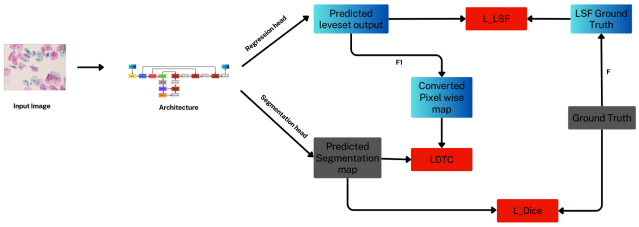


Figure 16: Overview of the proposed semi-supervised model

5 EXPERIMENTAL SET UP -MODEL TRAINING

Training the proposed models is a crucial step in developing deep learning techniques as it enables the model to learn the features and patterns in the data and perform accurate predictions. It includes optimizing the parameters to minimize the loss function and improve the performance. This section discusses the methods employed during the training and provides insights into the training process for each model.

5.1 Single cell nuclei segmentation - Phase 1

For this phase, the approach is supervised learning and the following are the hyperparameters used during the training process.

Early Stopping: This technique is implemented to prevent overfitting. It is a regularization technique that suspends the training process once the validation dataset performance starts to deteriorate. The performance of the model is tracked using the validation loss. The patience parameter which specifies the maximum number of epochs that the model is allowed to continue training without seeing any improvement in the validation loss is set to 3 and the delta parameter which specifies the minimum amount of improvement that is considered significant is taken as 0.0001.

Batch Normalization: A pre-processing technique used to standardize data is normalization. Batch Normalization is a normalization technique done between the layers of a Neural Network instead of in the raw data. This technique has been employed to improve the model's learning rate and prevent overfitting.

Optimizer: Adam optimizer is used. This optimization algorithm is an extension of the stochastic gradient descent (SGD) algorithm. Adam optimizer modifies the learning rate for each network weight separately, unlike SGD, which maintains a single learning rate. It is more computationally efficient than other optimization algorithms.

Activation Function: Softmax activation function is used as the activation function for all the implemented approaches. It is a mathematical function that creates a probability distribution of possible outcomes from a vector of real numbers of the same size. Mathematically softmax is defined as, $S(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$

The dataset is divided into 80% training, 10% validation, and 10% test sets

The proposed model uses Unet with the pre-trained EfficientNet b0 as the encoder with the encoder weight initialized to 'ImageNet'. The model is trained for 60 epochs with an initial learning rate of 0.0001, which is reduced to 0.00001 after 25 epochs. The learning rates are finalized after multiple runs for hyperparameter tuning. The validation set is processed in batches of 4, while the training set is processed in batches of 8. Early stopping, batch normalization, optimizer, and activation functions are selected as same as the parameters mentioned at the beginning of this section.

5.2 Multicell cell nuclei segmentation - Phase 2

Unsupervised learning

The dataset is divided into 70% training, 15% validation, and 15% test sets.

The proposed model uses the model trained on the pap smear dataset with additional layers to incorporate the level set function as mentioned in the previous section. The model is trained for 30 epochs with an initial learning rate of 0.0001. The learning rates are finalized after multiple runs for hyperparameter tuning. The validation set and training set are processed in batches of 4, Early stopping, batch normalization, optimizer, and activation functions are selected as same as the parameters mentioned before.

Semi supervised learning

To enhance the performance, this approach was experimented with. The experimental set-up is the same as the previous approach, the difference is the dataset used and the loss function. The dataset has two parts- labelled and unlabelled. The supervised loss functions dice loss(L_{dice}) and loss of level set results (L_{LSF}) was calculated for the labelled data and dual-task consistency loss was calculated for both types of dataset.

Half of the dataset was labelled. The labelled and unlabelled dataset is divided into 80% training, 10% validation and 10% test sets.

5.3 Evaluation Metrics

The performance of the implemented models is evaluated using widely accepted metrics and their overall effectiveness in accurately segmenting nuclei.

Dice loss is a widely-used loss function used to calculate the similarity between images and is similar to the Intersection-over-Union (IoU) heuristic. IoU is calculated by dividing the anticipated segmentation's area of the union by the ground truth's area of overlap.

Some of the other metrics used for evaluation are accuracy, precision, recall, F1 score, and Zijdenbos Similarity Index (ZSI). Pixel accuracy computes the per cent of pixels in the image that are classified correctly. Precision is defined as the number of predicted objects in an image that correspond to ground truth annotations. Recall is a metric that counts the number of ground truth annotations that are successfully classified as positive predictions. F1 score evaluates the model's accuracy. This accuracy measure counts the number of times a model correctly predicts throughout the full dataset. Another measurement, Zijdenbos Similarity Index (ZSI) is a mostly used evaluation metric in medical image segmentation proposed by Alex P. Zijdenbos [8]. The mathematical definitions of these metrics are given below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

$$ZSI = \frac{2 * TP}{2 * TP + FP + FN} \tag{10}$$

where TP is the number of pixels that are detected correctly, TN denotes a pixel that is accurately identified as not belonging to the specified class, FP means False Positive, and FN denotes False Negative. For the multicell dataset, the evaluation metrics used is the loss function as mentioned before (eq 2 and eq 5)

6 RESULTS

6.1 Phase1

The proposed approach was able to localize the nucleus within the slide image from the Herlev Pap smear dataset. The model's performance is evaluated using widely accepted metrics such as precision, recall, and F1 score to assess their overall effectiveness in accurately segmenting nuclei.

Table 1: Performance of test set - Single cell

IoU	Accuracy	Precision	Recall	ZSI	Dice Loss
0.938	95%	0.974	0.976	0.975	0.031

Unet with EfficientNetb0 encoder provided a training performance with loss of 0.02, 0.95 IoU score and 96% accuracy and a validation performance with loss of 0.028, IoU score 0.94 and 95% accuracy. The dice loss and IoU score of train and validation set over the epochs are shown in Figure 17, and Figure 18 respectively.

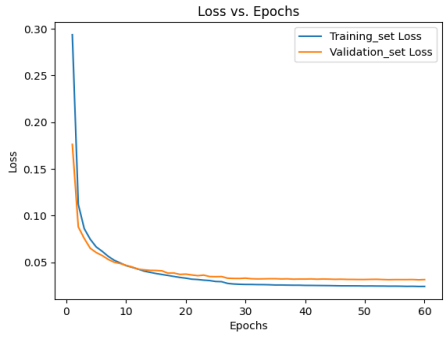


Figure 17: Loss vs Epoch- Single cell

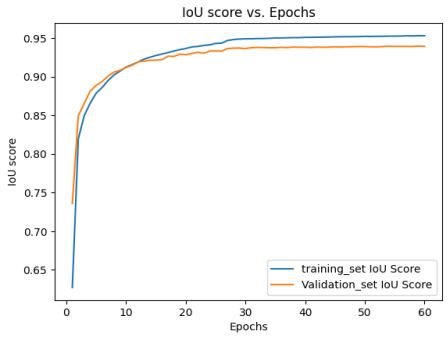


Figure 18: IoU score vs Epoch - Single cell

The test data performance provided a dice loss of 0.031, an IOU score of 0.938, and an accuracy of 95%. The image, ground truth, and segmented nuclei are shown in Figure 19 and the evaluation metrics of the test set are given in Table 1 for the performance analysis.

6.2 Phase 2

Unsupervised learning

The proposed model didn't provide a good result, but a pattern of the nuclei in the image was obtained. The models' performance was measured using metrics of dual consistency loss. The model provided a training performance with a loss of 0.0286, and a validation performance with a loss of 0.0285.

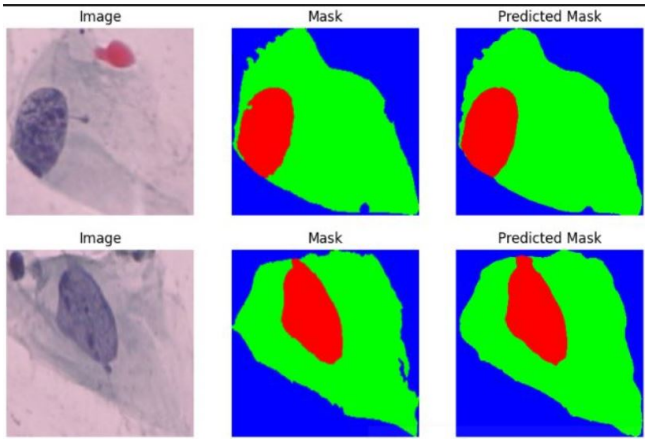


Figure 19: Segmented Nuclei

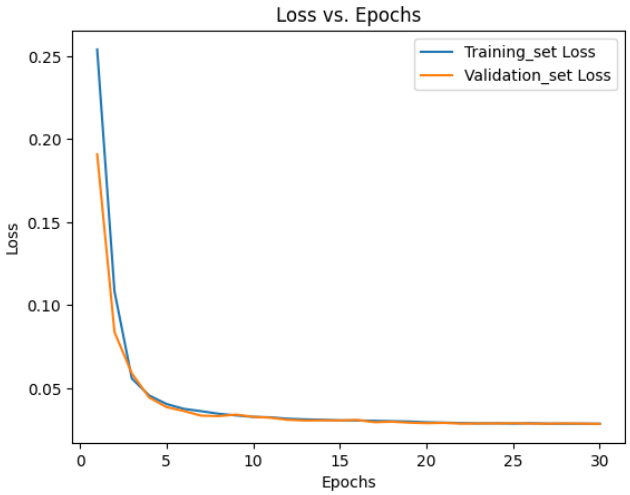


Figure 20: Loss Vs Eppoch- Multicell nuclei (unsupervised)

The loss of train and validation set over the epochs are shown in Figure 20

The test data performance provided a loss of 0.03. The image and predicted mask are shown in Figure 21 for the performance analysis.

Semi-supervised learning

Even though a semi-supervised approach was used to enhance the performance, the opposite occurred. The predicted mask was resembling the features of the single-cell masks and thus desired performance was not obtained. This model's performance was measured using the total loss. The model provided a training loss of 0.37 and a validation loss of 0.38.

The loss of train and validation set over the epochs are shown in Figure 22.

The test data performance provided a loss of 0.3. The image and predicted mask are shown in Figure 23 for the performance analysis.

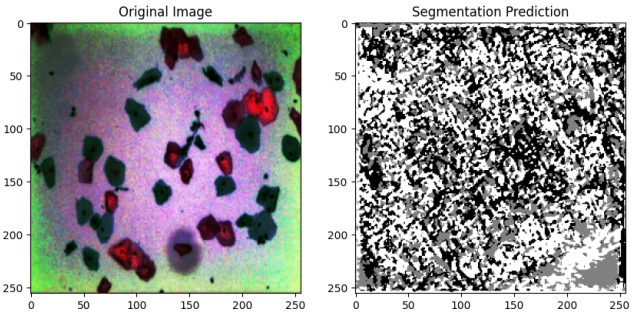


Figure 21: Image and its predicted mask (unsupervised)

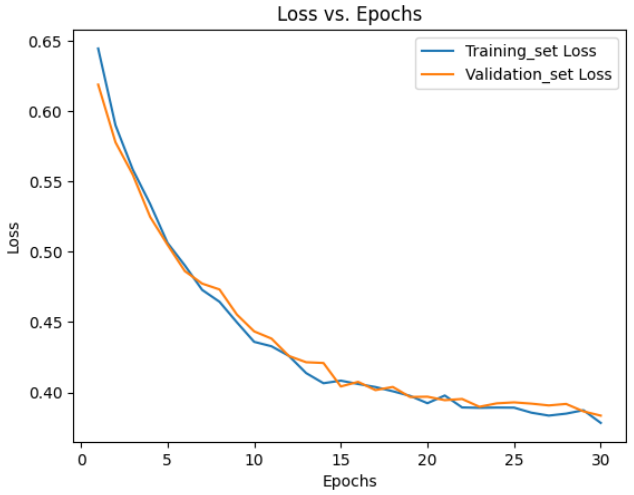


Figure 22: Loss Vs Eppoch- Multicell nuclei (semi-supervised)

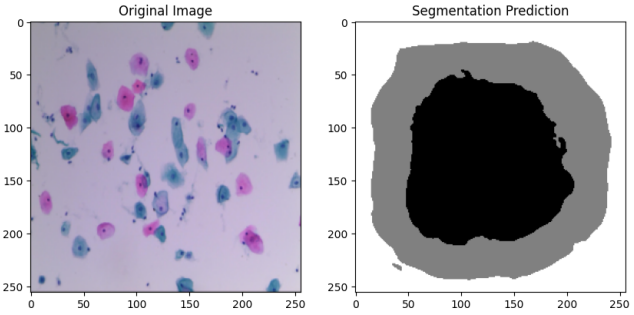


Figure 23: Image and its predicted mask(semi-supervised

7 DISCUSSION

The proposed model Unet with efficientNetb0 architecture provided good performance in segmenting the nuclei in the single slide image. The masks in the pap smear dataset only represented single nuclei, but some of the images contains more than one nuclei. The proposed model did not segment multiple nuclei within a single cell, but performed well according to the ground truth provided with an accuracy of 95%. The model trained on the single-celled pap smear dataset was then used to segment nuclei in multicell images of the CCEDD dataset by adapting some changes in the proposed system. But this didn't work as expected. The first experiment of unsupervised learning provided some patterns of the images and thus to improve the performance half of the dataset was labelled using the ground truth of the single-cell dataset. This approach also didn't give a good performance instead the predicted output had a greater influence of the single-cell nuclei mask. Thus to improve the first experiment the availability of labelled data for some images is required. The one with the use of single-cell masks didn't work well with the proposed model, so another method can be generating the masks of the multicell dataset using the edge-detected images that were available within the dataset. By using the exact masks of the images in the dataset may yield a good result for the semi-supervised approach.

8 CONCLUSION

Early detection and appropriate treatment are key to lowering the risk of cervical cancer. So the detection of the presence of mutated cells performs a significant role in treating this prevalent form of cancer in women. The performance of the implemented model the U-Net with EfficientNetb0 encoder for the single cell dataset indicates their potential to improve the efficiency and accuracy of cervical cancer screening. This is especially significant given that manual screening by pathologists is a time-consuming and challenging process due to the limited availability of certified pathologists. By automating the segmentation of nuclei from cervical cell images, the proposed model can improve the performance of automated cervical cancer detection, leading to faster and more accurate diagnoses.

The reported results for the single-cell dataset were obtained on the preprocessed masks, as the model could not segment the unidentified grey areas that were only present in some of the ground truth images. Additionally, the ground truth segmentation masks were only provided for individual nuclei, despite the presence of multiple nuclei in the input images. Therefore, future work could explore improving the model to segment multiple nuclei while also addressing the challenge of overlapping nuclei. For the case of multicell nuclei segmentation, the performance of the proposed approach did not meet the expected level. To enhance the performance through the semi-supervised approach, a potential future work is to generate masks for a subset of the dataset using the edge-detected images and to augment the model by incorporating attention layers. This approach aims to improve the segmentation results and address the challenges faced in accurately segmenting multicell nuclei.

REFERENCES

[1] Qaiser T. Raza S. E. Rajpoot N. M. Bashir, R. M. 2023. In *Consistency Regularisation in Varying Contexts and Feature Perturbations for Semi-Supervised Semantic Segmentation of Histology Images.*

[2] Tuan Le Dinh, Suk-Hwan Lee, Seong-Geun Kwon, and Ki-Ryong Kwon. 2022. Cell Nuclei Segmentation in Cryonuseg dataset using Nested Unet with EfficientNet Encoder. In *2022 International Conference on Electronics, Information, and Communication (ICEIC).* 1–4. https://doi.org/10.1109/ICEIC54506.2022.9748537

[3] Fischer P. Ronneberger, O. and Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://arxiv.org/abs/1505.04597

[4] Mingxing Tan and Quoc Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.), Vol. 97. PMLR, 6105–6114.

https://proceedings.mlr.press/v97/tan19a.html

[5] Xiangde Luo||Jieneng Chen||Tao Song||Guotai Wang. 2023. Semi-supervised Medical Image segmentation through dual-task consistency. https://aaai.org/papers/08801-semi-supervised-medical-image-segmentation-through-dual-task-consistency/

[6] Jing Zhao, Yong-Jun He, Si-Qi Zhao, Jin-Jie Huang, and Wang-Meng Zuo. 2022. AL-Net: Attention Learning Network Based on Multi-Task Learning for Cervical Nucleus Segmentation. *IEEE Journal of Biomedical and Health Informatics* 26, 6 (2022), 2693–2702. https://doi.org/10.1109/JBHI.2021.3136568

[7] Jie Zhao, Quanzheng Li, Xiang Li, Hongfeng Li, and Li Zhang. 2019. Automated Segmentation Of Cervical Nuclei In Pap Smear Images Using Deformable Multi-Path Ensemble Model. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 1514–1518. https://doi.org/10.1109/ISBI.2019.8759262

[8] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, and A.C. Palmer. 1994. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Transactions on Medical Imaging* 13, 4 (1994), 716–724. https://doi.org/10.1109/42.363096