Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.14, No.1, 2024, pp.17-28

DOI: https://doi.org/10.51983/ijiss-2024.14.1.3916

Performance Analysis of Five U-Nets on Cervical Cancer Datasets

Priyadarshini Chatterjee[1], Shadab Siddiqui[2], Giuseppe Granata[3], Prasanjit Dey[4] and Razia Sulthana Abdul Kareem[5]
[1]Research Scholar, [2]Assistant Professor,
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana, India
[3]Associate Professor, Business Management and Marketing, University Mercatorum, Rome, Italy
[4]Research Scholar, ADAPT SFI Research Centre, Ireland
[5]Senior Lecturer, Department of Computer Science, University of Greenwich Old Royal Naval College, United Kingdom
E-mail: jinipriya@klh.edu.in, cseshadabsiddiqui@gmail.com, giuseppe.granata@unimercatorum.it,
prasanjit.dey@adaptcentre.ie, razia.sulthana@gre.ac.uk
(Received 7 December 2023; Revised 26 January 2024, Accepted 5 February 2024; Available online 15 February 2024)

Abstract - **Image segmentation is crucial for precise analysis and classification of biomedical images, especially in the realm of cervical cancer detection. The accuracy of segmentation significantly influences the efficacy of subsequent image classification processes. While traditional algorithms exist for image segmentation, recent advancements in convolutional neural networks, particularly U-Nets have showcased exceptional effectiveness, especially in the realm of biomedical imaging. This research focuses on evaluating the accuracy of various U-Net architectures applied to three distinct cervical cancer datasets i.e., DSB containing 1340 images, SipakMed containing 1849 images and Intel Images for Screening containing 2000 images datasets taken from 2018 Data Science Bowl. The investigated U-Net architectures comprise the fundamental U-Net, Attention U-Net, Double U-Net, Spatial Attention U-Net, and Residual U-Net. The performance of the u-nets is judged on the metrics: Recall, Precision, F1, Jaccard and Accuracy. It is observed that Basic U-Net on the DSB dataset provides highest value on these metrices and accuracy obtained is 96%. The reason of high accuracy for DSB dataset can be attributed to the contrast of the images which by using co-occurrence matrix is calculated as 145.**
Keywords: **Cervical Cancer, Segmentation, U-Nets**

I. INTRODUCTION

Cervical region malignancy stands as the second most prevalent cancer among women. The mortality rate associated with this ailment among the female population is also a cause for concern. Cervical cancer can be attributed to over 200 variations of the HPV virus (Zhang, L, 2017). These types of cause genital or anus cancer most popularly known as cervical cancer. If not diagnosed at the proper time, the infection becomes persistent and nagging and also can become the cause of death. The virus causes changes in cells in and around cervix. The worst part is the symptom of cervical cancer is not known at the preset of the cancer. There is a high probability that the infection can be passed to others (Marinakis, 2009), (Dong. N, 2020). If the cancerous cells are detected earlier then they can be surgically removed to stop the spreading of the infection further. However, the detection of the cancer is not easy and simple (Gao. F, 2021).

Cervical cancer is detected by scanning the internal of the vagina and acquiring the Pap smears. These Pap smears are then sent to laboratory for screening (Lakshmi, 2016), (Bora, K, 2016). This process is time consuming and at the same time expensive. Statistical data shows that cervical cancer is predominant amongst women of all strata of society. The women belonging to lower strata of society finds it extremely difficult to go ahead for the process of cervical cancer screening. So, the only solution remains is to provide some tools that are less expensive and requires less expertise (Tan, 2015), (Ha, 2022). The proposed work finds the accuracy of segmentation analysis by implementing five different u-nets on cervical cancer datasets. This work identifies the best u-net that gives highest accuracy on all the three data sets. The identified u-net can be more improvised and can be used for various purposes in making an automated tool related to cervical cancer. Following three paragraphs in the present section identifies the gaps of existing research that can be reduced by the findings of the present work. The present work is based on existing u-nets but tries to see the u-nets in the perspective of datasets. This study also underscores the significance of image quality, particularly the contrast of an image, in the process of image analysis and classification (Harada, 2021), (William, 2018).

Researchers in the past have tried to find out many automated tools for screening of cancer. There are also automated tools for screening of cervical cancer as well (Mayrand, 2007), (Koss, 1989). However, there is scope of improvement in all such automated tools. The improvement can be done in terms of cost, efficacy, time and various other metrices. The automated tools are based on different technological advancements in the field of engineering. One such domain with whose help we can make automated tool is Digital Image Processing and Deep Learning (Mat-Isa,2005), (Chen, 2013).

Digital Image processing has four important steps. These are 1. Acquiring images, 2. Pre-processing images, 3. Segmentation of image and 4. Classifying images. This paper deals with the third step in image processing i.e image

segmentation (Zhou, 2024), (AlMubarak, 2019). Accurate image segmentation stands as the critical cornerstone in the realm of digital image processing and deep learning. The tools for doing image segmentation have evolved greatly over past decades (Song, 2014). Initially researchers used to use.

Traditional image segmentation algorithms like watershed. However, as the field of engineering advanced, so did the techniques of image segmentation also become progressive and innovative (Mahato, 2023), (Biscotti, 2005). It was well understood by the researchers that the entire efficacy of automated tools in detecting a disease was lying on an accurate segmentation. So, the thoughts of the researchers got divulged in making accurate tools for image segmentation (Hu, 2019). This was the reason of involving deep learning in making tools for image segmentation. This is important in the paradigm of this paper to describe image segmentation. Image segmentation can be broadly defined as the highlighting the region of interest by keeping the background of an image same (Sornapudi, 2018), (Malm, 2013). This region of interest is further sent for classification so that the objective of the research can be achieved.

Researchers were concentrating on improving the accuracy of segmentation (Sankaranarayanan, 2004), (Sukumar, 2016). This led them to develop a specialized tool utilizing a convolutional neural network known as U-Net. As the name suggests, the structure is U shaped. The left segment of the "U" constitutes the encoding section, while the lower middle and right segments of the "U" structure form the decoding section. U-Net is a total convolutional network that can work on all kind of medical images like CT scans, USG scans, MRI scans. It requires lesser amount of data and can provide the results in lesser time (Salembier,2000), (Sun,2017). The computational efficiency is beyond comparison for such kind of FCN. The researchers were elated, and they kept on making improved versions of U-Nets over the years. This study aims to identify the most effective U-Net based on performance indices such as Jaccard index, accuracy, recall, precision and F1 score. This paper will provide a comparison analysis by implementing five u-nets on three different cervical cancer datasets (Felzenszwalb,2004).

II. REVIEW OF LITERATURE

In 2008, approximately six million cases of cervical cancer were reported globally. Cervical cancer was in its peak in the developing countries, majorly the type of cervical cancer that was diagnosed was invasive cervical cancer. Cervical cancer and its lesion are detected by means of Pap Smear Screening. The Pap test is a type of colposcopy to inspect the cervix. It also includes to inspect the smears acquired microscopically by a pathologist. Nevertheless, to say that lots of expertise and experience is the common trait of any pathologist to proceed with such kind of tests (Zhang, 2017), (Uijlings, 2013). In order to support a pathologist's diagnosis, automated tool for screening of any type of cancer is always useful.

The initial step in the pathologists' diagnosis involves examining the presence of cervical squamous intraepithelial lesion (SIL) in the atypical cells within the cervix's epithelial region. This is done using visual inspection of the Pap Smear slides. This CIN is also of three types: CIN1 (mid dysplasia), CIN2 (moderate dysplasia), CIN3 (severe dysplasia). Some slides for reference, considered directly from the contribution of Dr. Khaled J (Li, 2020), (Song, 2016). Alkhateeb, of normal atypical cells to severe dysplasia is included below in Fig.1.

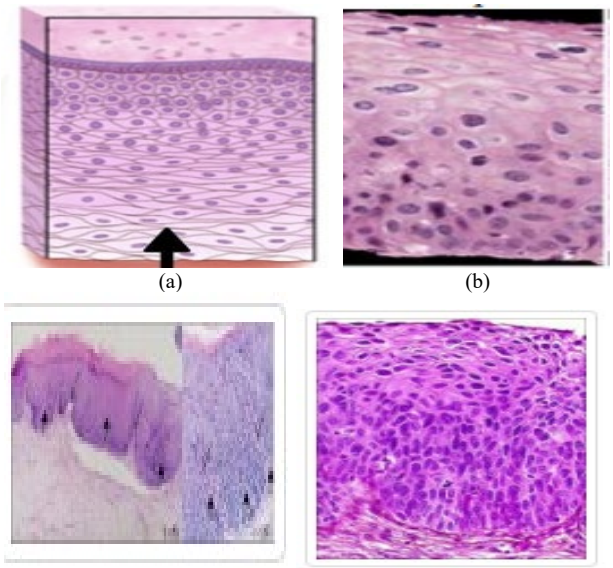

Fig. 1 (a) Normal atypical cells, (b) CIN1, (c) CIN2, (d) CIN3

We can see that as the CIN is gaining severity the epithelium is showing delay in maturity and there is also an increase of immature abnormal cells. Figures (b to d) clearly depicts the position of the immature atypical cells. In figure b (CIN1), the immature abnormal cells are in the one third part. In figure c (CIN2), the immature atypical cells are seen from above and covers almost two third of the image. Figure d (CIN3), shows entire image is covered with immature atypical cells (Van der Maaten, 2008), (Greenspan, 2008).

Image segmentation represents a crucial stage in digital image analysis, with the precision of image classification contingent upon accurate segmentation. The subsequent classification process relies on the correct delineation of the region of interest (Ojala, 2000), (Ronneberger, 2015). Conventional image segmentation has various types like a. image segmentation using edge detection, b. image segmentation using thresholding, c. background-based image segmentation, and d. image segmentation using clustering. These methods are traditional in nature, each with its own set of advantages and disadvantages. As mentioned earlier, the segmentation types are categorized based on two fundamental properties: similarity and discontinuity. When segmentation is done on the basis of discontinuity, the subdivision of the ROI is done on the basis of certain changes in the grey level of the pixels. This type mainly captures the secluded points, lines, and edges. In case of similarity-based

partition, the main focus is to group pixels of similar nature (Lu, 2020), (Jégou, 2017), (Bay, 2008). There are various algorithms like region growing, thresholding, merging and splitting. All of these approaches are conventional and are not that useful to find the ROI in a crowded cell in biomedical image processing.

U-Net structure based on the principles of convolutional neural network, is fairly adopted for performing bio medical image segmentation (Feng, 2019), (Ojala, 2002). The success of U-Net can be visualized on its adaptability on MRI images, CT scans, X-rays, and even microscopic images. One of the most important advantages of using U-Net is its computation time and requirement of less volume of data. With the advancement of deep learning in computer vision, U-Net can be extensively used in image analysis and segmentation. Researchers recommend using U-Nets especially in case of bio medical images, as the accuracy of any analysis is very important. U-Nets vouches to provide a high rate of accuracy (Ezzell, 2003), (Cortes, 1995).

Over the years, with the progress of medical sciences and computer vision, different variants of u-nets are developed by scientists (Biau, 2012), (Mackie, 2003). Depending on the type of disease to diagnose, the type of u-nets also changes. The researchers have always aimed to make the u-net structure more adaptable to the change in image types, so that there is no compromise on accuracy. Delving into the structure of U-Net, it primarily consists of two segments: the compressive path and the expansive path. Image analysis is performed by the compressive path, while image synthesis is the main function of the expansive path (Stapleford, 2010), (Yang, 2012), (Song, 2016). The U-Net's task involves merging the features extracted by the compressive path with the outcomes of the expansive segment. Fig. 2. below is the U-Net structure outlined that illustrates its dual components.

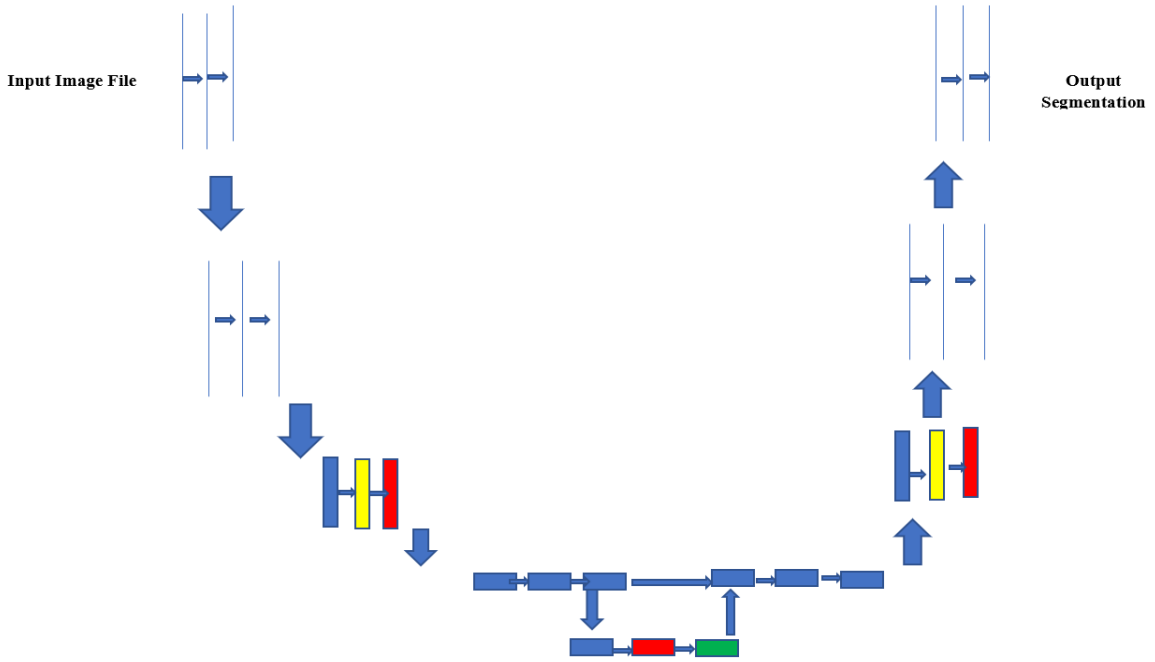

Fig. 2 Basic U-Net architecture

In the above figure, the coloured arrows represent different types of operations. The boxes at the bottom shows the feature maps that is achieved at each layer (Deng,2023), (Lakshmi,2013). Some of these feature maps are also cropped by the encoder part. U-Net provides a highly detailed segmentation which makes them useful in case of biomedical images. The U-Net accomplishes this by incorporating abrupt changes in elasticity within the training data. The network is designed to acquire knowledge of these changes autonomously, eliminating the requirement for pre-defined data (Bengtsson,2014), (Acosta-Mesa,2014). The U-net also applies overlapped objects that belong to the same class by applying weighted loss function. U-nets are faster to train than the other segmentation models. The U-net has evolved since its inception and there are various types of u-nets available for particular diseases. This paper provides a performance analysis taking into consideration five types of U-nets applied on cervical cancer datasets. While U-Net comes with notable advantages, it is important to acknowledge that, like any model, it is not without its drawbacks. Let's explore some of the limitations associated with the U-Net architecture:

1. High Memory Requirements
2. Fixed Input Size
3. Class Imbalance
4. Limited Feature Learning

The next section of this paper provides the readers with the metrics achieved after the implementation of the U-nets on the datasets.

III. METHODOLOGY

A. Data Description

The benchmark datasets used in this paper contains segmented images of nucleus. These images are acquired under varying conditions. These images also have different cell types, different quality like fluorescence and brightness. These datasets are designed in such a way that they can be used for testing using different algorithms. This paper does not focus on addressing class imbalance; rather, its primary objective is to investigate the influence of image contrast on segmentation accuracy. Moreover, as the datasets are secondary, there is no reference available on the data imbalance. There is a unique image id for each image in the data set. There are two folders, image, and mask. The image folder contains images, and the mask folder contains the segmented masks of the nucleus. The mask folder contains non overlapping images and are used for training. Table I below is about the number of images in three cervical cancer data sets used in this paper.

TABLE I NUMBER OF IMAGES IN THE SECONDARY DATASETS

Name of the Dataset	Number of Images
DSB	1340
SipakMed	1849
Intel Images	2000

B. Pre-Processing

As these datasets are bench marked dataset, so very little pre-processing is done. It is usually seen that when Pap smear stains are transferred into digital image, the moisture around the cervix is visible as white spots in the image. These spots try to camouflage the actual feature of interest. These spots also affect the segmentation process and reduces the accuracy of segmentation. Hence, it becomes very important to remove these spots. The process for doing this is given by the algorithm below.

Algorithm

1. The image of the cervix is separated into three planes, red, green, and blue.
2. The white pixels are selected and a "logical and" is done with the appropriate structuring element.
3. This gives a dilated image whose borders are converted into natural grey scale.
4. An interpolating algorithm is selected that interpolates from within the pixel to the boundary using Laplace's equation to find Δy and -Δy.
5. We select Laplace's equation, so that it provides the best interpolant.
6. We are trying to prove and find y(x) in Eq. (1) given below,

$$y(x) = \begin{cases} c_{12} \ln|x| + c_{22} & m = 2 \\ \frac{c_{12}}{(2-m)\,|x|^{m-2}} + c_{22} & m \geq 3 \end{cases} \quad (1)$$

As y(x) is a harmonic function, we are trying to prove on Rm y(x) = z(|x|). Above equation (1) on Rm provides Δy(x) = 0 for x=0 and for x not equal to 0, Δy(x) is undefined, irrespective of the value of constants $c1$ and $c2$.

C. U-Nets on Three Cervical Cancer Datasets

Selecting an architectural framework for segmenting biomedical images relies on the unique attributes of the data and task requirements. Numerous adaptations and advancements have been suggested to enhance the efficacy of the U-Net structure in the context of medical image segmentation. Previous studies addressing the segmentation of densely populated cells commonly employed the fundamental U-Net architecture, subsequently introducing variations to enhance accuracy. In the specific context of this paper, which utilizes Pap Smear stains and colposcopy images as datasets, prior research also utilized the basic U-Net on these datasets for segmentation. Additionally, the residual U-Net was employed for colposcopy image segmentation. However, the spatial attention U-Net, double U-Net, and attention U-Nets were not utilized in the analysis of cervical cancer datasets. The double U-Net architecture allows for a hierarchical extraction of features. The first U-Net can capture low-level features and spatial information, while the second U-Net can focus on learning high-level features and contextual information based on the features obtained from the first stage. The choice of the attention U-Net is primarily attributed to its gating mechanism, contributing to improved network learning. In Res U-Net skip connections allow the direct flow of gradients during backpropagation. This helps in mitigating the vanishing gradient problem, which can occur in deep neural networks. Spatial attention mechanisms enable the network to dynamically weigh different spatial locations based on their relevance to the task at hand. This improves the model's ability to focus on important regions and suppress irrelevant or distracting features. The main objective of this paper is to conduct a comparative study based on five metrics by selecting five U-Nets applicable to cervical cancer datasets, ultimately proposing a U-Net with the highest accuracy. It should be noted that the proposal of the U-Net with the highest accuracy is contingent on the outcomes of these five U-Nets and falls outside the scope of this paper.

D. Basic U-Net

The U-Net, characterized by its U-shaped structure, relies on a deep convolutional network. This structure consists of two essential elements referred to as the contracting pathway and the expanding pathway. The compressing pathway comprises successive 3×3 convolution operations, succeeded by a ReLu activation function and subsequently followed by a max-pooling operation. The above arrangement is repeated many a times. The expanding path or the decoder does the up-sampling the features using 2×2 up convolution. The corresponding feature from the encoder is cropped and added with the features of the decoder.

Before producing the output, a 1×1 convolution is utilized to reduce the number of feature maps to the required channel count. The cropping of the feature map present in the encoder path is required to be cropped as they contain least amount of contextual information. Furthermore, the connections that skip from the encoder to the decoder necessitate a U-shaped architecture, enabling the flow of contextual information both downward from the encoder and upward to the decoder. The U-net uses an energy function defined by Eq. (2) below.

$$\text{Energy} = \sum l(x) \log\left(t_{n(x)}(x)\right) \qquad (2)$$

In Eq. (2) above tn is the soft-max function calculated pixel wise on the final feature map. Table II below shows the result of implementing the basic U-Net on the datasets. The experiment is carried out after preprocessing the images under a software environment of Keras version 3.0x and Tensorflow version 2.7.0. The algorithm was run till 30 epochs and the total number of trainable parameters obtained were 29,298,600. The algorithm is divided into four modules, model, train, test, and eval. All the datasets are made into two folders, train, and test. Within the "train" directory, there are two subfolders: "images" and "mask." Training utilizes 80% of the total images, while the remaining 20% is allocated for testing. The results captured as five metrics is simulated in the Table II below. It is evident from observing the results that the implementation of the basic U-Net yields the highest accuracy for DSB across the three datasets.

TABLE II RESULT OF BASIC U-NET ON THREE CERVICAL CANCER DATASETS

DataSet	F1	JACCARD	RECALL	Precision	Accuracy
DSB	0.85	0.76	0.88	0.86	0.96
SipakMed	0.86	0.74	0.85	0.83	0.93
Intel Images	0.84	0.73	0.85	0.84	0.92

1. Graphical Analysis of the Result of the Basic U-Net

Fig. 3 below is a plot of the metrics (scores) i.e., F1, Jaccard, recall, precision and accuracy labelled in the y-axis and the three different datasets labelled in the x-axis. The accuracy is highlighted in violet colour, and it can be visualized from the graph below that for DSB datasets, the accuracy is the highest.

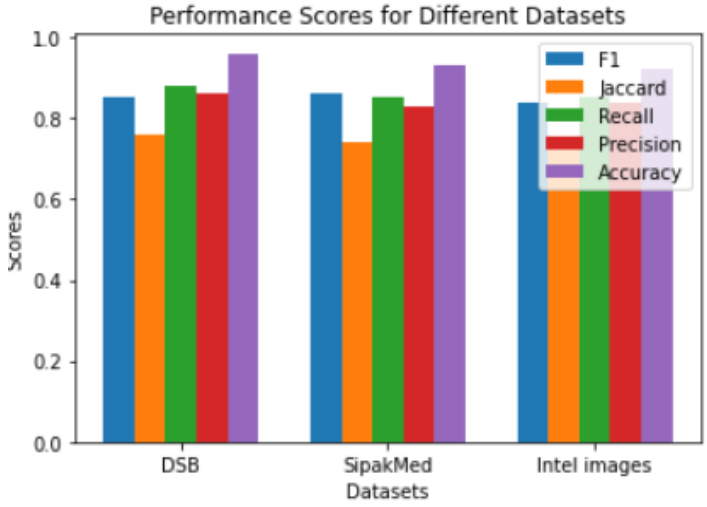

Fig. 3 Graphical Analysis of basic U-Net

D. Double U-Net

The double U-Net is another powerful architecture based on dense net. A dense network of skip connections is present between the input and output stages. The exchange of semantic information throughout the network is facilitated by the dense skip connection linking the encoder and the decoder.

It is called double U-Net as it as one skip connection per layer. The skip connections receive feature maps from higher layers, along with up-sampled features from the layer directly below. Thus, we can consider each level as a dense block. This kind of arrangements helps to retain the semantic information between the layers. Eq. (3) below depicts the work of the skip connection.

$$y^{m,n} = \begin{cases} W_o(y^{m-1n}) & n = 0 \\ W_o\left(\left[y_{k=n}^{m,l}\right]^{n-1} l_o(y^{m+1,n-1})\right) & n > 0 \end{cases} \qquad (3)$$

In the Eq. (3) above, y represents the feature map, and n, m are the indices of the encoder path and the skip connection as well. Here (W_o) is the convolution operation and (l_o) denotes the up-sampling operation. The number of skip connections as intermediate layer depends on the number of layers present.

This U-Net is implemented on the three cervical cancer datasets and the result of the implementation is given in Table III. below. Four modules are formed for running this algorithm i.e., model.py, eval.py, test.py and train.py. The total number of parameters after executing this model are 28,297,280. The count of learnable parameters is 29,290,200, while the number of fixed parameters is 8,296. The model is run for 30 epochs. The software environment remains similar to the working of the basic U-Net mentioned in description of Basic U-Net above. The results captured as five metrics is simulated in the Table III below. It can be seen that the accuracy of DSB is the highest after implementation of Double u-net on three datasets.

TABLE III RESULT OF DOUBLE U-NET ON THREE CERVICAL CANCER DATASETS

DataSet	F1	JACCARD	RECALL	Precision	accuracy
DSB	0.82	0.72	0.84	0.82	0.91
SipakMed	0.72	0.65	0.73	0.75	0.84
Intel Images	0.74	0.65	0.75	0.74	0.85

1. Graphical Analysis of the Result of the Double U-Net

Fig. 4 below is a plot of the metrics (scores) i.e., F1, Jaccard, recall, precision and accuracy labelled in the y-axis and the three different datasets labelled in the x-axis. The accuracy is highlighted in violet colour, and it can be visualized from the graph below that for DSB datasets, the accuracy is the highest.

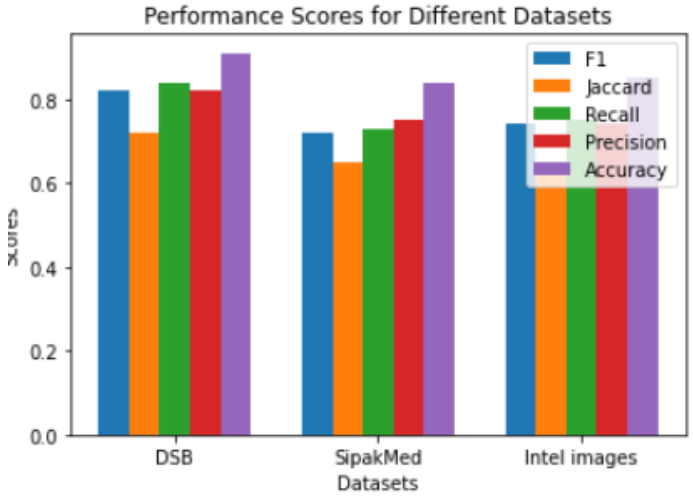

Fig. 4 Graphical Analysis of Double U-net

E. Residual U-Net

The main motive behind a ResNet is to minimize the hinderances in training a deep convolutional network. Deep neural network has an advantage that they converge to a solution faster, but it is also seen that these networks have performance degradation as the number of layers are increased. Their performance gets saturated. The deep neural networks suffer from a problem of vanishing gradient problem in their weights and thus lose their feature identities in the deeper layers. ResNet tries to reduce the vanishing gradient problem by considering the features from upper layer and adds it to the next layer. This way the network preserves the feature maps in a much better way.

In a residual U-Net, each block integrates skip connections, where the input to one layer is combined with the output from the preceding layer. These skip connections are implemented in both the reduction path and the expansion path. The residual block can be denoted by the following Eq. (4) and Eq. (5) described below.

$$x_1 = {}^{h}_{0}(y_1) + {}^{\zeta}_{0}(y_1, A1) \quad (4)$$

$$y_{l+1} = f_{0(x_1)} \quad (5)$$

In the above equation y_l and y_{l+1} are the input and output of the ResNet. ζ_o is the residual function, f_o is the activation function and h_o is the identity mapping function. All the images in the train and test folders along with the number of images used as a mask is in the ratio of 70% and 30%. Total number of trainable parameters are 1,900,106 and there are no trainable parameters obtained after implementing the algorithm on the three cervical cancer datasets when run for 30 epochs. Table IV below shows the implementation of ResNet on the cervical cancer datasets, keeping the execution environment similar to the execution environment mentioned in section above. It can be seen that the accuracy of DSB is the highest after implementation of ResNet on three datasets.

TABLE IV RESULT OF RESNET ON THREE CERVICAL CANCER DATASETS

DataSet	F1	JACCARD	RECALL	Precision	Accuracy
DSB	0.85	0.77	0.86	0.82	0.93
SipakMed	0.73	0.76	0.74	0.84	0.85
Intel Images	0.82	0.75	0.87	0.79	0.91

1. Graphical Analysis of the result of the ResNet:

Fig. 5. below is a plot of the metrics (scores) i.e., F1, Jaccard, recall, precision and accuracy labelled in the y-axis and the three different datasets labelled in the x-axis. The result is after implementation of ResNet. The accuracy is highlighted in violet colour, and it can be visualized from the graph below that for DSB datasets, the accuracy is the highest.

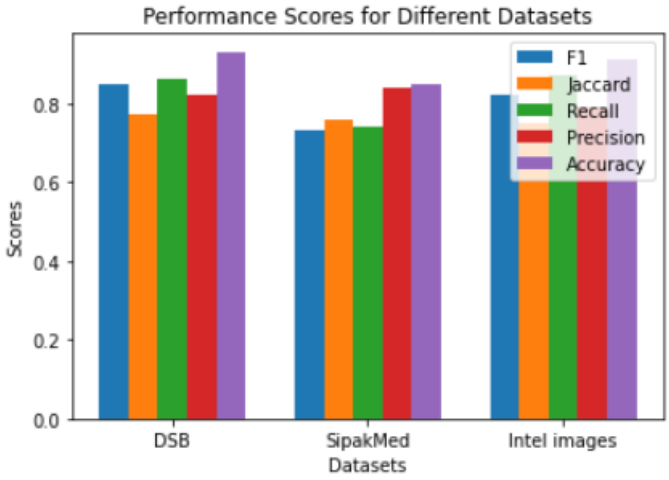

Fig. 5 Graphical Analysis of ResNet

F. Attention U-Net

It is always important in case of digital image processing to concentrate on specific object of interest, and it is required to ignore the unnecessary objects. This important property of the image processing is achieved by the attention u-net by using attention gates. The attention gate crops the features that are not important and preserves the features which are of utmost important. The layers at the expanding path have attention gate. These gates forward the features from the encoder path. This step is executed prior to combining the features with the up-sampled features of the expanding path.

The gates are utilized iteratively after each layer, contributing to enhanced segmentation performance while maintaining low computational complexity. The attention part acts as an encoder-decoder model, and it helps in localized classification information. If the training data is labelled, it tunes the specific objects in the image. In case of digital image processing, the additive gates are used in attention u-net making it more powerful in terms of accuracy. The additive attention gate can be described by Eq (6) and (7) below.

$$r_{att}^{t} = \psi^{T}\left(\sqcap_{1}\left(W_{x}^{T}\alpha_{i}^{1} + W_{g}^{T}\sigma_{i} + b_{g}\right)\right) + b_{\psi} \quad (6)$$

$$\lambda_{i}^{1} = \sqcap_{2}\left(r_{att}^{t}(\alpha_{i}^{1}, \sigma_{i}; \Theta_{att})\right) \quad (7)$$

In the above equation αi^{l} is the feature of the encoder path and g is denoted as the gating signal. Below, Table V presents the outcomes of the attention U-Net applied to the three cervical datasets. The software used are similar to that of the other u-nets described above. The model is run for 30 epochs with total number of parameters being 1,900,104 and total number of trainable parameters are 1,88,100 and number of non-trainable parameters being 2500. The images are divided into test and train folders as described in the above sections.

Table V below shows the simulation of attention u-net on three cervical cancer datasets. As we see in the below table, DSB dataset shows highest accuracy.

TABLE V RESULT OF ATTENTION U-NET ON THREE CERVICAL CANCER DATASETS

DataSet	F1	JACCARD	RECALL	Precision	Accuracy
DSB	0.83	0.70	0.88	0.82	0.92
SipakMed	0.72	0.73	0.73	0.78	0.84
Intel Images	0.83	0.73	0.81	0.74	0.78

1. Graphical Analysis of the Result of the Attention U-Net

Fig. 6 below is a plot of the metrics (scores) i.e., F1, Jaccard, recall, precision and accuracy labelled in the y-axis and the three different datasets labelled in the x-axis. The result is after implementation of Attention U-Net. The accuracy is highlighted in violet colour, and it can be visualized from the graph below that for DSB datasets, the accuracy is the highest.

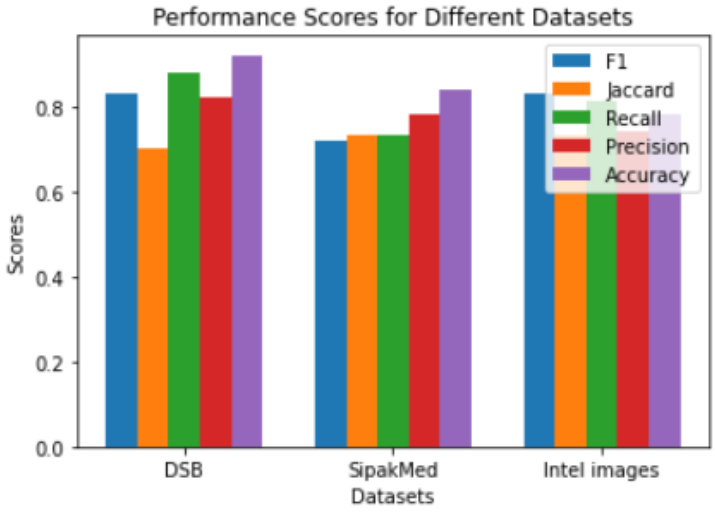

Fig. 6 Graphical Analysis of Attention U-Net

G. Spatial Attention U-Net

The spatial attention U-Net employs a U-shaped layout, with the decoder situated on the right side and the encoder on the left side. At each stage, it incorporates a fusion of a convolutional block and a 2×2 down sampling operation. The convolutional layer within the block contains a drop block, a rectified linear unit, and a batch normalization operation. The reduction in resolution is accomplished through the down sampling operation with a stride of 2.

Additionally, a spatial attention module is incorporated, leveraging the spatial relationships among features to generate a mapping. In order to calculate the spatial components, firstly max pooling is applied followed by mean pooling operation along the axis of the channels. These operations are then added to output clear feature descriptor. Then a convolution operation is applied along with sigmoid activation on the feature descriptors to produce the attention map. The output of the spatial attention net can be described by Eq. (8),(9) and (10) below. The software requirements for implementing this u-net are similar to the nets described above.

The total parameters obtained are 537,707, out of which total trainable parameter is 536,707 and total nontrainable parameter is 1409. The model is run for 30 epochs and the images are divided into test and train folders in a ratio of 70: 30.

$$SA^s = SA.N^s\,(SA) \quad (8)$$

$$= SA_0\Theta(e^{7\times7}\big(MaxPool(SA); AvgPool(SA)\big)) \quad (9)$$

$$= SA_0\Theta(e^{7\times7}(SA^S_{MP}; SA^S_{AP})) \quad (10)$$

In Eq. (9) and (10) above, e7×7 is a filtering operation with a filter size of 7and $\Theta(_o)$ is a sigmoid function.

Table VI below shows the simulated results of this U-net on three cervical cancer data sets. We see below, the DSB dataset provides highest accuracy.

TABLE VI RESULT OF SPATIAL ATTENTION U-NET ON THREE CERVICAL CANCER DATASETS

DataSet	F1	JACCARD	RECALL	Precision	Accuracy
DSB	0.87	0.88	0.87	0.87	0.94
SipakMed	0.85	0.80	0.81	0.82	0.89
Intel Images	0.80	0.79	0.81	0.80	0.85

1. Graphical Analysis of the result of the Spatial Attention U-Net:

Fig. 7 below is a plot of the metrics (scores) i.e., F1, Jaccard, recall, precision and accuracy labelled in the y-axis and the three different datasets labelled in the x-axis. Clearly it is seen from the plot that DSB data set shows much promising result than the other two dataset in terms of accuracy and the other metrics.

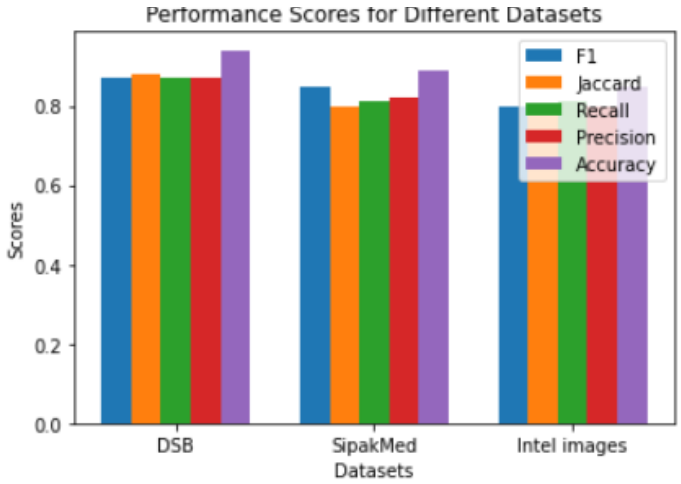

Fig. 7 Graphical Analysis of Spatial Attention U-Net

IV. U-NET ON THREE CERVICAL CANCER DATASETS

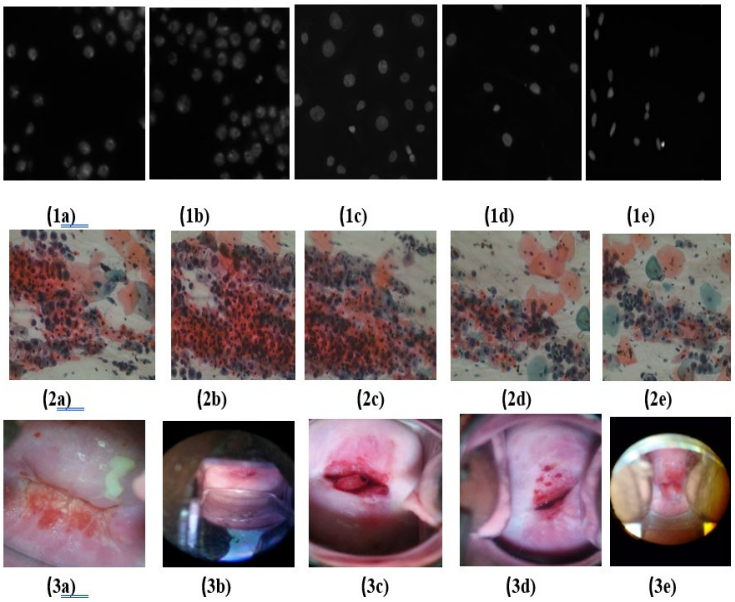

Fig. 8 Output Images

Fig. 8. shows the output images of three datasets after implementing the five U-nets. These images are segmented images which clearly shows the nucleus spread over the region.

Fig. (1a to 1e). is the resultant image obtained in the train folder for DSB data set. Fig. 1a. is the output of basic U-net, figure 1b is the output of double U-net, figure 1c is the output of residual u-net, Fig.1d. is the output of attention u-net and Fig. 1e. is the result of SA u-net.

Fig. (2a to 2e). is the output image obtained in the train folder for SipakMed data set. Fig. 2a. is the resultant of basic U-net, Fig. 2b. is the result of double U-net, Fig. 2c. is the resultant of residual u-net, Fig. 2d. is the result of attention u-net and Fig. 2e. is the output of SA u-net.

Fig. (3a to 3e). is the resultant image obtained in the train folder for Intel data set. Fig. 3a. is the outcome of basic U-net, Fig. 3b. is the effect of double U-net, Fig. 3c. is the outcome of residual u-net, Fig. 3d. is the output of attention u-net and Fig. 3e. is the output of SA u-net.

V. RESULT ANALYSIS

Table VII below shows the combined result of simulation of five u-nets on the three cervical cancer secondary data set. The table clearly indicates that the DSB dataset exhibits the highest level of accuracy. As this paper does a comparative study of five u-nets and tries to identify the best u-nets in terms of accuracy, Therefore, accuracy emerges as a pivotal factor in the choice of the U-Net.It can be seen that the accuracy of DSB dataset after implementation of five u-nets is the highest comparative to the other two datasets.

TABLE VII RESULT OF FIVE U-NETS ON THREE CERVICAL CANCER DATASETS

U-Net	Data Set	F1	JACCARD	RECALL	Precision	Accuracy
Basic U-Net	DSB	0.85	0.76	0.88	0.86	0.96
	SipakMed	0.86	0.74	0.85	0.83	0.93
	Intel Images	0.84	0.73	0.85	0.82	0.92
Double U-Net	DSB	0.82	0.72	0.84	0.82	0.91
	SipakMed	0.72	0.65	0.73	0.75	0.84
	Intel Images	0.74	0.65	0.75	0.74	0.85
Res U-Net	DSB	0.85	0.77	0.86	0.82	0.93
	SipakMed	0.73	0.76	0.74	0.84	0.85
	Intel Images	0.82	0.75	0.87	0.79	0.91
Attention U-Net	DSB	0.83	0.70	0.88	0.82	0.92
	SipakMed	0.72	0.73	0.73	0.78	0.84
	Intel Images	0.83	0.73	0.81	0.74	0.78
SA U-Net	DSB	0.87	0.88	0.87	0.87	0.94
	SipakMed	0.85	0.80	0.81	0.82	0.89
	Intel Images	0.80	0.79	0.81	0.80	0.85

As discussed above, at the time of converting Pap Smear stains into digital images, the moisture around cervix comes as white spots in the images. For image segmentation, it is very important to remove these spots. Moreover, removing these bright spots also improves the texture of the image. Texture is the spatial arrangement of the intensity values. The region that needs to be segmented, should be texturally accurate and not only bright. This paper focuses on how to capture the texture of image datasets and how texture affects the accuracy of segmentation. The grey level co-occurrence matrix approach taken from Mudassir Rafi et al; finds out the grey level matrix of each data set. This paper tries to find the size of the matrix in a given threshold. The matrix size is contingent upon the grey level of the image. The size is normalized using the formula given in Eq. (11) below.

$$M(x,y) = f' \frac{f'(x,y)}{\sum_{i=0}^{L} \times \sum_{j=0}^{L} f'_{(i,j)}}, x, y = 0 \ldots\ldots L \quad (11)$$

In the given equation, 'M (x, y)' represents the normalized matrix. f'(i, j) denotes the count of pixel pairs with intensity values zi and zj. These pixels take part in computing the size of the matrix. This normalized matrix of the colour images is converted into greyscale images using OpenCv. The Co-occurrence matrix calculated above is used to calculate 14 statistical methods. Amongst the 14 statistical methods, this paper uses the calculation of contrast of three image data set. The intensity of the pixels of the coloured images are considered as the new values obtained after the colour images are converted into grey scale images. The formula for finding contrast is given in Eq. (12) below.

$$C = \sum_{m=1}^{l} \sum_{n=1}^{l} (m-n)^2 \, N_{(m,n)} \quad (12)$$

The value of C gives the contrast of whole image of three data sets. The value of contrast is calculated by selecting fifty images from each data set. Table VIII below shows the contrast of three datasets and average of the contrast.

TABLE VIII CONTRAST AND AVERAGE OF THREE DATASETS

Data Sets	Contrast Value (C)	Average (50 images)
DSB	145	2.9
SipakMed	137	2.74
Intel	129	2.58

For this paper we have considered a threshold of 3 for the value of C. It is seen that DSB data set is having a value of 2.9 that is almost close to the threshold value. The other two data sets have a lower contrast value than the DSB data set. Given that it is also substantiated that texture significantly contributes to the precision of the segmentation operation, moreover, we have found using co-occurrence matrix the value of the contrast of three datasets, DSB dataset having highest value of C, we can justify that texture of the images and contrast is the main contributing factors in obtaining the highest accuracy value of segmentation.

VI. CONCLUSION

The inference drawn from this study is that image quality holds a crucial role in enhancing the precision of digital image segmentation. The motivation behind this research stems from exploring the potential use of GLCM as a texture descriptor for images. Furthermore, it is observed that the contrast of the image significantly influences the accuracy of image segmentation. Upon implementing five U-Nets on three cervical cancer datasets, a consistent observation emerged, indicating that the DSB dataset consistently exhibited the most promising segmentation results across all five U-Nets. The calculated average contrast value for the DSB dataset is 2.9, as determined from the normalized co-

occurrence matrix in the aforementioned Section 3.5. The co-occurrence matrix effectively captures and reflects the texture of the image with precision. The "C" value for the DSB dataset closely approaches the threshold value and surpasses that of the other two datasets. These findings lead us to the conclusion that texture, and contrast are integral factors in the analysis of digital image segmentation. This work does not focus on the class imbalance of the dataset as class imbalance is required for image classification. Out of five metrics, this work focuses on accuracy and makes accuracy as a predominant factor for choosing a particular net for cervical cancer images. In the future we are in the process of making a new U-Net and we would also like to gather real time images so that they can be tested with the new U-net that we design. It is important to say at this juncture that the five u-nets have their own best features. The new u-net that we will design will be architecture taking into consideration the best features of these five u-nets and also considering the results obtained in this paper, so that our ultimate aim to increase the accuracy by reducing the number of trainable parameters can be achieved.

Data Accessibility

1. DSB data set is available 2018 Data Science Bowl. The link for the data is: https://www.kaggle.com/competitions/data-science-bowl-2018/data.
2. SipakMed data is the largest cervical cancer dataset maintained by Kaggle. The data is available in the following link: https://www.kaggle.com/datasets/prahladmehandiratta/cervical-cancer-largest-dataset-sipakmed.
3. Intel & Mobile ODT Cervical Cancer Screening dataset is also maintained by Kaggle. The data is available in the following link: https://www.kaggle.com/competitions/intel-mobileodt-cervical-cancer-screening/data.

REFERENCES

[1] Acosta-Mesa, H. G., Rechy-Ramírez, F., Mezura-Montes, E., Cruz-Ramírez, N., & Jiménez, R. H. (2014). Application of time series discretization using evolutionary programming for classification of precancerous cervical lesions. *Journal of Biomedical Informatics, 49*, 73-83.

[2] AlMubarak, H. A., Stanley, J., Guo, P., Long, R., Antani, S., Thoma, G., ... & Stoecker, W. (2019). A hybrid deep learning and handcrafted feature approach for cervical cancer digital histology image classification. *International Journal of Healthcare Information Systems and Informatics (IJHISI), 14*(2), 66-87.

[3] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding, 110*(3), 346-359.

[4] Bengtsson, E., & Malm, P. (2014). Screening for cervical cancer using automated analysis of PAP-smears. *Computational and Mathematical Methods in Medicine,* 2014.

[5] Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research, 13*, 1063-1095.

[6] Biscotti, C. V., Dawson, A. E., Dziura, B., Galup, L., Darragh, T., Rahemtulla, A., & Wills-Frank, L. (2005). Assisted primary screening using the automated ThinPrep Imaging System. *American Journal of Clinical Pathology, 123*(2), 281-287.

[7] Bora, K., Chowdhury, M., Mahanta, L. B., Kundu, M. K., & Das, A. K. (2016, December). Pap smear image classification using convolutional neural network. In *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing* (pp. 1-8).

[8] Chen, Y. F., Huang, P. C., Lin, K. C., Lin, H. H., Wang, L. E., Cheng, C. C., ... & Chiang, J. Y. (2013). Semi-automatic segmentation and classification of pap smear cells. *IEEE Journal of Biomedical and Health Informatics, 18*(1), 94-108.

[9] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273-297.

[10] Deng, Y., Du, S., Wang, D., Shao, Y., & Huang, D. (2023). A calibration-based hybrid transfer learning framework for RUL prediction of rolling bearing across different machines. *IEEE Transactions on Instrumentation and Measurement, 72*, 1-15.

[11] Dong, N., Zhao, L., Wu, C. H., & Chang, J. F. (2020). Inception v3 based cervical cell classification combined with artificially extracted features. *Applied Soft Computing, 93*, 106311.

[12] Ezzell, G. A., Galvin, J. M., Low, D., Palta, J. R., Rosen, I., Sharpe, M. B., ... & Yu, C. X. (2003). Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee. *Medical Physics, 30*(8), 2089-2115.

[13] Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision, 59*, 167-181.

[14] Feng, X., Qing, K., Tustison, N. J., Meyer, C. H., & Chen, Q. (2019). Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. *Medical Physics, 46*(5), 2169-2180.

[15] Gao, F., Liu, S., Zhang, X., Wang, X., & Zhang, J. (2021). Automated Grading of Lumbar Disc Degeneration Using a Push-Pull Regularization Network Based on MRI. *Journal of Magnetic Resonance Imaging, 53*(3), 799-806.

[16] Greenspan, H., Gordon, S., Zimmerman, G., Lotenberg, S., Jeronimo, J., Antani, S., & Long, R. (2008). Automatic detection of anatomical landmarks in uterine cervix images. *IEEE Transactions on Medical Imaging, 28*(3), 454-468.

[17] Ha, A. Y., Do, B. H., Bartret, A. L., Fang, C. X., Hsiao, A., Lutz, A. M., ... & Hurt, B. (2022). Automating scoliosis measurements in radiographic studies with machine learning: Comparing artificial intelligence and clinical reports. *Journal of Digital Imaging, 35*(3), 524-533.

[18] Harada, G. K., Siyaji, Z. K., Mallow, G. M., Hornung, A. L., Hassan, F., Basques, B. A., ... & An, H. S. (2021). Artificial intelligence predicts disk re-herniation following lumbar microdiscectomy: Development of the "RAD" risk profile. *European Spine Journal, 30*(8), 2167-2175.

[19] Hu, L., Bell, D., Antani, S., Xue, Z., Yu, K., Horning, M. P., ... & Schiffman, M. (2019). An observational study of deep learning and automated evaluation of cervical images for cancer screening. *JNCI: Journal of the National Cancer Institute, 111*(9), 923-932.

[20] Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 11-19).

[21] Koss, L. G. (1989). The Papanicolaou test for cervical cancer detection: a triumph and a tragedy. *Jama, 261*(5), 737-743.

[22] Lakshmi, G. K., & Krishnaveni, K. (2016). Feature extraction and feature set selection for cervical cancer diagnosis. *Indian Journal of Science and Technology, 9*(19), 1-7.

[23] Lakshmi, K., & Krishnaveni, K. (2013). Automated extraction of cytoplasm and nuclei from cervical cytology images by fuzzy thresholding and active contours. *International Journal of Computer Applications, 73*(15).

[24] Li, L., Li, X., Yang, S., Ding, S., Jolfaei, A., & Zheng, X. (2020). Unsupervised-learning-based continuous depth and motion estimation with monocular endoscopy for virtual reality minimally invasive surgery. *IEEE Transactions on Industrial Informatics, 17*(6), 3920-3928.

[25] Lu, J., Zheng, X., Sheng, M., Jin, J., & Yu, S. (2020). Efficient human activity recognition using a single wearable sensor. *IEEE Internet of Things Journal, 7*(11), 11137-11146.

[26] Mackie, T. R., Kapatoes, J., Ruchala, K., Lu, W., Wu, C., Olivera, G., ... & Mehta, M. (2003). Image guidance for precise conformal

radiotherapy. *International Journal of Radiation Oncology, Biology, Physics, 56*(1), 89-105.

[27] Mahato, G. K., & Chakraborty, S. K. (2023). Securing edge computing using cryptographic schemes: a review. *Multimedia Tools and Applications*, 1-24.

[28] Malm, P., Balakrishnan, B. N., Sujathan, V. K., Kumar, R., & Bengtsson, E. (2013). Debris removal in Pap-smear images. *Computer Methods and Programs in Biomedicine, 111*(1), 128-138.

[29] Marinakis, Y., Dounias, G., & Jantzen, J. (2009). Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm-based feature selection and nearest neighbor classification. *Computers in Biology and Medicine, 39*(1), 69-78.

[30] Mat-Isa, N. A., Mashor, M. Y., & Othman, N. H. (2005). Seeded region growing features extraction algorithm; its potential use in improving screening for cervical cancer. *International Journal of The Computer, the Internet and Management, 13*(1), 61-70.

[31] Mayrand, M. H., Duarte-Franco, E., Rodrigues, I., Walter, S. D., Hanley, J., Ferenczy, A., ... & Franco, E. L. (2007). Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *New England Journal of Medicine, 357*(16), 1579-1588.

[32] Ojala, T., Pietikäinen, M., & Mäenpää, T. (2000). Gray scale and rotation invariant texture classification with local binary patterns. In T. Brodatzky, T. Feldman, T. Mäenpää, J. H. Roth, & L. H. Shapley (Eds.), *Computer Vision-ECCV 2000: 6th European Conference on Computer Vision Dublin, Ireland, June 26–July 1, 2000 Proceedings, Part I 6* (pp. 404-420). Springer Berlin Heidelberg.

[33] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(7), 971-987.

[34] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.

[35] Salembier, P., & Garrido, L. (2000). Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing, 9*(4), 561-576.

[36] Sankaranarayanan, R., Basu, P., Wesley, R. S., Mahe, C., Keita, N., Mbalawa, C. C. G., ... & IARC Multicentre Study Group on Cervical Cancer Early Detection. (2004). Accuracy of visual screening for cervical neoplasia: Results from an IARC multicentre study in India and Africa. *International Journal of Cancer, 110*(6), 907-913.

[37] Song, D., Kim, E., Huang, X., Patruno, J., Muñoz-Avila, H., Heflin, J., ... & Antani, S. (2014). Multimodal entity coreference for cervical dysplasia diagnosis. *IEEE Transactions on Medical Imaging, 34*(1), 229-245.

[38] Song, Y., Cheng, J. Z., Ni, D., Chen, S., Lei, B., & Wang, T. (2016, April). Segmenting overlapping cervical cell in pap smear images. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), IEEE,* 1159-1162.

[39] Song, Y., Tan, E. L., Jiang, X., Cheng, J. Z., Ni, D., Chen, S., ... & Wang, T. (2016). Accurate cervical cell segmentation from overlapping clumps in pap smear images. *IEEE Transactions on Medical Imaging, 36*(1), 288-300.

[40] Sornapudi, S., Stanley, R. J., Stoecker, W. V., Almubarak, H., Long, R., Antani, S., ... & Frazier, S. R. (2018). Deep learning nuclei detection in digitized histology images by superpixels. *Journal of Pathology Informatics, 9*(1), 5.

[41] Stapleford, L. J., Lawson, J. D., Perkins, C., Edelman, S., Davis, L., McDonald, M. W., ... & Fox, T. (2010). Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. *International Journal of Radiation Oncology, Biology, Physics, 77*(3), 959-966.

[42] Sukumar, P., & Gnanamurthy, R. K. (2016). Computer aided detection of cervical cancer using pap smear images based on adaptive neuro fuzzy inference system classifier. *Journal of Medical Imaging and Health Informatics, 6*(2), 312-319.

[43] Sun, G., Li, S., Cao, Y., & Lang, F. (2017). Cervical cancer diagnosis based on random forest. *International Journal of Performability Engineering, 13*(4), 446.

[44] Tan, S. Y., & Tatsumura, Y. (2015). George Papanicolaou (1883-1962): discoverer of the Pap smear. *Singapore Medical Journal, 56*(10), 586.

[45] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International Journal of Computer Vision, 104*, 154-171.

[46] Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*(11).

[47] William, W., Ware, A., Basaza-Ejiri, A. H., & Obungoloch, J. (2018). A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. *Computer Methods and Programs in Biomedicine, 164*, 15-22.

[48] Yang, J., Wei, C., Zhang, L., Zhang, Y., Blum, R. S., & Dong, L. (2012). A statistical modeling approach for evaluating auto-segmentation methods for image-guided radiotherapy. *Computerized Medical Imaging and Graphics, 36*(6), 492-500.

[49] Zhang, L., Lu, L., Nogues, I., Summers, R. M., Liu, S., & Yao, J. (2017). DeepPap: deep convolutional networks for cervical cell classification. *IEEE Journal of Biomedical and Health Informatics, 21*(6), 1633-1643.