IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Deploying Patch-based Segmentation Pipeline for Fibroblast Cell Images at Varying Magnifications

Hafizi Malik[1], Ahmad Syahrin Idris[2], Siti Fauziah Toha[1], Izyan Mohd Idris[3], Muhammad Fauzi Daud[4], and Mohammad Osman Tokhi[5]

[1]Healthcare Engineering and Rehabilitation Research, Department of Mechatronics Engineering, International Islamic University Malaysia, Gombak, Selangor, Malaysia
[2]Department of Electrical and Electronic Engineering, University of Southampton Malaysia, Iskandar Puteri, Johor, Malaysia
[3]Institute for Medical Research (IMR), National Institutes of Health (NIH), Ministry of Health Malaysia, Jalan Setia Murni U13/52, Seksyen U13 Setia Alam, 40170 Shah Alam, Selangor, Malaysia
[4]Institute of Medical Science Technology, Universiti Kuala Lumpur, Kajang, Selangor, Malaysia
[5]School of Engineering, London South Bank University, United Kingdom

Corresponding author: Ahmad Syahrin Idris (e-mail: A.S.Idris@soton.ac.uk).

This work is partially funded by the University of Southampton Malaysia SDG Fund UoSM/SDG2022/5. We would like to acknowledge our research collaborators at the Institute for Medical Research (IMR) Malaysia which provided materials and equipment for data collection and at the Institute of Medical Science Technology of Universiti Kuala Lumpur Malaysia for providing valuable information, datasets, and guidance needed to complete this study.

ABSTRACT Cell culture monitoring necessitates thorough attention for the continuous characterization of cultivated cells. Machine learning has recently emerged to engage in a process, such as a microscopy segmentation task; however, the trained data may not be comprehensive for other datasets. Most algorithms do not encompass a wide range of data attributes and require distinct system workflows. Thus, the main objective of the research is to propose a segmentation pipeline specifically for fibroblast cell images on phase contrast microscopy at different magnifications and to achieve reliable predictions during deployment. The research employs patch-based segmentation for predictions, with U-Net as the baseline architecture. The proposed segmentation pipeline demonstrated significant performance for the UNet-based network, achieving an IoU score above 0.7 for multiple magnifications, and provided predictions for cell confluency value with less than 3% error. The study also found that the proposed model could segment the fibroblast cells in under 10 seconds with the help of OpenVINO and Intel Compute Stick 2 on Raspberry Pi, with its optimal precision limited to approximately 80% cell confluency which is sufficient for real-world deployment as the cell culture is typically ready for passaging at the threshold.

INDEX TERMS Cell confluency, deep learning, fibroblast, microscopy segmentation, phase contrast

I. INTRODUCTION

Skin is the largest organ in the human body and also serves as a barrier against outside pollutants [1]. It is mostly composed of the protein collagen, which is produced by fibroblast cells. When the skin is injured, the wound-healing process stimulates the production of new fibroblasts, which create collagen to close the wound. Therefore, fibroblast cell culture is critical for growing wound grafts and helping tissue regeneration in surgery.

Cells are commonly divided into three categories which are lymphoblast-like, epithelial-like, and fibroblastic. As previously stated, a skin cell is in the form of fibroblastic. When compared to other categories, fibroblastic has the most difficult shape representation, especially when imaging analysis is required. It is because fibroblasts have irregular shapes and dimensions over their growth period, whereas others have significant shapes and regular dimensions. This makes it more challenging to accurately segment fibroblasts, as their appearance may vary significantly over time. Additionally, since cell observation is done in a closed environment, proper lighting adjustments are needed for optimal utilization.

Cell segmentation is the process of segmenting microscopic image pixels into the binary class where true and false classes represent cell regions and unoccupied regions, respectively. A well-segmented image can capture

biologically relevant morphological information such as health and contamination; however, fuzzy boundaries can cause segmentation leaks in which a segmented region expands into neighboring structures [2]. However, there are still some occupations that use traditional techniques to segment the cells, which results in inconsistency, especially when a wide range of expertise, from biology specialists to interns, is involved. It is because human eyes are prone to optical illusions [3], and ineffective and unstable at measuring something quantitatively. U-Net, a deep learning-based segmentation algorithm, has been shown to perform well on biomedical images, such as cells, and after years, researchers identified several limitations of U-Net architecture, which were then improved with various enhancement mechanisms and has shown significant performance on their own case study [4].

Recent research on skin segmentation has largely focused on the outer areas of the skin, such as skin pores [5], chronic wounds [6], and lesions or cancer [7], [8], or the outer layer of the skin, such as epidermal tissue [9] rather than the cellular structure of the skin. Furthermore, most research on cellular segmentation has focused on cells that have less significant shape changes over time, such as blood cells [10], [11], cervical cells [12], [13], and nucleus [14], and has been conducted at single magnification level. These cells tend to have circular shapes and regular dimensions. This has limited the scope of study to the visible features of the skin rather than the underlying cells and cells with more consistent and predictable shapes, rather than cells that exhibit more complex and varied shape changes over time. The previous studies differ in that they introduce new models or approaches to address similar object segmentation on various image modalities.

The study utilizes phase-contrast microscopy images of fibroblast skin cells as the test subject and only requires binary segmentation for its confluency, which refers to the area that the cells cover within an image. Thus, the study does not aim to classify the types of cells or characterize the fibroblast cell for any health conditions or contamination. Following the scope of this study, the objectives are to identify a proper implementation of a particular deep learning enhancement mechanism that can innovate a better model based on this case study and find a correlation between the segmentation complexity and the phase contrast microscope's image quality with respect to magnification levels.

This paper presents a preliminary work of segmenting different magnification levels of fibroblast cells and providing precise and consistent confluency value predictions, with a focus on achieving high performance with low complexity before deploying the model to an edge device like a Raspberry Pi. Despite the development of various complex models by researchers, this study focuses only on a single, relevant deep-learning enhancement mechanism to improve the performance of a pre-existing model rather than developing an entirely new model. We show that our approach has both advantages and disadvantages in relation to the characteristics of the datasets.

The main contribution of this paper is to demonstrate the improved performance and minimal complexity of the modified model on the new application of segmenting the features of fibroblast cell images at different magnification and confluency levels and to provide guidance on the necessary precautions to be taken in the segmentation pipeline for future research in this area. The workflow of this study is shown in Fig. 1.

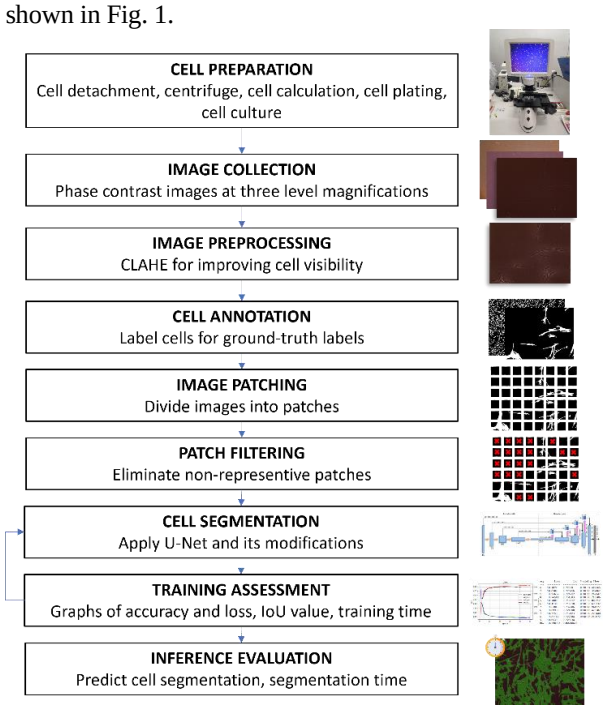


FIGURE 1. A workflow diagram of the proposed study.

II. RELATED WORK

According to the literature review, the majority of biomedical and medical research on multi-scale segmentation has primarily focused on enhancing the model architecture rather than exploring its applications. Moreover, while researchers claim to have developed multi-scale architectures, they were mainly tested on single-scale object images based on the record. Below are the kinds of literature that are directly involved with multi-scale object images.

[15] rethought U-Net architecture for multi-scale biomedical image segmentation, introducing MultiResUNet. Multiple sizes of lesion images are one of the datasets tested. The researchers increased the number of 3×3 filters of each convolutional block in the successive three layers. The feature maps of each successive filter are concatenated and added with the feature maps of a residual connection from the input along with 1×1 filters. They also replaced the original skip connection of U-Net with a sequence of four successive 3×3 filters, each accompanied by a short residual connection consisting of 1×1 filters. However, the researchers opted for a lower number of filters in the added convolutional layers compared to the original U-Net. As a

result, the MultiResUNet model featured only 7,262,750 parameters.

[16] proposed a U-Net architecture with a modified receptive field with a feature-fused module and attention gate mechanism called as FF-UNet. The feature-fused module replaced traditional U-Net convolutions with two branches of convolution operation of 3x3 kernel filter and one with dilation size 2, where both operations are fused using element-wise addition. The model resulted in 3.94 million parameters which are 49% less than the classical U-Net. The model was experimented on datasets of skin lesions, colorectal polyps, and nuclei images. The study found that the FF-UNet model without an attention gate performs better on most of the open datasets.

In contrast to the previously mentioned studies, the literature discussed further utilized multi-scale images which are merely based on different image resolutions. While it is worth discussing these findings, it should be noted that the multi-scale images in the study exhibited a lack of feature variations in their internal structure. However, the conceptual framework employed in this research holds significant potential, even if the implementation falls short in terms of capturing diverse features at various scales. [17] proposed PMED-Net that consists of six small encoder-decoder networks (six levels) where each network generates coarse predictions and they are upsampled with stride 2, concatenated with different scale input images and used as input for next level network. Each instance of the proposed network has only three stages with a much fewer number of feature maps as 16, 32, and 64. Overall, the proposed architecture comprises 1,465,974 parameters for its six-pyramid level training.

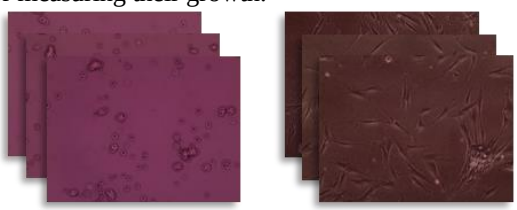
[18][19] introduced a scale-aware transformer that allows the system to learn local and global representations from digital whole slide histopathological images of skin tissue images at multiple scales in an end-to-end fashion. Their proposed model has three main steps, including learning patch embeddings using MobileNetv2 for varying input scales, learning contextualized patch embeddings for each input scale using transformers, and learning scale-aware embeddings of concatenated multiple input scales using transformers. In essence, the model initially requires separate training paths for specific input scales and subsequently combines the outputs of each path to generate a unified scale-aware model.

Through the literature review, two concerns have emerged. Firstly, the degree difference of feature variations of diverse biomedical objects. Secondly, the feasibility of the developed model for a modular deployment. Thus, it is crucial for another research to delve into the feasibility of the multi-scale model on different complex objects and devices.

III. RESEARCH METHODOLOGY

A. DATA ACQUISITION AND ANNOTATION

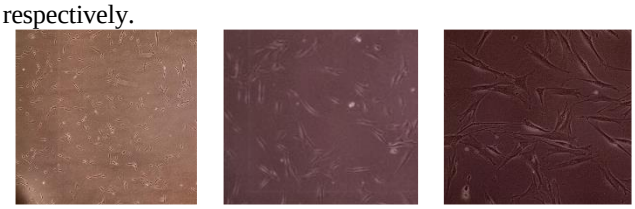
This study utilized two datasets on fibroblast cell growth, comprising a total of 400 images with a dimension of 2048 x 1536 pixels. These images were obtained using phase-contrast microscopy from the Institute for Medical Research (IMR) Malaysia and the Institute of Medical Science Technology of Universiti Kuala Lumpur. In the experiments, the fibroblast cells were allowed to grow and reach maximum confluency for up to 7 days, based on the initial number of cells placed in the petri dish. The datasets show significant differences in terms of their features, particularly the shape of the cells, as they grow. Specifically, cells in dataset A have spherical shapes while cells in dataset B have elongated shapes, as shown in Fig. 2. Additionally, dataset A only includes images from the initial culture day, while dataset B includes images from the following days. This research will focus on dataset B, which consists of 300 images and is the most accurate representation of cell growth from day 1 forward. Meanwhile, dataset A is only suitable for counting the number of cells, not for measuring their growth.



(a) Dataset A: Day 0 (b) Dataset B: Day 1-7

FIGURE 2. A collection of fibroblast cell datasets based on their shape differences.

EVOS™ XL Core Imaging System was used to capture cell images. It is commonly utilized for cell or tissue culture using phase contrast method. Based on the zoom capability of the microscope, Dataset B is divided into three subsets based on the image magnifications: x4, x10, and x20, with approximately 100 images in each subset. Their distinct features can be seen in Fig. 3. Brightness, contrast, and saturation are three manipulated variables that the microscope can adjust. To define the experiment's standard procedure, their values are set to be constant for all magnification levels. Therefore, both contrast and saturation are set to 50, which is the default value of the microscope and brightness is set between 35 to 60 depending on the magnification levels. As a result, images appear darker at higher magnification levels. The implementation setup is chosen because our subsequent study aims to deploy the model to an edge device that can be easily used by individuals with little experience. Next, images are chosen from dataset B and split for training-validation set and test set to 80% and 20%, respectively. The training-validation set is then randomly divided into 70% and 30%, respectively.



(a) x4 (b) x10 (c) x20

FIGURE 3. Dataset B with three different magnification levels.

B. IMPLEMENTATION STRATEGIES ON SPECIFIC DATASET CONDITIONS

Model performance and complexity can be influenced by the conditions of the dataset. In this study, we consider deep learning approaches to address these challenges. Our dataset consists of images where the background and cell color are quite similar, making it difficult for traditional computer vision and unsupervised machine-learning methods to accurately segment the cells. While deep learning methods are better at segmenting a wide range of objects, they can still be challenged by low imaging quality, high-resolution images, and a lack of images. To address these issues, we believe that the problems can be solved before training by applying color correction and image patching techniques as needed.

1) COLOR CORRECTION FOR LOW IMAGING QUALITY

Common issues with biomedical images are that they have low contrast and high noise [20]. Contrast Limited Adaptive Histogram Equalization (CLAHE) is chosen to correct the image color. Compared to standard histogram equalization, it often requires experiment-specific parameter tuning. CLAHE is applied per tile to improve the visibility level of objects in foggy images by stretching the histogram, limiting the contrast, and performing bilinear interpolation at the edges to match the next tiles. First, the images are converted to LAB space, where the L channel contains luminosity or intensity information, and the A and B channels contain color information. The channels will be separated, and the filter will be applied only to the L channel, with the channels being merged back in for the original image.

2) IMAGE PATCHING FOR HIGH-RESOLUTION IMAGES

Each image has a resolution size of 2048x1536 pixels based on data collection. For training, the images are deemed too large to be handled by the computational power available at the time. To avoid heavy loss of model training, the images will be patched into a number of smaller input sizes, each of which can still represent the object features; thus, the study proposed that the images are patched into 256x256 pixel images in which each original image yields 48 image patches. Based on the experiment, it was found that using patch sizes below 256 pixels could negatively impact the representative features of the patches. This method reduces the computational power needed for training and indirectly expands the number of images available for the training and validation sets [21].

To avoid an imbalanced dataset, we filtered several non-representative patches that had either too many background pixels or too many cell pixels, and an average of 688 patches of each magnification level were available for the training-validation process in total. Ideally, each patch should have an equivalent number of cell and background pixels so that the model can learn significant feature differences. This step is critical because, as previously stated, data for the training-validation process is divided at random and the training set may contain fewer complex images than the validation set for particular epochs, affecting training robustness and performance.

C. BASELINE AND NETWORK ARCHITECTURE

U-Net CNN is used as the baseline for comparison because it previously demonstrated satisfactory performance for biomedical images with low training images, which is relevant to our study. As mentioned earlier, some images from the subset of 100 images were excluded from the training data because they lacked sufficient feature information. It also has simpler architecture, which makes it relatively less complex to deploy [22] and allows for more refinement possibilities to ensure compatibility with data conditions, compared to other existing cell segmentation models, based on the literature review. Therefore, any modification of a model will be done on the standard U-Net. However, a reduced version of U-Net is preferred because it has fewer parameters with 1,940,817 weights and is proven to perform more or less than standard U-Net with 26,590,833 trainable parameters. Thus, this becomes the first step to reducing the model complexity of trainable parameters while maintaining its performance. In this experiment, U-Net starts with 16 filters of convolution layer and the filters are doubled for each layer at the encoding stage and are halved at the decoding stage.

D. COMPUTER HARDWARE AND TRAINING SETTINGS

The Anaconda Python distribution is recommended to accommodate the program and libraries on Windows 10 due to its ease of package movement and deployment. Each model is implemented and trained on an NVIDIA GeForce RTX 3060 GPU with 12 GB VRAM using the TensorFlow GPU v2.8.0/ Keras framework. Adam optimizer is used, with a learning rate of 1e-3. A hybrid loss of focal loss and dice loss is used as the metrics to control and evaluate the model performance during training and overcome the effects of imbalanced datasets where the cells only occupy a small region in most images [23]. All models are trained for a maximum of 50 epochs with a batch size of 24 training samples, with checkpoints used after each epoch to save only the best model with the best validation loss as the stopping criteria.

For deployment hardware, Raspberry Pi with ARMv7 Processor rev 3 CPU is utilized alongside Intel Neural Compute Stick (NCS2) as a vision processing unit.

E. EVALUATION CRITERIA

In this study, Intersection over Union (IoU), also known as the Jaccard index, is currently used as the main evaluation indicator to select the best result. F-score components, precision and recall, are also considered as the alternative metric to identify segmentation errors, such as under-segmentation and over-segmentation, on test data predictions. The selected metrics are derived from four elements of a confusion matrix which are true positive, false positive, true

positive, and false negative. In this study context, the terms are described in Table I.

TABLE I
CONFUSION MATRIX ELEMENTS

Terms			Descriptions	Remarks
True Positives	(TP)	-	True predictions on cell pixels	Merit
False Positives	(FP)	-	Not predicted cell pixels	Demerit
False Negatives	(FN)	-	Wrong predictions on background pixels	Demerit
True Negatives	(TN)	-	True predictions on background pixels	Merit

Thus, the considerable metrics are computed based on the confusion matrix elements as (1-4). IoU focuses on indicating the quality of localization of cell predictions. While F-score provides the harmonic mean of precision and recall, both focus on over-segmentation and under-segmentation errors of cell predictions, respectively.

$$\text{IoU} = \frac{\text{TP}}{\text{TP+FP+FN}} = \frac{\text{labels} \cap \text{predictions}}{\text{labels} \cup \text{predictions}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{Fscore} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision+Recall}} \quad (4)$$

IV. RESULTS AND DISCUSSION

A. MODEL DEVELOPMENT PLAN

The model development is focused on x4 and x20 magnifications because they have significant differences in terms of feature and image settings for segmentation. This notion also expedites the laborious annotation process. Referring to the available dataset, this study aims to develop a segmentation process that can adapt all magnifications of fibroblast cell growth using a relevant deep learning enhancement mechanism on U-Net architecture.

Thus, the implementation of multi-scale convolution blocks on U-Net is chosen to extract features at multiple scales or resolutions from microscopy images. This helps the model capture both low-level and high-level information for the segmentation task. The multi-scale mechanism has been long introduced by the Inception network and several works have applied the mechanism on U-Net [24], [25]. Based on the review, the architecture is developed by replacing the U-Net convolutional blocks with different concurrent kernel sizes. Then, the feature map of convolutional paths is concatenated to be an output for the following layers. However, the selection of concurrent kernel sizes seems to be varying in different experiments. In this case study, we proposed a different strategy from previous work to identify the best combinations for a multi-scale model, and trials are done to investigate their potential to what extent.

Rather than randomly combining multi-scale blocks, Table II shows that the trials are carried out by replacing the original kernel size 3x3 of U-Net convolutional layers with 5x5, 7x7, and 9x9, and their performance is recorded separately. The best of the allocated kernel sizes will be selected and paired in the final model. For training, three configurations of datasets are prepared: x4 only, x20 only, and a combination of both datasets. Each configuration was subjected to ten trials. Based on the table, models with kernel sizes 3x3 and 5x5 provide significant performance with higher IoU values and lower deviation for both datasets.

TABLE II
IOU COMPARISONS OF DIFFERENT KERNEL SIZES ON U-NET

Kernel sizes	Intersection over Union (IoU)		
	Training datasets based on magnification levels		
	x4	x10	x20
3x3	0.707±0.002	0.661±0.024	0.626±0.017
5x5	0.709±0.002	0.670±0.036	0.678±0.030
7x7	0.704±0.009	0.652±0.032	0.596±0.074
9x9	0.695±0.012	0.592±0.048	0.613±0.054

A dropout layer is placed between the first and second convolution layers of the convolution path to improve the generalization of the model. The output of both kernel sizes is then concatenated to be an input for the subsequent layer. However, the concatenation dramatically increases the depth dimension of the input to the subsequent layers as well as results in considerably more parameters and higher complexity. In order to reduce the depth, each concatenated feature map will be followed by a 1x1 convolution layer. The multi-scale model with and without the 1x1 convolution layer produces 7,063,937 and 9,238,705 trainable parameters, respectively. In terms of training time, the former takes approximately 1 minute less than the latter to train for 50 epochs. As a result, this initiative prevents larger space use and lesser time complexity. The illustration of the multi-scale block and the final model design are shown in Fig. 4 and Fig. 5.

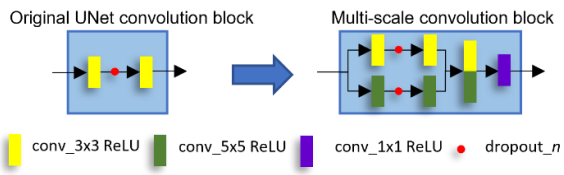


FIGURE 4. The modification of the baseline U-Net convolution block into a multi-scale version.

B. SEGMENTATION PIPELINE PERFORMANCE

In Fig. 6, the mean training graphs of the train and validation IoU score of (A) the standard U-Net and (B) its multi-scale are displayed, and the shaded region represents the deviation of five trials. The bar graph below the training graphs shows the

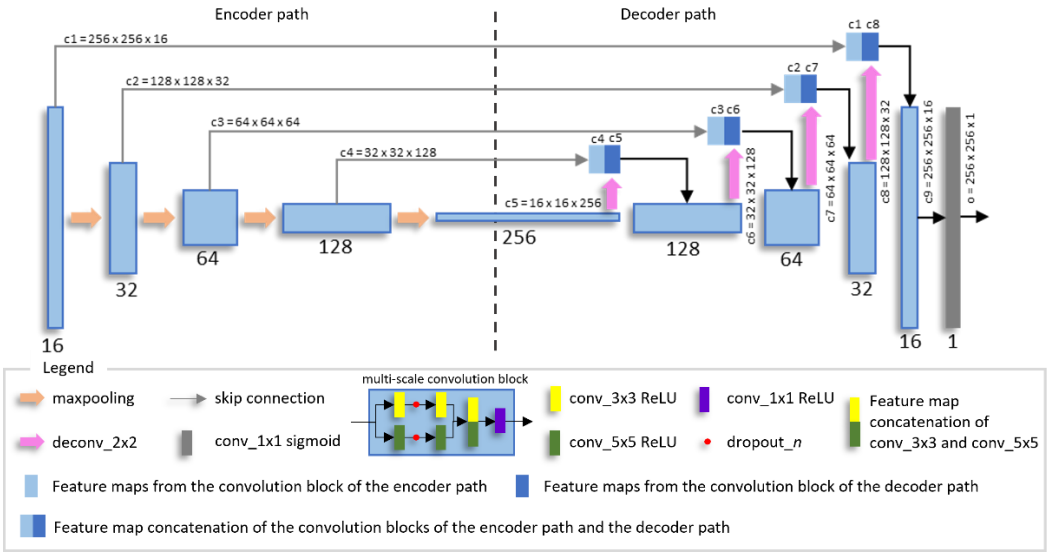


FIGURE 5. **The proposed model for different magnification levels of fibroblast cell segmentation.**

overfitting scenarios of the models across the epochs as the score differences are calculated by subtracting mean train IoU with mean validation IoU where the increasing positive and negative values indicate overfitting and underfitting, respectively. It can be observed that the multi-scale model is able to significantly reduce the degree of overfitting of the baseline U-Net, especially on x10 and x20 datasets. It can also be concluded that the higher the magnification levels, the harder the model is to learn and converge.

As mentioned earlier, although there are existing models, such as Mask RCNN and DeepLabV3+, that have demonstrated superior performance in cell segmentation tasks like U-Net, their resource-intensive nature makes them unsuitable for training using the specified settings and for deployment on edge devices, making them irrelevant to be in comparisons, as they are not easily optimizable for training or deployment. Thus, we have evaluated the performance of two common lightweight architectures for mobile implementations, MobileNet and EfficientNet [26], in comparison to the proposed multi-scale U-Net model for our cell segmentation. Both models were primarily designed for image classification tasks. Therefore, a decoder architecture, comprising of a sequence of upsampling or transposed convolution layers, is added to perform the segmentation tasks where the modifications are highlighted in Table III, including the complex aspects of the models. It is worth noting that the aspects presented are merely the minimum requirements necessary or initial precautions for running the models. In reality, training deep learning models involves several overheads, such as the amount of data and data complexity, that can significantly impact the memory requirements and overall cost. For deployment, the inferencing time is rather concerning compared to the training time. Regardless of the model performance, while both, the multi-scale U-Net and EfficientNet, have competitive results in terms of parameter and feature memory, the multi-scale U-Net has an advantage when it comes to inferencing time. Despite the fact that the multi-scale U-Net requires more training time than EfficientNet, its lower inferencing time makes it better suited for deployment purposes. On the other hand, while EfficientNet may have lower training time, it requires more inferencing time, which could be a hindrance.

TABLE III
MODEL COMPLEXITY COMPARISONS

Model	U-Net	Multi-scale	MobileNet	EfficientNet
Number of layers	40	85	99	250
Trainable parameters	1,940,817	7,063,937	6,002,817	7,327,677
Memory for features (MB)	1695	3891	2739	3689
Minimum memory required (GB)	1.66	3.83	3.79	4.87
Model size (MB)	23	83	71	87
Mean training time (mins)	2.9	7.4	3.3	5.2
GPU inference time (ms/ patch)	40	56	56	72

Their performance results are presented in Table IV, which clearly demonstrates that the proposed multi-scale U-Net model outperformed other models in terms of IoU score. It is observed that the performance improvement of the multi-scale model was particularly prominent on x10 and x20 cell images. This is likely due to the fact that the cell images significantly exhibit varying structural dimensions across their morphologies, which the refinement approach, multi-scale convolution path, was able to effectively leverage. On the

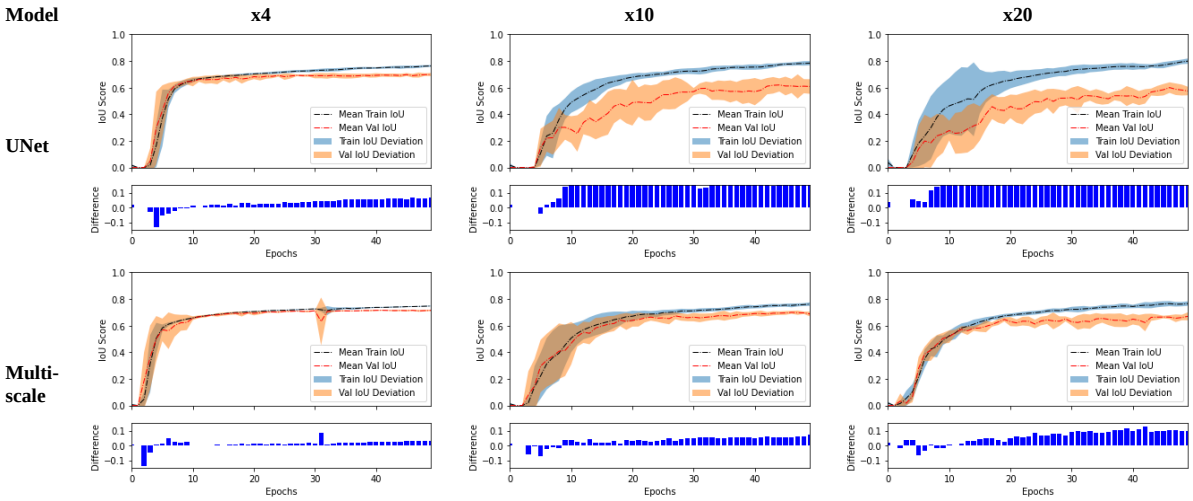


FIGURE 6. **Graphs of the proposed model training performance on different datasets.**

other hand, cells in x4 mostly possess consistent dimensions of individual cells, which may attribute why the performance improvement was less pronounced for this image set. Overall, the results clearly highlight the effectiveness of our proposed multi-scale U-Net model in accurately segmenting cell images, particularly with inconsistent structural dimensions.

TABLE IV
IOU COMPARISONS OF DIFFERENT MODELS

Model	Intersection over Union (IoU)		
	Training datasets based on magnification levels		
	x4	x10	x20
U-Net	0.707±0.002	0.661±0.024	0.626±0.017
Multi-scale	0.717±0.002	0.703±0.009	0.701±0.009
MobileNet	0.474±0.010	0.554±0.010	0.603±0.012
EfficientNet	0.421±0.035	0.487±0.031	0.589±0.025
Relative performance (%)			
Multi-scale	**+ 1.41**	**+ 6.35**	**+ 11.98**
MobileNet	-32.96	-16.19	-3.67
EfficientNet	-40.45	-26.32	-5.91

Other metrics that are also taken into account are precision and recall, indicating the degree of under-segmentation and over-segmentation errors of the models. Fig. 7 shows the visual representation of patch segmentation on the test dataset including their precision and recall. The observations made from the figure reveal that the models have distinct combinations of IoU, precision, and recall scores, which implies that each model has its own strengths and weaknesses in segmenting the objects of interest. The observations made from the figure revealed that the models have distinct combinations of IoU, precision, and recall scores, which implies that each model has its own strengths and weaknesses in segmenting the objects of interest. Based on the visual, red and blue pixels represent under-segmentation and over-segmentation errors, respectively. Overall, the multi-scale model outperforms all other models, with an average precision of 78.71 and an average recall of 85.16 across all datasets. Based on the two metric scores, the multi-scale model is expected to have more over-segmentation errors than under-segmentation errors when segmenting the cells.

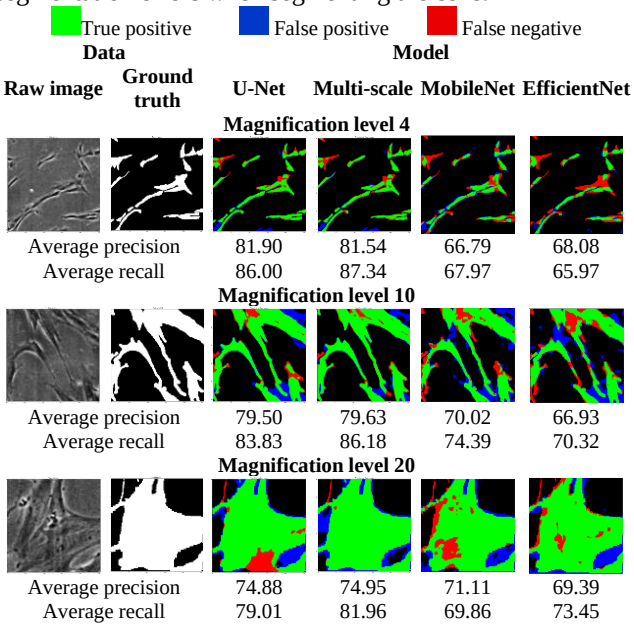


FIGURE 7. **Precision and recall of patch segmentation for the models.**

C. SEGMENTATION FOR DEPLOYMENT ANALYSIS

The original images are 2048x1536 pixels in size, and the model is trained using the selected patch size during training. The images for segmentation should have similar dimensions as the images used during training. Thus, for predictions, it is also appropriate to employ patch-based segmentation instead of resizing the images for model compatibility, which the approach can jeopardize the discriminative features of the cells. The original images will be patched first, and cells will

be segmented by patches. After each segmented patch is completed, the patches are unpatched to obtain the original images with true class labels as shown in Fig. 8. Essentially, the predictions will return the probability of each pixel. In this experiment, pixels with values of 0.5 and above are considered a true class and vice versa.

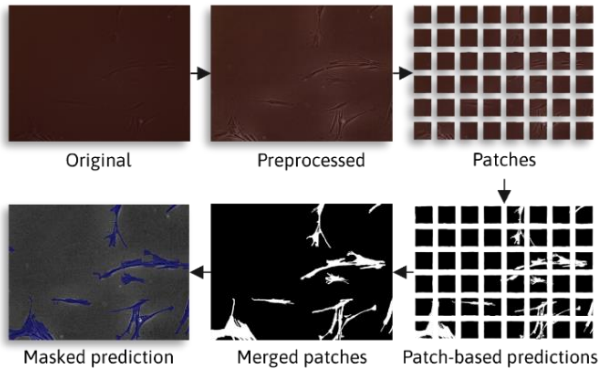


FIGURE 8. Patch-based segmentation procedures.

However, when the patches are directly stitched, the segmented images have edge effects that cause discontinuity between the next tiles when they are combined, especially when the tiles have high cell confluency or fuzzy regions as illustrated in Fig. 9. The effects are clearly visible when the segmentation results are plotted in binary. To address the issue, overlapping patches were used during the prediction phase, and then the blending method was applied to merge the overlapping regions into an intensity mask using element-wise addition. The image is patched into segments of 256 pixels in dimension, with a stride of 128 pixels between each segment. As a result, the overlapped segmentation is more visually pleasing compared to the direct stitching. Certainly, as the overlapping patches were employed, the number of patches per image to be segmented also increased from 48 patches to 165 patches, making the segmentation time approximately 3 times longer than the average segmentation of 4 seconds per image, as shown in Table 5, to 12 seconds.

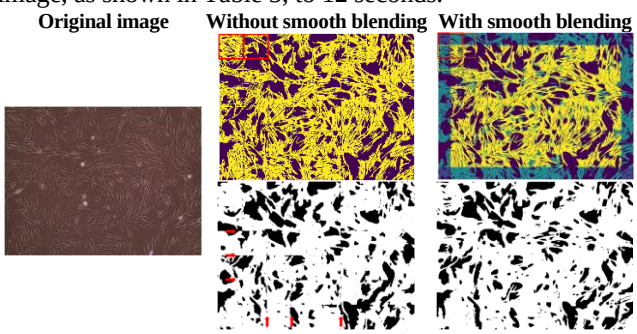


FIGURE 9. Comparisons between segmentation with and without smooth blending.

Fig. 10 displays examples of confluency segmentation at different magnification levels, utilizing the smooth blending segmentation approach with the multi-scale model. As previously hypothesized, the predicted confluency values tend to be higher than the ground truth values due to the model's lower precision than recall. Nevertheless, the average relative difference between the predicted and ground truth confluency values is only ±2.94%, which can be considered negligible, when deployed.

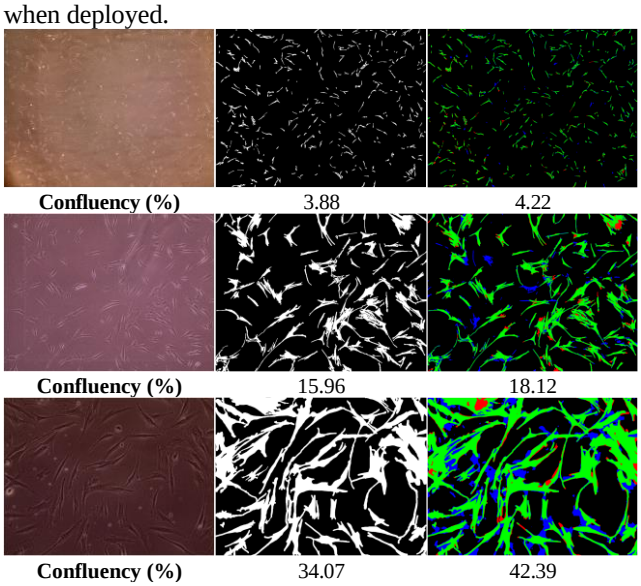


FIGURE 10. Confluency segmentation of large images at different magnification levels using the proposed models.

Table V compares the average segmentation or inference time of the baseline U-Net and the multi-scale mechanism on both a personal computer and a Raspberry Pi with a specific processing unit. One of the significant findings is that the OpenVINO framework is more efficient in reducing the inference time compared to the Keras framework on both platforms. However, using the Neural Compute Stick (NCS2) only results in improved inference time on the Raspberry Pi as an edge device, not on the personal computer, which already has higher computational capabilities. Briefly, the multi-scale model has a segmentation time that is approximately twice as long as the baseline U-Net, but it also has a significant relative performance, especially on x20 cell images which is above 10%, as depicted in Table IV.

D. DISCUSSION AND FUTURE WORKS

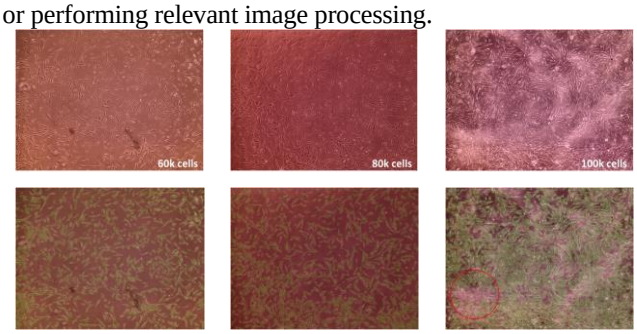
As stated earlier, this research serves as a preliminary work aimed at integrating the segmentation model into an automated cell growth monitoring system, which could potentially involve various components such as image analysis tools, remote monitoring, and sample handling. Therefore, it is imperative to consider the practical implementation of the segmentation model in real-world scenarios.

One important consideration is the adaptability of the segmentation model to variations over time. In addition to the lighting settings previously mentioned, sharpness is another variable that cannot be digitally adjusted. One finding from this study is that the higher the confluency in a petri dish, the harder it is to sharpen the microscope image. It is believed

TABLE V
SEGMENTATION TIME COMPARISONS OF U-NET AND ITS MULTI-SCALE

Platform	Framework	Processing Unit	Model	Time Taken (secs)
PC	Keras	CPU (Intel i5-11400F)	U-Net	3.07±0.07
			Multi-scale	5.07±0.21
		GPU (Nvidia RTX 3060)	U-Net	2.21±0.34
			Multi-scale	2.59±0.23
	OpenVINO	CPU (Intel i5-11400F)	U-Net	0.62±0.08
			Multi-scale	1.85±0.06
		VPU (Intel Movidius Myriad X)	U-Net	3.70 ±0.02
			Multi-scale	7.84±0.02
Raspberry Pi	Keras	CPU (ARMv7 rev 3)	U-Net	28.58±0.37
			Multi-scale	88.47±2.03
	OpenVINO	VPU (Intel Movidius Myriad X)	U-Net	4.84 ±0.21
			Multi-scale	8.91±0.12

because the more cells there are in the specific area, the more complex and cluttered the image becomes, making it more difficult to distinguish the features of the individual cells as the light passing through the sample may be scattered or diffracted, resulting in a blurred or hazy image, as depicted in Fig. 11. The scenario mostly occurs when the confluency nears 80%, and the cells are not evenly distributed across the petri dish. Therefore, as the confluency increases, it is necessary to adjust the sharpness of the microscope. This issue must be considered if fully automated cell growth monitoring is to be developed as the confluency increases over time. Thus, in future, it may be beneficial to collect and train on images with cluttered or crowded cells. Another approach could be to develop a feedback system based on reinforcement learning to detect changes in image quality over time so that corrective actions can be executed such as adjusting the sharpness level or performing relevant image processing.



Confluency: 17.86% **Confluency:** 22.03% **Confluency:** 55.58%

FIGURE 11. Large image segmentation of different confluency levels.

The second consideration is a streamlined workflow for the end users. At present, the segmentation model is trained specifically on individual datasets. As a result, three models were created in this study, each representing a different magnification level but with the same CNN architecture. To streamline the workflow, instead of users having to manually select the appropriate model, one potential approach is to develop a cell image classification system that automatically identifies the corresponding cell image and selects the appropriate model. Another alternative is to develop a new adaptable model for the merged dataset. Transfer learning was utilized to test the multi-scale model on the merged dataset, aiming to ensure all datasets were trained equally in each batch. Unfortunately, the outcomes showed varying model behavior, particularly between x4 and x20 images, where both contain cells with distinct features in terms of dimension and inner structure, with x20 images having larger dimensions and more intricate inner structures than x4 images.

V. CONCLUSION

A multi-scale mechanism was applied on the baseline U-Net to adapt various magnification levels and inconsistent object structural dimensions. The results demonstrated improved performance in terms of IoU, precision and recall, especially on the x20 cell image set. The models were compared to common models for mobile deployment, MobileNet and EfficientNet, where both have been outperformed despite the reduced version of U-Net having simpler architecture. Generally, biomedical images have high resolution, thus, patch-based segmentation was employed and overlapping patches were implemented to reduce the edge effects of neighboring patches. The segmentation pipeline successfully met the needs of end-users by providing reliable predictions of confluency values at varying magnification levels with less than 3% error compared to the ground truth values.

While the segmentation pipeline was successful, the performance of the pipeline was limited at higher confluency levels, which may affect the segmentation of crowded cells towards the end of the lag phase of cell growth. However, this limitation seems trivial during the deployment of the automated cell growth monitoring system since the cell culture is usually passaged when it reaches 80% confluency where the scenario commonly appears.

Based on this study, it can be concluded that future work should focus on higher magnification and confluency levels for more critical analysis. In terms of deployment hardware, the use of the OpenVINO framework and Intel Movidius Neural Compute Stick (NCS2) definitely improved the inferencing time of Raspberry Pi as an edge device. Therefore, further research on its capabilities and limitations upon various deep learning mechanisms will be beneficial.

ACKNOWLEDGMENT

This work is partially funded by The University of Southampton Malaysia SDG Fund UoSM/SDG2022/5. We would like to acknowledge our research collaborators at the Institute for Medical Research (IMR) Malaysia which provided materials and equipment for data collection and at the Institute of Medical Science Technology of Universiti

Kuala Lumpur Malaysia for providing valuable information, datasets, and guidance needed to complete this study.

REFERENCES

[1] A. Kaur, S. Midha, S. Giri, and S. Mohanty, "Functional skin grafts: Where biomaterials meet stem cells," *Stem Cells Int.*, pp. 1–20, 2019.

[2] J. Cao, "Accurate Cell Segmentation Based on Biological Morphology Features," *2018 IEEE Int. Conf. Syst. Man, Cybern.*, pp. 3380–3383, 2018.

[3] A. Korzynska, L. Roszkowiak, J. Zak, and K. Siemion, "A Review of Current Systems for Annotation of Cell and Tissue Images in Digital Pathology," *Biocybern. Biomed. Eng.*, vol. 41, no. 4, pp. 1436–1453, 2021.

[4] M. Z. Khan, G. S. Member, M. A. Khan, and S. Member, "Deep Neural Architectures for Medical Image Semantic Segmentation : Review," *IEEE Access*, vol. 9, pp. 83002–83024, 2021.

[5] K. Kachi, Y. Iwahori, H. Usami, S. Fukui, A. Wang, and M. K. Bhuyan, "Pore Detection From Human Skin Image Using U-Net," *2020 9th Int. Congr. Adv. Appl. Informatics*, pp. 463–466, 2020.

[6] P. Gholami, M. A. Ahmadi-Pajouh, N. Abolftahi, G. Hamarneh, and M. Kayvanrad, "Segmentation and Measurement of Chronic Wounds for Bioprinting," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 4, pp. 1269–1277, 2018.

[7] G. G. De Angelo, A. G. C. Pacheco, and R. A. Krohling, "Skin lesion segmentation using deep learning for images acquired from smartphones," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2019-July, no. July, pp. 1–8, 2019.

[8] M. A. Anjum, J. Amin, M. Sharif, H. U. Khan, M. S. A. Malik, and S. Kadry, "Deep Semantic Segmentation and Multi-Class Skin Lesion Classification Based on Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 129668–129678, 2020.

[9] K. R. J. Oskal, M. Risdal, E. A. M. Janssen, E. S. Undersrud, and T. O. Gulsrud, "A U-Net Based Approach to Epidermal Tissue Segmentation in Whole Slide Histopathological Images," *SN Appl. Sci.*, vol. 1, no. 7, pp. 1–12, 2019.

[10] M. Zhang, S. Member, X. Li, and M. Xu, "Automated Semantic Segmentation of Red Blood Cells for Sickle Cell Disease," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 11, pp. 3095–3102, 2020.

[11] K. Al-Dulaimi, J. Banks, K. Nugyen, A. Al-Sabaawi, I. Tomeo-Reyes, and V. Chandran, "Segmentation of White Blood Cell, Nucleus and Cytoplasm in Digital Haematology Microscope Images: A Review-Challenges, Current and Future Potential Techniques," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 290–306, 2021.

[12] J. Huang, G. Yang, B. Li, Y. He, and Y. Liang, "Segmentation of Cervical Cell Images Based on Generative Adversarial Networks," *IEEE Access*, vol. 9, pp. 115415–115428, 2021.

[13] D. Jia, Z. Li, and C. Zhang, "A Parametric Optimization Oriented, AFSA Based Random Forest Algorithm: Application to the Detection of Cervical Epithelial Cells," *IEEE Access*, vol. 8, pp. 64891–64905, 2020.

[14] N. Kumar, R. Verma, D. Anand, Y. Zhou, O. F. Onder, E. Tsougenis, H. Chen, P. A. Heng, J. Li, Z. Hu, Y. Wang, N. A. Koohbanani, M. Jahanifar, N. Z. Tajeddin, A. Gooya, N. Rajpoot, X. Ren, S. Zhou, Q. Wang, D. Shen, C. K. Yang, C. H. Weng, W. H. Yu, C. Y. Yeh, S. Yang, S. Xu, P. H. Yeung, P. Sun, A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, O. Smedby, C. Wang, B. Chidester, T. V. Ton, M. T. Tran, J. Ma, M. N. Do, S. Graham, Q. D. Vu, J. T. Kwak, A. Gunda, R. Chunduri, C. Hu, X. Zhou, D. Lotfi, R. Safdari, A. Kascenas, A. O'Neil, D. Eschweiler, J. Stegmaier, Y. Cui, B. Yin, K. Chen, X. Tian, P. Gruening, E. Barth, E. Arbel, I. Remer, A. Ben-Dor, E. Sirazitdinova, M. Kohl, S. Braunewell, Y. Li, X. Xie, L. Shen, J. Ma, K. Das Baksi, M. A. Khan, J. Choo, A. Colomer, V. Naranjo, L. Pei, K. M. Iftekharuddin, K. Roy, D. Bhattacharjee, A. Pedraza, M. G. Bueno, S. Devanathan, S. Radhakrishnan, P. Koduganty, Z. Wu, G. Cai, X. Liu, Y. Wang, and A. Sethi, "A Multi-Organ Nucleus Segmentation Challenge," *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1380–1391, 2020.

[15] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.

[16] A. Iqbal, M. Sharif, M. A. Khan, W. Nisar, and M. Alhaisoni, "FF-UNet: a U-Shaped Deep Convolutional Neural Network for Multimodal Biomedical Image Segmentation," *Cognit. Comput.*, vol. 14, no. 4, pp. 1287–1302, 2022.

[17] A. Khan, H. Kim, and L. Chua, "PMED-Net : Pyramid Based Multi-Scale Encoder-Decoder Network for Medical Image Segmentation," vol. 9, 2021.

[18] W. Wu, S. Mehta, S. Nofallah, S. Knezevich, C. J. May, O. H. Chang, J. G. Elmore, and L. G. Shapiro, "Scale-Aware Transformers for Diagnosing Melanocytic Lesions," *IEEE Access*, vol. 9, pp. 163526–163541, 2021.

[19] S. Nofallah, B. Li, M. Mokhtari, W. Wu, S. Knezevich, C. J. May, O. H. Chang, J. G. Elmore, and L. G. Shapiro, "Improving the Diagnosis of Skin Biopsies Using Tissue Segmentation," *Diagnostics*, vol. 12, pp. 1–18, 2022.

[20] S. Naidu, P. Parvatkar, A. Quadros, K. C. Kumar, A. Natekar, and S. Aswale, "Medical Image Enhancement based on Statistical and Image Processing Techniques," *Int. J. Eng. Res. Technol.*, vol. 10, no. 5, pp. 509–515, 2021.

[21] A. C. Li, S. Vyas, Y. H. Lin, Y. Y. Huang, H. M. Huang, and Y. Luo, "Patch-Based U-Net Model for Isotropic Quantitative Differential Phase Contrast Imaging," *IEEE Trans. Med. Imaging*, vol. 40, no. 11, pp. 3229–3237, 2021.

[22] A. M. Kist and M. Dollinger, "Efficient Biomedical Image Segmentation on EdgeTPUs at Point of Care," *IEEE Access*, vol. 8, no. Figure 1, pp. 139356–139366, 2020.

[23] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net : Context Encoder Network for 2D Medical Image Segmentation," *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[24] R. Su, D. Zhang, J. Liu, and C. Cheng, "MSU-Net: Multi-Scale U-Net for 2D Medical Image Segmentation," *Front. Genet.*, vol. 12, pp. 1–14, 2021.

[25] H. Hu, Y. Zheng, Q. Zhou, J. Xiao, S. Chen, and Q. Guan, "MC-Unet: Multi-scale Convolution Unet for Bladder Cancer Cell Segmentation in Phase-Contrast Microscopy Images," *Proc. - 2019 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2019*, no. 1, pp. 1197–1199, 2019.

[26] D. Ramachandram, J. L. Ramirez-GarciaLuna, R. D. J. Fraser, M. A. Martínez-Jiménez, J. E. Arriaga-Caballero, and J. Allport, "Fully Automated Wound Tissue Segmentation Using Deep Learning on Mobile Devices: Cohort Study," *JMIR mHealth uHealth*, vol. 10, no. 4, pp. 1–19, 2022.

HAFIZI MALIK (B. Eng'21) received B. Eng (Hons) in Mechatronics Engineering and is currently pursuing his Master of Science in Mechatronics Engineering at International Islamic University Malaysia. During his undergraduate studies, he primarily involved in electronics and programming for robotic and IoT applications, and his final year project was about developing clustering- and stochastic-based driving cycle prediction method while also developing a LoRaWAN-enabled electronic device for geo-fencing, tracking and communication. His current research interest includes machine learning implementation to computer vision problems.

AHMAD SYAHRIN IDRIS (B.Eng'03-M.Phil'11-Dr.Eng'18), is currently an Assistant Professor and the Deputy Head of Program at the Department of Electrical and Electronic Engineering, University of Southampton Malaysia (UoSM). He received his B. Eng (Hons) in Electrical and Electronics Engineering from University Technology Petronas, Malaysia and received his M.Phil from The University of Sheffield, UK in Electronic and Electrical Engineering. He later received his PhD in Opto-electronics from Kyushu University, Japan.

After his B.Eng degree, he joined Intel as a Product Development Engineer specializing in developing design-for-test solutions for Intel chipset products. While in the UK, he was also a researcher at the University of Sheffield specializing in the fabrication and characterization of III-V semiconductors for APD and PIN photodetectors. After his M.Phil degree, he joined Freescale Semiconductor as a Senior Test Development Engineer developing test solutions for automotive and industrial microcontrollers. His current research interests are in fabrication and characterization of opto-electronic devices and developing design-for-test solutions for microelectronic circuits. Dr. Idris is also a Professional Engineer with the Board of Engineers Malaysia and The Institute of Engineers Malaysia.

SITI FAUZIAH TOHA (B.Eng'03-M'06-PhD'10), is currently a Professor at the Department of Mechatronics Engineering, International Islamic University Malaysia (IIUM). She received B. Eng (Hons) in Electrical and Electronics Engineering from University Technology Petronas and received MSc from Universiti Sains Malaysia in electrical engineering. She then completed her Ph.D in Automatic Control and Systems Engineering from The University of Sheffield in 2010.

She was later joining the Perusahaan Otomobil Nasional Berhad (PROTON) Malaysia as a control expert consultant, working on AI-based battery management system for electric vehicle. Her current research interest spanning over the Modelling and Analysis of Complex System (MACS), Control Algorithms and Artificial Intelligence Optimisation, Assistive Devices and Bio-inspired robotics, Green Renewable Energy. Dr Toha is a senior member of IEEE and also a Professional Engineer with the Board of Engineers Malaysia (BEM) as well as a Chartered Engineer with Engineering Council, The Institution of Engineering and Technology, United Kingdom. She also served as the Engineering Accreditation Council (EAC) panel for BEM. She is also an active member of Young Scientist Network, Academy of Sciences Malaysia (YSN-ASM) and appointed as Co-chair for Science Policy Working Group (2018).

IZYAN MOHD IDRIS (MB BCh BAO, MSc) is currently a Medical Officer at the Nutrition, Metabolic & Cardiovascular Research Center (NMCRC), Institute for Medical Research (IMR) located at the National Institutes of Health, Malaysia. She received her medical degree from the Royal College of Surgeons in Ireland (RCSI) and MSc from The University of Sheffield, UK. Previously, she completed her housemanship training at Hospital Kajang, Selangor. She is currently undertaking a Ph.D in Medicine in the field of Tissue Engineering. As a medical officer in IMR, she has experience working in a biochemistry diagnostic laboratory and worked to establish a primary cell culture laboratory and bank for Inborn Errors of Metabolism (IEM). She has been involved in nationwide cardiovascular studies, including the MAVERIK study, involving field work across 12 hospital sites in Malaysia. Her research interests are in the field of cell culture and stem cells, particularly induced pluripotent stem cells and its application in disease modelling, regenerative medicine, and tissue engineering.

Izyan Mohd Idris is currently a Medical Officer at Nutrition, Metabolic & Cardiovascular Research Center (NMCRC), Institute for Medical Research (IMR), National Institutes of Health (NIH), Malaysia.

MUHAMMAD FAUZI DAUD currently holds a senior lecturer position at the Institute of Medical Science Technology, Universiti Kuala Lumpur. He joined the institute in 2013 after completing his Ph.D. at Kroto Research Institute, University of Sheffield, working on peripheral nerve engineering research. Before that, he received his undergraduate training in BSc. Biomedical Science from the University of Sheffield in 2008. His research interest includes bioscaffolds for peripheral nerve engineering, regenerative neurobiology, and microcarrier technology for cell biomanufacturing.

He was appointed the Head of Research and Innovation at the Institute of Medical Science Technology, Universiti Kuala Lumpur, and has held the position since 2019. He also contributes actively to the Tissue Engineering and Regenerative Medicine Society of Malaysia (TESMA) as the deputy honorary secretary since 2019.

MOHAMMAD OSMAN TOKHI received the BSc degree in Electrical Engineering from Kabul University, Afghanistan, in 1978, and the PhD degree from Heriot-Watt University, United Kingdom in 1988. He is currently a Professor with the School of Engineering, London South Bank University, U.K. His current research interests include Active Noise and Vibration Control, Adaptive/intelligent Control, Assistive Robotics, and bio/nature-inspired computational intelligence. He has published extensively and has executed numerous research projects to successful completion in these areas. He is a Chartered Engineer, a Fellow of Institution of Engineering and Technology (IET), and a member of International Institute of Acoustics and Vibration (IIAV) and of Climbing and Walking Robots (CLAWAR) Association. He has worked in various academic positions and in the industry.