Segmentation of Cytology Images to Detect Cervical Cancer Using Deep Learning Techniques

Betelhem Zewdu Wubineh [1[0000-0002-4790-7449]], Andrzej Rusiecki [1[0000-0003-2239-1076]], and Krzysztof Halawa[1[0000-0001-6508-0468]]

[1] Wroclaw University of Science and Technology, Faculty of Information and Communication Technology, Wroclaw, Poland

betelhem.wubineh@pwr.edu.pl, andrzej.rusiecki@pwr.edu.pl, krzysztof.halawa@pwr.edu.pl

Abstract. Cervical cancer is the fourth most common cancer among women. Every year, more than 200,000 women die due to cervical cancer; however, it is a preventable disease if detected early. This study aims to detect cervical cancer by identifying the cytoplasm and nuclei from the background using deep learning techniques to automate the separation of a single cell. To preprocess the image, resizing and enhancement are adopted by adjusting the brightness and contrast of the image to remove noise in the image. The data is divided into 80% for training and 20% for testing to create models using deep neural networks. The U-Net serves as baseline network for image segmentation, with VGG19, ResNet50, MobileNet, EfficientNetB2 and DenseNet121 used as backbone. In cytoplasmic segmentation, EfficientNetB2 achieves a precision of 99.02%, while DenseNet121 reaches an accuracy of 98.59% for a single smear cell. For nuclei segmentation, EfficientNetB2 achieves an accuracy of 99.86%, surpassing ResNet50, which achieves 99.85%. As a result, deep learning-based image segmentation shows promising result in separating the cytoplasm and nuclei from the background to detect cervical cancer. This is helpful for cytotechnicians in diagnosis and decision-making.

Keywords: Cytoplasm, Deep Learning, Nuclei, Segmentation, U-Net.

1 Introduction

Cervical cancer describes the development of malignant tumors of normal cells that initially covered the upper part of the cervix [1], which is the fourth most common cancer among women [2]–[4]. Every year, more than 200,000 women die of cervical cancer; approximately three-quarters of these deaths occur in developing countries [5], mainly due to a lack of medical resources and experts. However, it is one of the preventable diseases if detected early through screening [6]. Cytological tests, such as Pap smear, are crucial screening procedures to identify abnormalities in the cervical region [7]. To find and identify nuclear and cytoplasmic atypia, the diagnostic process requires a cellular-level examination under a microscope by a cytologist or pathologist [8]. However, manual analysis is labor-intensive, error-prone, and time-consuming [9].

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_25

Furthermore, due to factors such as the shortage of pathologists and regional economic differences, manual analysis has not been able to meet the urgent needs of patients. The presence of blood clots, mucus, overlapping cells, and other types of tissue and debris are factors that determine the quality of an image [2]. These may not be clearly visible to humans to identify abnormal cells from normal ones. Improving the screening capacity is the most effective way to reduce the incidence of cancer and save lives. This helps experts in diagnosis, reduces errors and workload, and speeds up the screening process [9]. Segmentation is one of the crucial tasks in the screening process because it can help better understand the morphological properties of cells by analyzing their constituent parts, such as the nucleus and cytoplasm [10].

The physical properties of the cytoplasm and nucleus are essential to determine whether a cell is normal or abnormal. Accurate cell segmentation helps experts identify normal and malignant cells within a Pap smear. Therefore, the screening process helps to detect abnormalities early before they become malignant [11]. To automate cervical cell segmentation and improve cervical cancer detection accuracy, deep learning-based computer-aided diagnostic techniques have been used [3]. The most challenging aspect of automating cervical cell screening is the precise segmentation of the nuclei and cytoplasm. In this study, we divide the image between the background and the cytoplasm and nuclei. In image segmentation and recognition, the U-Shape Network Structure (U-Net) has been shown to be extremely superior [12].

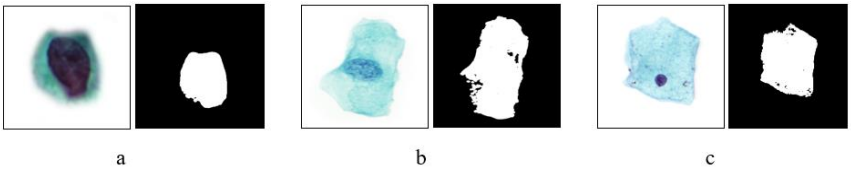
The aim of this study is to develop and evaluate deep learning (DL) models for the image segmentation of cytoplasmic and nucleic cervical cells, aiming to simplify and automate the separation of single cells. The remainder of this paper is organized as follows: Section 2 introduces the methods used, Section 3 presents the segmentation results and discusses the findings, and Section 4 provides concluding remarks.

2 Materials and Methods

In this section, the data collection, preprocessing technique, implementation detail and model development are discussed.

2.1 Dataset Collection

The dataset used for this study was collected from the Pomeranian Medical University in Szczecin, Poland. It consists of 419 cytological images with a resolution of 1130 x 1130 pixels in BMP format. In most images, the nuclei of abnormal cells appear large with blurred borders and low contrast between the cytoplasm and the background, making them highly susceptible to false predictions. The images from different classes along with their corresponding masks are depicted in Fig. 1.



a b c

Fig. 1 Sample images with corresponding mask: a) HSIL, b) LSIL, and c) NSIL

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_25

In contrast, the nuclei of normal cervical cells are tiny and exhibit high contrast. The dataset includes images classified as high-grade squamous intraepithelial lesion (HSIL), low-grade squamous intraepithelial lesion (LSIL), and normal squamous intraepithelial lesion (NSIL), with 124, 61, and 234 images respectively.

2.2 Pre-processing of the Data

Medical images are often inconsistent, noisy, and incomplete; therefore, preprocessing becomes crucial to improve the performance of the model [13]. In the preprocessing phase, we begin by creating masks for the samples collected from the hospital. Masks are created using OpenCV. The initial step involves converting the image to grayscale. Subsequently, thresholding is applied to generate a binary mask using the Otsu method. Morphological operations, such as erosion and dilation, are performed to eliminate noise and enhance the cell boundary in the mask image. Finally, mask images are generated, with the cytoplasm represented in white and the background in black. Afterwards, the images are separated into the image data and masks (labels). Following this, the dataset is divided into a training set and a testing set with a ratio of 80% and 20%, respectively. All images are resized to a size of 224 × 224 pixels. Additionally, image enhancement techniques such as adjusting brightness and contrast are applied to reduce noise. Data augmentation is then used to improve the performance and generalization of DL models and to reduce overfitting problems. In this study, geometric transformations, such as rotation, flipping, and scaling were applied. Sample images after enhancement and augmented images are shown in Fig. 2.

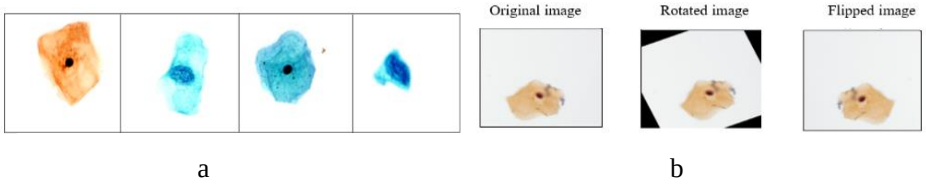


Fig. 2 Sample image a) after enhancement, b) data augmentation

2.3 Implementation Details

In this U-Net structure, the objective is to automatically extract a region of interest (ROI) around the tumor in the cervix. To train the model, it is necessary to define the experimental setup and hyperparameters. The DL framework used to implement the model is Keras with TensorFlow in Python. The experiment was conducted on a Windows 11 operating system, utilizing a 12th Gen Intel(R) Core (TM) i5-12500H processor, 32GB of RAM, and an NVIDIA GeForce RTX 4060 GPU model. The training hyperparameters used for the study are listed as follows: a batch size of 18 and training for 100 epochs without early stopping. The Adam optimizer [14] was employed with a learning rate set to 0.003. The Jaccard distance is utilized as the loss function, measuring the dissimilarity between predicted and actual sets. Minimizing the Jaccard distance aids in producing predictions that closely match the actual mask. Performance evaluation metrics for the model include accuracy, sensitivity, precision, Intersection over Union (IoU), and Dice coefficient.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_25

2.4 Training the Model

In this section, we discuss model training using a U-shaped neural network that employs data consisting of five down-sampling modules and five up-sampling modules. The down-sampling (encoder) reduces the spatial dimension of the feature maps while increasing their depth, and the up-sampling (decoder) is to decode the encoded data, utilizing information from the concatenation, and increase the spatial dimension back to the original input size. The concatenation path helps to retain spatial information during the up-sampling process. As the image size is reduced on the encoder path, the decoder path increases the image size. The down sampling includes two convolutional layers, max pooling, batch normalization, and activation function. The filter size is 3×3 for each convolution layer and 2×2 for each max pool layer. The input image is 224 x 224 x 3, and the number of filters increases in each block: 64, 128, 256, 512 and 1024. The central block has 1024 filters for 2 convolution layers. In the decoder path, transposed convolution (up sampling), concatenation, convolution layers, batch normalization and activation functions are employed. As the image size increases, the number of filters is decreased: 1024, 512, 256, 112 and 64. The final output size is 224 x 224 x1 which is the segmented image.

We then train U-Net as the base, with VGG19, ResNet50, MobileNet, EfficientNetB2, and DenseNet121 as the backbone. These pre-trained models are employed for feature extraction and are most widely used in literature. To connect the pre-trained model to the U-Net architecture, we follow the following procedures: In the encoder, the initial layer of the pre-trained model backbones captures features and reduces the spatial dimension of the input image. The layers with reduced spatial dimension connect to the deepest layer of pre-trained backbone, used for capturing abstract features and retaining high-level semantic information while reducing the spatial dimensions. Subsequently, the decoders up-sample the feature maps and reconstruct the spatial information. The transposed convolutions increase the spatial resolution, and the features of the corresponding layer in the encoder are concatenated during the up-sampling process. The cervical cell segmentation model is shown in Fig. 3.

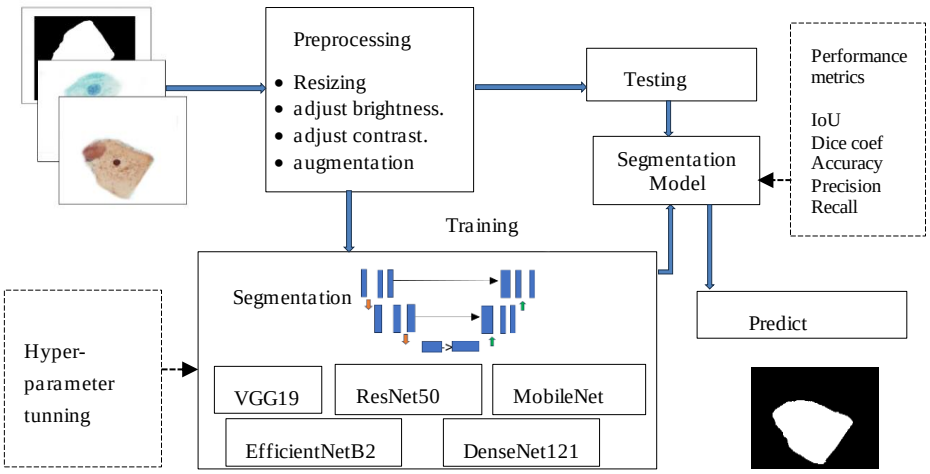


Fig. 3 The cervical cell segmentation model

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_25

In this U-Net architecture, there are two convolutional layers in each block: the first one includes a 3 x 3 convolutional layer, batch normalization, and a ReLU activation function. The second consists of a 3x3 convolutional layer, batch normalization, a ReLU activation function, and spatial dropout. For down-sampling and up-sampling, a filter size of 2x2 is used. In addition, a sigmoid activation function is applied in the last layer. These pretrained models serve as the backbone of the U-Net. The VGG 19 architecture comprises 16 convolutional layers and 3 fully connected layers, totaling 19 layers to learn weights [15]. ResNet50 can achieve a very deep network of up to 152 layers by inserting a skip connection, to pass the input from the previous layer to the next layer without altering it [16]. MobileNet is an architecture designed for mobile devices that combines efficient computation and separable convolution in depth [17] with a depth layer of 28 [18]. EfficientNet is a scaling method that uniformly scales all depth, resolution, and breadth parameters using a compound coefficient. EfficientNet substantially outperforms other convolutional networks in various tasks [19]. The baseline network, EfficientNetB0, has a subsequent network until EfficientNetB7 and EfficientNetB2 is one of these networks. DenseNet is a CNN architecture of 121 layers and is used to improve the information flow by connecting each layer to the other layer behind it. As a result, the decision is based on all layers rather than just the final layer [20]. Finally, we tested the performance of the model, which can be used to predict entire image masks and obtain relevant information about the images. This can be helpful for medical professionals who diagnose diseases and make decisions.

3 Results and Discussion

In this study, we segmented the pixels of the images into cytoplasm, nuclei, and background. Data were divided into training, validation, and testing sets. The U-Net-shaped structure served as the base model, and VGG19, ResNet50, MobileNet, EfficientNetB2, and DenseNet121 were used as backbones to segment the images. The backbones are typically used to extract the features. Accuracy, sensitivity, precision, IoU, and dice coefficient were used to evaluate performance and compare different models. Table 1 presents a comparison of the performance of cytoplasm segmentation using various approaches.

Table 1 Results of cytoplasmic segmentation

Methods	Accuracy	Sensitivity	Precision	IoU	Dice coef
U-Net	97.03	78.6	93.01	97.03	84.77
U-Net +Vgg19	98.64	93.91	92.83	97.9	80.53
U-Net +Resnet50	97.43	82.66	97.28	97.89	87.81
U-Net +MobileNet	97.67	79.75	96.61	97.18	78.52
U-Net +EfficientNetB2	98.10	83.36	99.02	97.98	88.93
U-Net +Densenet121	98.59	91.68	94.56	98.48	87.5

In Table 1, U-Net itself is used as the original network and serves as a backbone for the other pretrained architectures to segment the image. U-Net, as a base model, achieves better performance in the cervical cell segmentation task using EfficientNetB2 as the backbone, with a precision value of 99.02% and an accuracy of 98.59% in

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_25

DenseNet121. The result of the segmented nuclei is shown in Table 2. In nuclei segmentation, both EficientNetB2 and ResNet50 produced impressive results with accuracies of 99.86% and 99.85%, respectively. The U-Net architecture, with pretrained models, was employed to extract crucial features related to cervical cancer. U-Net was used to segment the Pap smear images, with the aim of identifying specific areas within cervical cells such as the cytoplasm and nuclei.

Table 2 Nuclei segmentation result

Methods	Accuracy	Sensitivity	Precision	IoU	Dice coef
U-Net	99.25	80.01	95.10	99.50	84.95
U-Net +Vgg19	99.69	86.23	99.76	99.60	84.25
U-Net +Resnet50	99.85	92.07	95.72	99.77	87.20
U-Net +MobileNet	99.78	91.63	95.17	99.71	90.34
U-Net +EfficientNetB2	99.86	97.75	98.13	99.80	89.33
U-Net +Densenet121	99.82	87.51	98.66	99.76	88.09

The ground truth (actual mask) comprises three components of pixels: the background, the cytoplasm, and the nucleus. Fig. 4 shows the results obtained from various classes of segmentation of the test set.

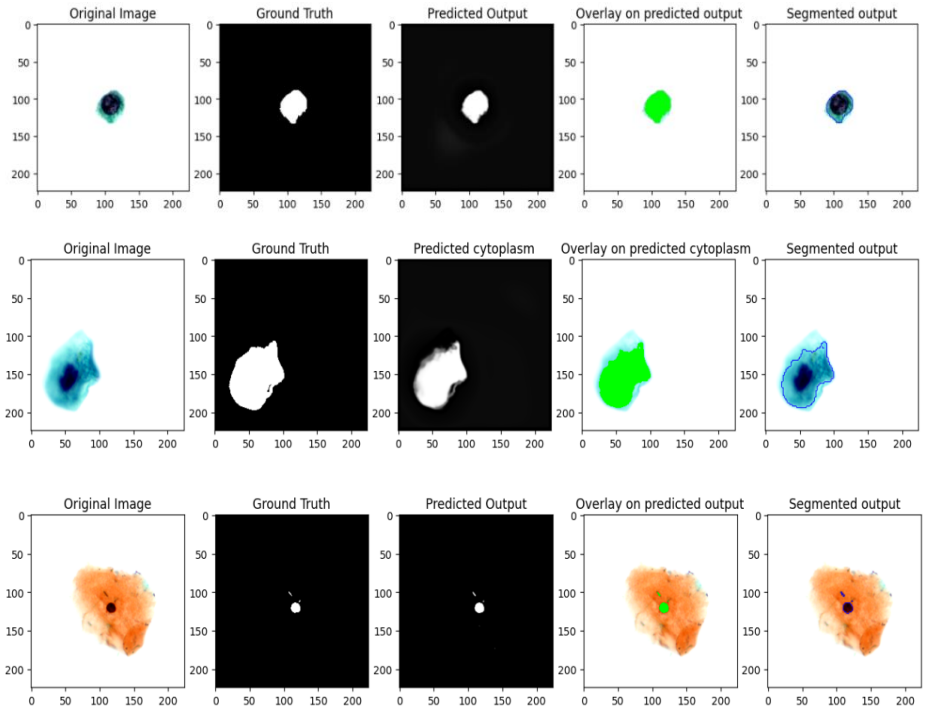


Fig. 4 Results of the test set in order of original image, actual mask, predicted mask, overlay on predicted cytoplasm/nuclei, and segmented cytoplasm/nuclei

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_25

The first two rows are the resulting cytoplasm, whereas the third is for the nuclei segmentation. The first column displays images of cervical cells. The second column presents the actual masks for the cytoplasm and nuclei, while the third column displays the model's prediction for the cytoplasm and nuclei. In the fourth column, overlays are applied to the predicted outputs, and the fifth column shows the segmented output of the cytoplasm and nuclei, respectively. The predicted column closely resembles the actual images, indicating that the model correctly separated the cytoplasm and nuclei from the background images and removed unwanted information. These results aid medical experts in identifying abnormal cells from normal ones. To fine-tune the hyperparameters, we explored different epoch values, and 100 epochs consistently produced superior outcomes.

4 Conclusions

In this study, we used the U-Net shaped structure with transfer learning algorithms as a backbone for accurate segmentation of cells in Pap smear images to detect cervical cancer in the early stages. Deep learning-based methods require large amounts of data for effective training; however, in medical imaging there is a lack of data that may impact the model's performance. When the number of images is small, transfer learning algorithms are effective in improving model performance. This involves transferring weights from a pretrained model to a specific task. We evaluated U-Net as a baseline with EfficientNetB2, which yields an accuracy of 99.02% and 99.86% for the cytoplasm and nuclei, respectively. Furthermore, DenseNet121 is the basis for this new dataset, achieving a precision value of 99.02% and an accuracy value of 98.85% in Resnet50. These results are promising for segmenting cervical cells to identify the cytoplasm and nuclei of the background. The result of this study is helpful for radiologists in making decisions in the cervical screening system.

References

[1] Xia M, Zhang G, Mu C, Guan B, & Wang M, "Cervical Cancer Cell Detection Based on Deep Convolutional Neural Network," *Chinese Control Conference, CCC*, 2020, 2020-July, pp. 6527–6532.

[2] Del Moral-Argumedo M J, Ochoa-Zezzati C A, Posada-Gómez R, & Aguilar-Lasserre A A, "A Deep Learning approach for automated Cytoplasm and Nuclei cervical segmentation," *Biomedical Signal Processing and Control*, 2023, 81(August 2022), p. 104483.

[3] Youneszade N, Marjani M, & Pei C P, "Deep Learning in Cervical Cancer Diagnosis: Architecture, Opportunities, and Open Research Challenges," *IEEE Access*, 2023, 11(December 2022), pp. 6133–6149.

[4] Alisha S & Vinitha Panicker J, "Cervical Cell Nuclei Segmentation On Pap Smear Images Using Deep Learning Technique," *2022 IEEE 3rd Global Conference for Advancement in Technology, GCAT 2022*, 2022, , pp. 1–5.

[5] Nguendo Y H B & Tchinda F C, "Descriptive epidemiology of uterine cervix cancer at the medical oncology unit of the Yaoundé general hospital-Cameroon," *GSC Biological and Pharmaceutical Sciences*, 2019, 9(1), pp. 083–091.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_25

[6] Zeleke S, Anley M, Kefale D, & Wassihun B, "Factors associated with delayed diagnosis of cervical cancer in tikur anbesa specialized hospital, Ethiopia, 2019: Cross-sectional study," *Cancer Management and Research*, 2021, 13, pp. 579–585.
[7] Harangi B, Toth J, Bogacsovics G, Kupas D, Kovacs L, & Hajdu A, "Cell detection on digitized Pap smear images using ensemble of conventional image processing and deep learning techniques," *International Symposium on Image and Signal Processing and Analysis, ISPA*, 2019, 2019-Septe, pp. 38–42.
[8] Wan T, Xu S, Sang C, Jin Y, & Qin Z, "Accurate segmentation of overlapping cells in cervical cytology with deep convolutional neural networks," *Neurocomputing*, 2019, 365, pp. 157–170.
[9] Li G, Sun C, Xu C, Zheng Y, & Wang K, "Cervical Cell Segmentation Method Based on Global Dependency and Local Attention," *Applied Sciences (Switzerland)*, 2022, 12(15).
[10] Conceição T, Braga C, Rosado L, & Vasconcelos M J M, "A review of computational methods for cervical cells segmentation and abnormality classification," *International Journal of Molecular Sciences*, 2019, 20(20).
[11] Shanthi P B, Hareesha K S, & Kudva R, "Automated Detection and Classification of Cervical Cancer Using Pap Smear Microscopic Images: A Comprehensive Review and Future Perspectives," *Engineered Science*, 2022, 19, pp. 20–41.
[12] Zhou W, Chen F, Zong Y, Zhao D, Jie B, Wang Z, *et al.*, "Automatic Detection Approach for Bioresorbable Vascular Scaffolds Using a U-Shaped Convolutional Neural Network," *IEEE Access*, 2019, 7, pp. 94424–94430.
[13] Bnouni N, Amor H Ben, Rekik I, Rhim M S, Solaiman B, Essoukri N, *et al.*, "Boosting CNN Learning by Ensemble Image Preprocessing Methods for Cervical Cancer," 2021, , pp. 264–269.
[14] Park J, Yang H, Roh H J, Jung W, & Jang G J, "Encoder-Weighted W-Net for Unsupervised Segmentation of Cervix Region in Colposcopy Images," *Cancers*, 2022, 14(14).
[15] Sharma N, Gupta S, Koundal D, Alyami S, Alshahrani H, Asiri Y, *et al.*, "U-Net Model with Transfer Learning Model as a Backbone for Segmentation of Gastrointestinal Tract," *Bioengineering*, 2023, 10(1).
[16] Ji Q, Huang J, He W, & Sun Y, "Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images," *Algorithms*, 2019, 12(3), pp. 1–12.
[17] Widiansyah M, Rasyid S, Wisnu P, & Wibowo A, "Image segmentation of skin cancer using MobileNet as an encoder and linknet as a decoder," *Journal of Physics: Conference Series*, 2021, 1943(1).
[18] Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017, [Online]. Available: http://arxiv.org/abs/1704.04861.
[19] Waziry S, Wardak A B, Rasheed J, Shubair R M, & Yahyaoui A, "Intelligent Facemask Coverage Detector in a World of Chaos," *Processes*, 2022, 10(9), pp. 1–12.
[20] Gottapu R D & Dagli C H, "DenseNet for anatomical brain segmentation," *Procedia Computer Science*, 2018, 140, pp. 179–185.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_25