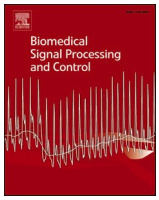
Biomedical Signal Processing and Control 77 (2022) 103783

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



# Deep learning based semantic segmentation and quantification for MRD biochip images

Fatma Çelebi [a,c], Kasim Tasdemir [c], Kutay Icoz [a,b,c,*]

[a] BioMINDS (Bio Micro/Nano Devices and Sensors) Lab, Electrical and Electronics Engineering Department, Abdullah Gül University, Kayseri 38080, Turkey
[b] Bioengineering Department, Abdullah Gül University, Kayseri 38080, Turkey
[c] Computer Engineering Department, Abdullah Gül University, Kayseri 38080, Turkey

ARTICLE INFO

*Keywords:*
Deep learning
Semantic segmentation
Transfer learning
MRD biochip
Microfluidics
Bright-field microscopy

ABSTRACT

Microfluidic platforms offer prominent advantages for the early detection of cancer and monitoring the patient response to therapy. Numerous microfluidic platforms have been developed for capturing and quantifying the tumor cells integrating several readout methods. Earlier, we have developed a microfluidic platform (MRD Biochip) to capture and quantify leukemia cells. This is the first study which employs a deep learning-based segmentation to the MRD Biochip images consisting of leukemic cells, immunomagnetic beads and micropads. Implementing deep learning algorithms has two main contributions; firstly, the quantification performance of the readout method is improved for the unbalanced dataset. Secondly, unlike the previous classical computer vision-based method, it does not require any manual tuning of the parameters which resulted in a more generalized model against variations of objects in the image in terms of size, color, and noise. As a result of these benefits, the proposed system is promising for providing real time analysis for microfluidic systems. Moreover, we compare different deep learning based semantic segmentation algorithms on the image dataset which are acquired from the real patient samples using a bright-field microscopy. Without cell staining, hyper-parameter optimized, and modified U-Net semantic segmentation algorithm yields 98.7% global accuracy, 86.1% mean IoU, 92.2% mean precision, 92.2% mean recall and 92.2% mean F-1 score measure on the patient dataset. After segmentation, quantification result yields 89% average precision, 97% average recall on test images. By applying the deep learning algorithms, we are able to improve our previous results that employed conventional computer vision methods.

## 1. Introduction

In leukemia treatment, there is a consensus on monitoring the patients for minimal residual disease (MRD) to determine the risk of relapses [1,2]. To monitor the patient's response to the treatment, the identification and quantification of cancer cells are required. Common methods for monitoring MRD are multiparameter flow cytometry (MFC), real-time quantitative polymerase chain reaction (RT- qPCR) [3], droplet digital PCR [4] and next-generation sequencing [5]. Alternatively, microfluidic systems for MRD monitoring were also been recently reported in the literature [6–8]. High-throughput images were recorded from the microfluidic platforms by using the microscopes equipped with digital cameras, later the recorded images were analyzed by powerful processors and algorithms [9,10].

Previously, we developed a microfluidic system that incorporates micron-size immunomagnetic beads to capture and separate the target leukemic cells. The magnetic bead-loaded cells were introduced into a microfluidic platform and captured by the antibodies that were immobilized on the square-shaped gold micropads [7]. Our design incorporated immunomagnetic beads and micropad arrays which functionalized with different types of antibodies, resulted in screening target cells for multiple antigens. Image recording and analysis were implemented as a readout method for the microfluidic platform [9]. An automated detection and quantification of not only cells but also the micropads and the immunomagnetic beads were needed. Our main goal was to implement robust computer vision approaches to detect and quantify the leukemia cells, the micropad arrays, and the immunomagnetic beads with high accuracy and low error rate. We reported our

* Corresponding author at: BioMINDS (Bio Micro/Nano Devices and Sensors) Lab, Electrical and Electronics Engineering Department, Abdullah Gül University, Kayseri 38080, Turkey.
*E-mail address:* kutay.icoz@agu.edu.tr (K. Icoz).

https://doi.org/10.1016/j.bspc.2022.103783
Received 13 December 2021; Received in revised form 18 January 2022; Accepted 27 April 2022
Available online 6 May 2022

previous efforts for developing image processing algorithms to quantify cells, magnetic beads, clusters [11] and micropads from the images of the microfluidic system recorded using a bright-field microscope [9]. To reduce processing time, we developed the image processing algorithms based on the identification of variations in color, size and shape of the objects in the images [9]. However, separate algorithms to detect and quantify the leukemia cells, immunomagnetic beads and micropads having different colors, shapes, and sizes had to be developed. For instance, the watershed segmentation algorithm [12] was used to detect and quantify immunomagnetic beads but watershed did not perform well for the cells and clusters of immunomagnetic beads. Also, the optimized parameters, for instance, structuring element size and shape for morphological operations differed for cells, immunomagnetic beads, and micropads and even some parameters differed from patient-to-patient samples. Therefore, all objects had to be detected separately and the results had to be combined for a final evaluation. For some low-quality images such as blurry and nonuniform light-effected images, a preprocessing was required before implementing the developed algorithm. To improve the results of the previous methods and to have a more generalized system, we decided to apply deep learning algorithms to the same patient dataset.

Deep learning techniques are widely used on biomedical images for classification, segmentation, and object detection. In [13], a broad overview of how the optimization of deep learning architectures used for medical image classification and segmentation was reported. In fluorescence images, new evaluation strategy to measure accuracy, types of errors and computational efficiency for nucleus segmentation using DeepCell and U-Net architectures were proposed [14]. 20.000 high-content microscopy yeast cell images were classified based on their yeast protein localization by using deep convolutional neural network (DeepLoc) [15]. In a review [16], intersection between deep learning and cellular image analysis were explained extensively. One of the important intersections was the segmentation of microscopy cell images [17–20]. By using a scanning electron microscope (SEM), blood cell images, red blood cells (RBCs), white blood cells (WBCs) and platelets were segmented by a novel convolutional encoder-decoder framework. VGG-16 was used as pixel-level feature extraction model and class-wise accuracy was stated as 97.45%, 93.34%, and 85.11% for RBCs, WBCs and platelets, respectively [21]. Another intersection was the classification of the cell images. Large number of single-cell label free bright-field images where each image belongs to one of the four types of the cell samples were classified by using convolutional neural network (CNN) with above 99% accuracy. Then, CNN performance was compared with k-nearest neighbors (kNN) and support vector machine (SVM) based methods [22]. In another study, the performance of detecting Plasmodium by different machine learning and deep learning techniques were proposed. CNN architecture was used for classification and performance of AlexNet, ResNet, VGG-16, and DenseNet architectures were compared on the dataset by transfer learning approaches [23]. U-Net architecture with a pre-trained model on single-cell was adapted to ImageJ image processing tool that enables non-machine-learning experts to analyze their own data [24].

In [13], it was also mentioned that deep learning applications offer solutions to problems such as diagnosis and monitoring the cancer. In microscopic images, Attention DeepLabv3 + deep learning model was applied to detect and segment multiple stained myeloma plasma cells in microscopy images [25]. For the detection of leukemia from microscopic images, transfer learning with AlexNet was applied. Dataset had 2820 images of stained cells and classification accuracy was reported as 100% [26]. Classification of Acute Lymphoblastic Leukaemia (ALL) to B-cell type and T-cell type by using AlexNet and LeukNet which were pre-trained convolutional neural networks was proposed. 336 images of both B-cells and T-cells were obtained from publicly available image databank of American Society of Haematology (ASH) and classification accuracy was reported as 94.12% [27]. In another study, it was stated that label-free, bright-field, and dark-field images of single leukemic cells were identified by ResNet deep learning architecture with more than 88% accuracy [28].

Although there are numerous studies employing deep learning techniques were reported for the segmentation and the classification of all biological objects such as leukemia cells, nucleus, normal and abnormal cells, to our best knowledge, the segmentation of mixture of object such as patient leukemia cells, immunomagnetic beads, bead clusters and micropads from the images acquired using a bright-field microscope have not been reported.

In this study, as a readout mechanism of the microfluidic platform, we propose an automated semantic segmentation algorithm for the bright-field microscopy images of cells, immunomagnetic beads and micropads. We used U-Net [29] architecture by adding batch normalization layer after each convolution layer and by optimizing some hyper-parameters of architecture. Then, we implemented DeepLabV3 [30] and Fully Convolutional Networks for Semantic Segmentation(FCN) [31] architectures to compare the results with proposed U-Net.

The major contribution of our study is to segment leukemia cells, immunomagnetic beads, and micropads from bright-field microscopy patient images, which does not have any staining, using U-Net deep learning segmentation approaches with parameter optimization, solving unbalanced dataset problem, and adding batch normalization layer after each convolution layer, then to quantify cells automatically after segmentation. The proposed approach gives better results compared to the other deep learning segmentation methods.

The paper is organized as follows: Section II is devoted to materials, and in section III, methods are reported. In section IV, the results are presented followed by the conclusion and discussion.

## 2. Materials

The details of the MRD Biochip fabrication and principle of operation were reported in [7], briefly photolithography, gold deposition and lift-off process were used to fabricate micropads on a glass substrate. Then the glass substrate including the matrix of pads were placed inside a microfluidic chamber formed by layers of glass (bottom) and polymethyl methacrylate (PMMA) (middle and top). PMMA layers were cut by using a laser cutter to and double-sided adhesive tapes sealed the layers to prevent from any fluid leakage.

Monitoring a cancer patient who received treatment for possible relapses in the future is critical for survival. Screening the cells for one type of biomarker is not sufficient to detect blast cells so detection of multiple markers is required. Flow cytometry is conventionally used to detect multiple markers on the cell membrane which can indicate the blast cell, however trained personnel and expensive equipment are needed to perform analysis. We showed that the MRD Biochip is able to screen cells having three different markers simultaneously. Micro and nano size magnetic particles were conjugated with different types of antibodies and another type of antibody was immobilized on the micropads. Double immunomagnetic separation of target cells from patient samples were followed by capturing cells in the microfluidic system. The captured cells were imaged and analyzed for quantification.

For this study, we extended the dataset introduced in [9]. Minimal residual disease (MRD) biochip bright-field microscopy images were acquired by DS-Ri 1 model CCD color camera and 20X objective (Nikon Instruments, Melville, NY). The size of the images were 3116 × 4076 pixels and total number of images was ninety-one.

Proposed deep learning algorithms were implemented on Matlab (R2020a, The MathWorks Inc., Natick, MA), and Python. Matlab codes were run on a PC with an Intel i7 − 7820HQ, 2.90 GHz CPU, and 8 GB RAM on it. Python codes were run on high-performance-computing cluster. While on Matlab only U-Net architecture was implemented and tested, on Python U-Net and DeepLabV3 and FCN architectures were implemented, and results were compared.

## 3. Methods

### 3.1. Image annotation

In a recorded image, the process of classifying each pixel to the particular label is known as semantic segmentation. During the semantic segmentation, deep learning network model is trained with train dataset, which contains images and their corresponding pixel-labeled ground-truth images. Therefore, ninety-one patient images were manually annotated using Adobe Photoshop software to get ground-truth images. In Fig. 1. Top row: a general view of high-resolution image, cropped image and cropped image annotation is given for better visualization. In the images, there are 5 different object classes to segment. These object classes are 1) cells, 2) immunomagnetic beads, 3) micropads, 4) background, and 5) bead clusters that refer to a cell that was fully covered by immunomagnetic beads. When the cluster occurs, cells are covered with many beads and become invisible, but we must identify those cells too. That's why, we defined a new class as cluster in the images. However, all bead clusters don't indicate underlying cells because, some beads attract to each other due to the magnetization and form a linear chain like structure [32], thus while annotating we were careful about this issue, and we determined clusters based on these properties; three or more beads had to bound each other, and after bounding their shape had to be circular. In addition, sometimes less than three beads can be bound to micropad and made cells invisible. This case was also annotated as a cluster. In Fig. 1. bottom row, some examples of clusters were shown. Cells have irregular shapes and sizes in the image. Beads have uniform circular shape. Micropads have regular square shapes, but different sizes of micropads were tested for different patient samples.

### 3.2. Overview of the MRD biochip image segmentation workflow

Our general segmentation workflow is given in Fig. 2. Raw images were captured from MRD biochip by the bright-field microscope. Then, raw images were divided into small images without overlapping. Input image size for the network was 224x224x3. At the end of the segmentation network, each pixel of the input image was classified as one of the five groups: cell, cluster, bead, micropad or background.

### 3.3. Semantic segmentation with U-Net architecture

U-Net is a semantic segmentation network, and it is commonly used in biomedical image segmentation [29]. U-Net is a frequently used deep-learning architecture for cell detection and shape measurements in biomedical images [24]. U-Net is a convolutional neural network model, and it consists of contracting path and expansive path. It is stated that even with small dataset it performs well [29]. Our U-Net network had convolutional layers followed by batch normalization layers [33,34] and activation function (ReLU) [35] layer, max-pooling layers and up-convolutional layers. The general view of the architecture is shown in Supplementary Fig. 1.

We have modified the original U-Net structure by adding batch normalization (BN) layer after each convolution layer as regularization. The input image size was set to 224x224x3, and weighted softmax loss function was used.

### 3.4. Unbalanced dataset problem and transfer learning

Training with unbalanced class dataset is a common problem for semantic segmentation models [36]. This problem leads to poor network performance on minority class, which is the center of attention for our case. In our images, coverages in terms of pixel area ratio are 72% background, 22% micropads and 5% immunomagnetic beads, 0.2% cells and 0.8% clusters. Because of unbalanced class, cells were segmented as background and clusters were segmented as micropads or as immunomagnetic beads. To overcome this issue, transfer-learning-oriented method was used. Method consisted of firstly training the network architecture on a balanced dataset, then using class weights of pre-trained network as initial weights for fine tuning training the unbalanced

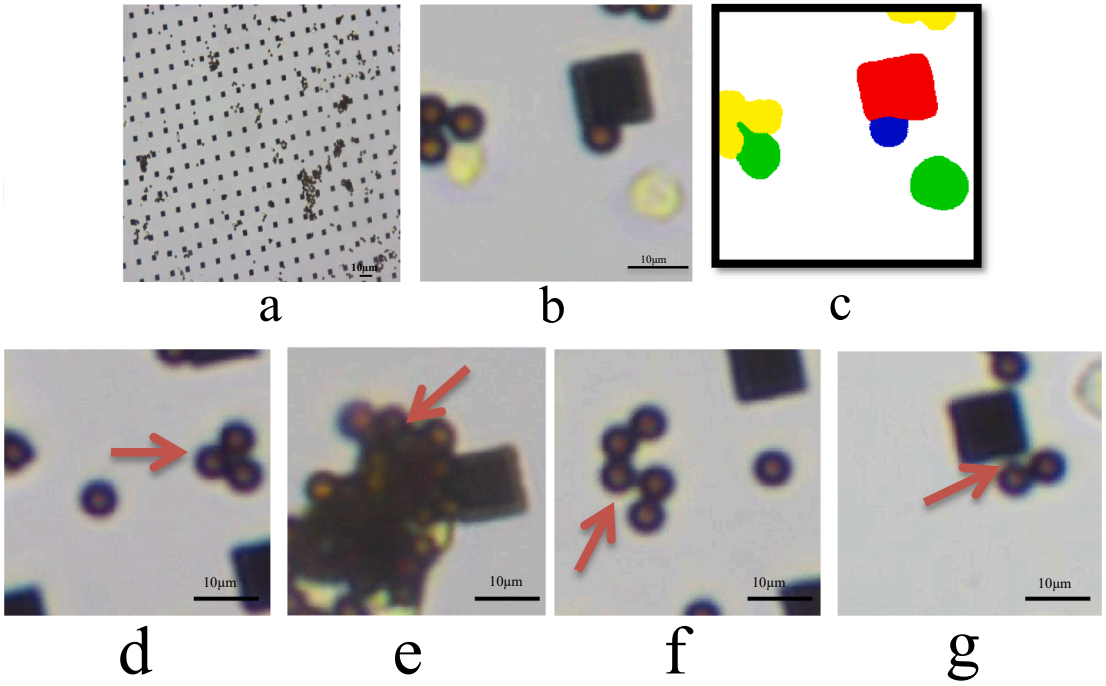



**Fig. 1.** Top row: The general view of image and its annotation. a) Original bright-field microscopy image obtained from the microfluid platform. b) Cropped image c) Annotated cropped image, green, yellow, red, blue, and white colors, respectively, represent cell, cluster, micropad, immunomagnetic beads, and background. Bottom row: Clusters can appear in 4 different forms as indicated by arrows. d) Three beads, e) more than three beads are sticked together and the shape is circular. f) More than three beads sticked together, but shape is not circular, so it is not cluster. g) Less than three beads bounded to micropad and occluded the cell.

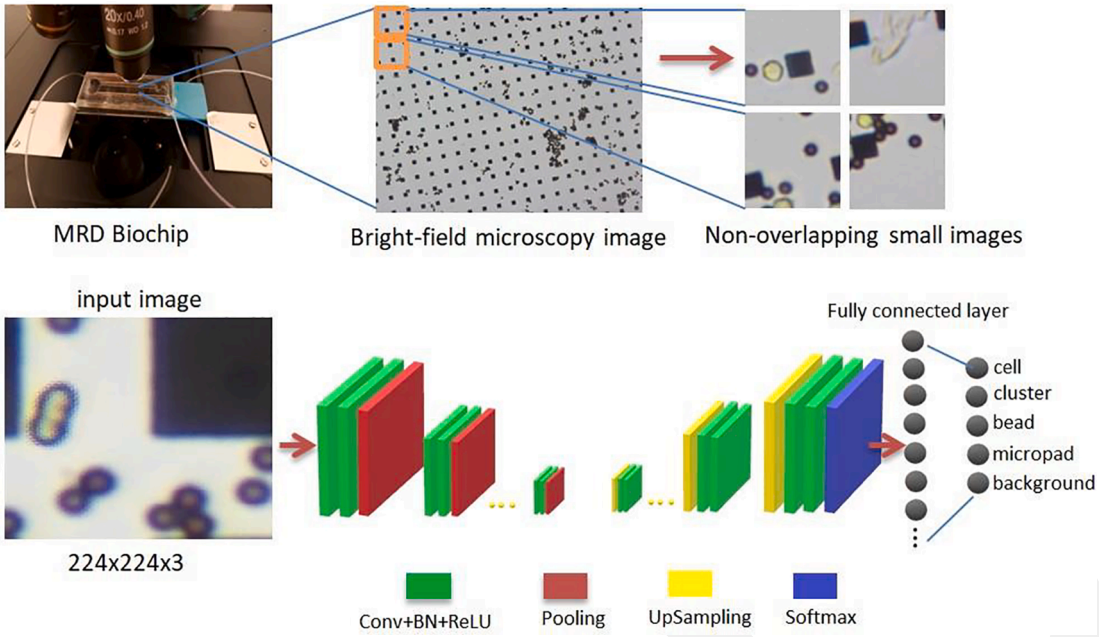


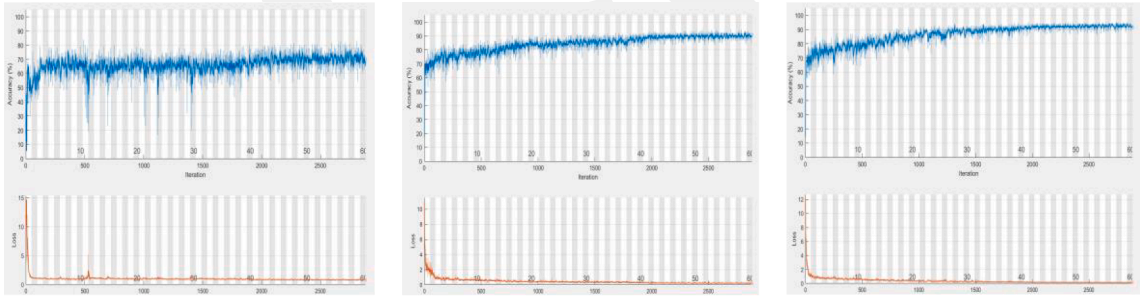

**Fig. 2.** Architecture of the pixel-wise classification network. The network was trained to classify the pixels into one of the 5 classes: cell, cluster, bead, micropad, background.

dataset [37]. To gather balanced class datasets, isolated (single) cells, clusters, micropads and immunomagnetic beads images was cropped from the big size images. Then, cropped images were resized with bicubic interpolation. Finally, network was trained on balanced dataset and weights were transferred to train an unbalanced dataset.

### 3.5. Experimental setup

#### 3.5.1. Dataset

To create balanced dataset, cropped image number for each class is presented in Supplementary Table 1. Initial weights for the original dataset training were obtained from training on a balanced dataset. We had 90 images of patient samples that are of size $3116 \times 4076 \times 3$. 65 high-resolution images were used for training, 7 high-resolution images were used for validation and 18 high-resolution images were used for



a b c

| | Section Depth | Initial Learning Rate | Max Epoch | Mini Batch Size | L2Regularization |
|---|---|---|---|---|---|
| **a** | 3 | 0.0006422 | 56 | 8 | 0.0012376 |
| **b** | 5 | 0.00037569 | 92 | 7 | 0.00068294 |
| **c** | 5 | 0.00041314 | 78 | 4 | 0.00018771 |

d

**Fig. 3.** (a-c): Accuracy and Loss result graphs of U-Net with different hyper-parameters. (d): Parameter values for each training in a, b, and c graphs. The graph shows that with the given hyper-parameter selection, the model (c) can learn the class representations without any overfitting or underfitting of the data.

testing which roughly corresponds to 80%/20% train-test, 90%/10% train-validation distribution. This percentages are commonly used in deep learning studies [38–40]. Our network accepted input image of size 224 × 224 × 3 pixels. First, our images were resized to 2912 × 4032 × 3 to be able to divide them smaller non-overlapping windows of 224 × 224 × 3 pixels. After that, all ninety-one high resolutions and their annotated images were equally and without overlapping divided into 224 × 224 × 3. Finally, 21,294 image patches were obtained for dataset.

#### 3.5.2. Hyper-Parameter optimization

Hyper-parameters such as initial learning rate, maximum epoch, mini batch size can dramatically affect the performance of the network (Fig. 3). It can be seen that optimum parameters can increase the accuracy of network by 20%. Thus, it is important to tune the hyper-parameters of the network. To find the optimum parameters, Bayesian optimization algorithm, which uses the information of the previous experiments to enable optimal choice of next parameters to test was run on the balanced dataset [41]. We applied the optimization algorithm for the depth of U-Net architecture, the initial learning rate, the maximum epoch, the mini batch size and the L2Regularization hyper-parameters.

Adam optimizer was used as a solver for training the network [42]. Training dataset was reshuffled at every epoch. Learning rate scheduler was used to decrease learning rate of the training dynamically with 0.1 drop factor and with 40 drop period.

### 3.6. U-Net, DeepLabV3 and FCN architectures

Codes were implemented using the Pytorch library. Having noisy images and different patient images in both train and test set were crucial for a reliable model. For that reason, we ensured that the test set contains samples from each image variations. After the split, 15,444 images were used for training, 1638 images were used for validation, and 4212 images were used for testing. For the maximum epoch 32, 50, 64, and 128 and for the mini batch size 16, 32, 54, 64 were tested.

We wanted to compare the segmentation performance of modified U-Net architecture with DeepLabV3 that is semantic segmentation model uses atrous convolution in parallel for multi-scale context capturing [30], and FCN architectures.

DeepLabV3 model with a ResNet-50 and ResNet-101 backbone, and FCN model with ResNet-50 and ResNet-101 backbone were used for comparison. Initial learning rate, L2Regularization, and mini batch size hyper-parameters were the same for all 5 architectures. All parameter values are reported in Supplementary Table 8. We have tried different epoch number and we have chosen the epoch number for each network based on achieving the minimal loss and higher cross-validation performance. This approach is given in similar studies [40]. As maximum epoch, 50 for U-Net architecture, 16 for DeepLabV3 architectures, and 32 for FCN architectures were used.

### 3.7. Quantification of cells and clusters

It is important to quantify the number of cells in the given image to be able to calculate the density of leukemia cells. The cells appear in two distinct ways. One of them is isolated and clearly visible cells, and the other one is occluded cells that are partially or fully covered by the immunomagnetic beads, which is named as “cluster” in the text. The workflow of quantification process in given in Fig. 4. To be able to quantify number of cells and clusters, segmentation result images were used. First of all, high-resolution images were cropped to small sized image patches and segmentation algorithm is applied on cropped image patches. The resulted patches are again merged to get high-resolution segmentation maps and they are used to quantify number of cells and clusters. Quantification was done on high-resolution images, because some part of the same cell and cluster can be seen in one patch, while the other part can be seen in another patch. In addition, it is needed to get total number of cells and clusters on the high-resolution images which refers our original dataset image. After merging patches to their original high-resolution image, on image, segmented cells and clusters obtained in binary image format. After that, to merge cell parts which are close to each other and belong to the same cells and cluster parts which are close to each other and belong to the same cluster, by using disk-shape structuring element morphological closing operation is applied. After closing for cleaning small cell parts, cluster parts and noisy parts, disk-shaped structuring element is used and morphological opening operation is applied. Finally, on the binary image, object quantification is applied and number of cells and clusters is found. For evaluation of quantification, the results are compared with the expert’s evaluation results.

## 4. Experimental results

21294 images with 224 × 224 × 3 size, which belong to four different patients and were collected at different times such as on diagnosis and after 15 days from the diagnosis [9], were segmented by the proposed U-Net algorithm. Images were annotated pixel by pixel according to their classes. With the proposed U-Net segmentation algorithm cells, immunomagnetic beads, micropads, clusters and background were segmented. In Fig. 5, an example of original 224x224x3 images, their annotated images and their segmentation results are shown.

### 4.1. Experimental setup using U-Net and test results

To determine the performance of the proposed segmentation algorithms, accuracy, intersection over union (IoU) [43] and BF-Score [44] were considered.

Accuracy shows the percentage of the correctly classified pixels and calculated as follows;

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP refers to true positive, TN refers true negative, FP refers false positive, and FN refers false negative.

The intersection over union (IoU) metric used in segmentation algorithms and it is a ratio between overlap and union of ground-truth labels and prediction labels. IoU is calculated as follows;

$$IoU = \frac{ground\ \ truth\ \ labels\ \ \cap\ \ prediciton\ \ labels}{ground\ \ truth\ \ labels\ \ \cup\ \ prediciton\ \ labels} \tag{2}$$

BF score means boundary F1 score, and it indicates how well-predicted boundary of each class aligns with the true boundary and calculated as follows;

$$BF\ \ Score = \frac{Precision^{*}Recall^{*}2}{Precision + Recall} \tag{3}$$

Results were obtained with Adam optimizer with initial learning rate 0.00041314, maximum epoch 16, and mini batch size 8. The result of matrices for each class on the test dataset is given in Supplementary Table 2. The tests were implemented on Matlab.

Annotation was done on high resolution images, and then they were divided into smaller images. On high-resolution images, the part which beads are sticked together and the shape is circular cannot be seemed as circular after non-overlapping crop. Therefore, in smaller images those cluster parts were missed and segmented as bead. IoU for cluster is lower than the others because, IoU value is calculated by division of the overlap area between ground truth image and segmentation result image by the union area between ground truth image and segmentation result image. The average value of matrices for all classes on the test dataset is given in Table 1.

As a result, the accuracy for each class was more than 85%, IoU was more than 50%, which is close to the state-of-the-art methods in biomedical image datasets, and BFScore was also more than 50%. Global

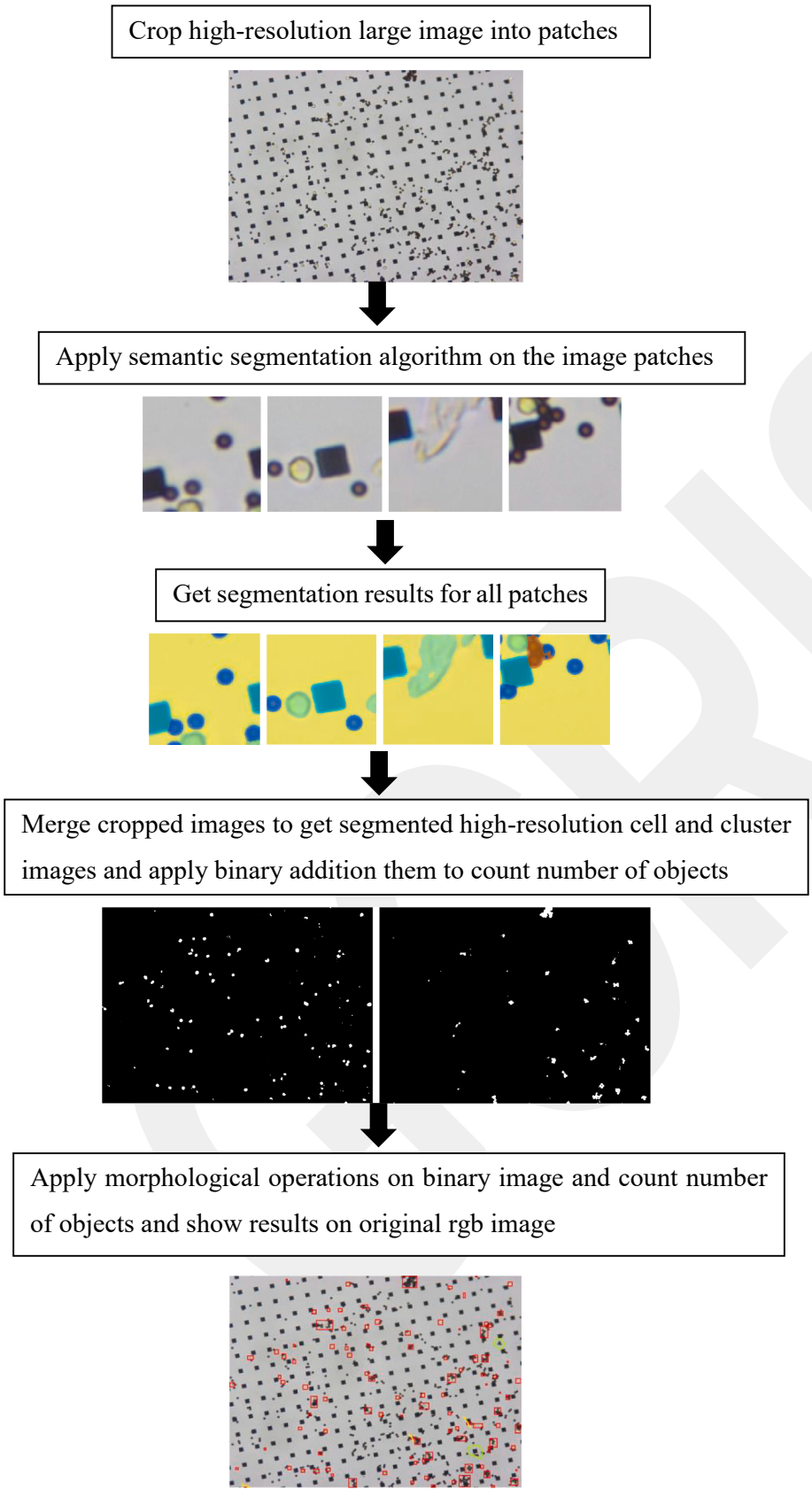



**Fig. 4.** Quantification of cells and clusters workflow. High-resolution RGB image is cropped into small image patches and segmentation is done on them. Then, after segmentation, patches are merged and obtained segmented cell and cluster high-resolution binary image. After that because clusters are also cell in our case, two segmented images are added to find total number of cells and by using morphological operations object counting on binary image is completed. Finally, counting results are shown on original RGB image and evaluation is done by manual counting.

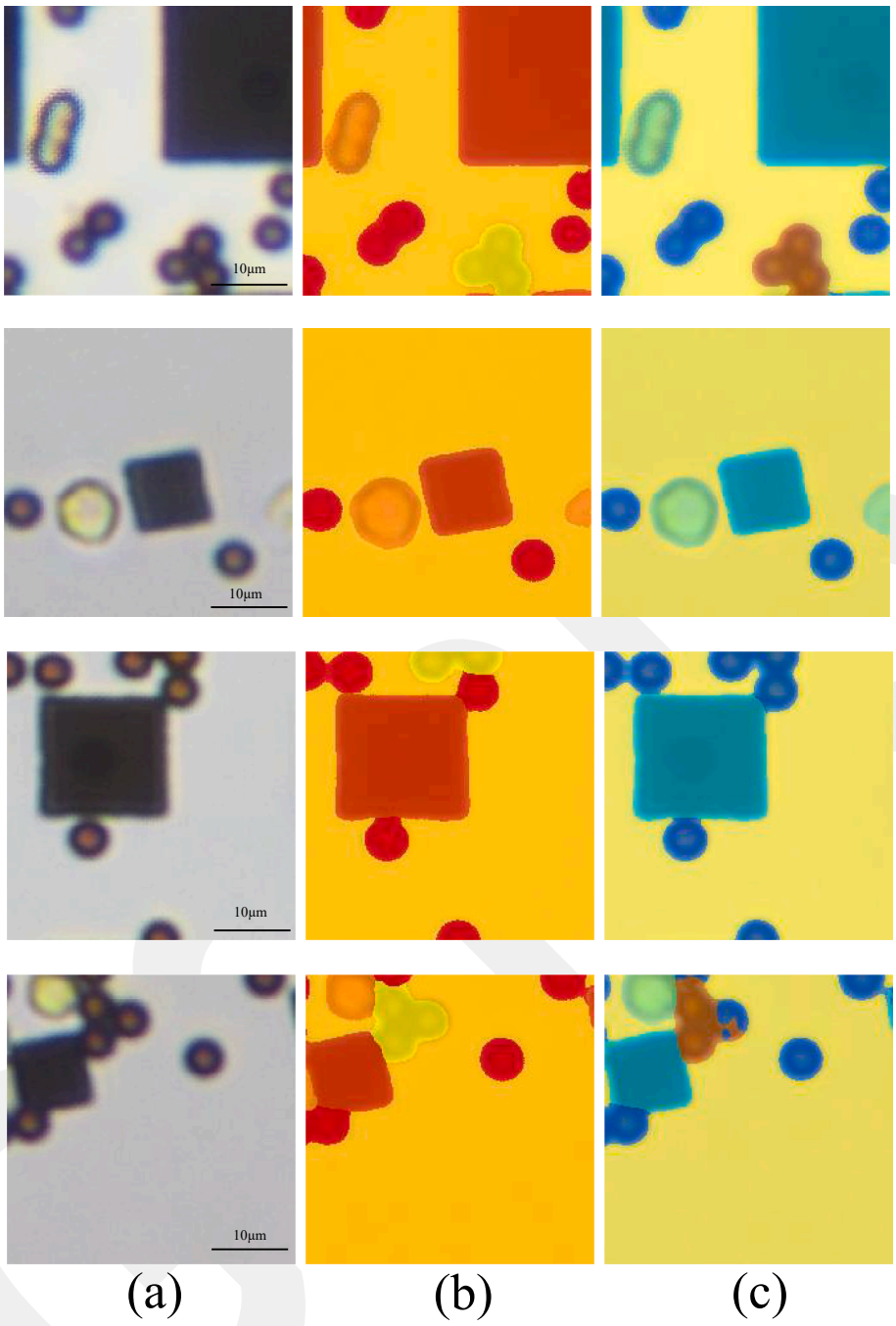



**Fig. 5.** Results of U-Net segmentation on test images. Left to right (a) input images, (b) ground truth images, (c) segmentation results.

**Table 1**
U-Net segmentation average metrics for 5 classes.

| Global Accuracy | Mean Accuracy | Mean IoU | Weighted IoU | Mean BF Score |
|---|---|---|---|---|
| 0.99 | 0.93 | 0.84 | 0.98 | 0.95 |

accuracy was 99% that is more accurate than similar studies [37,45].

The proposed network performance on training and validation dataset is given in Supplementary Fig. 2.

From Supplementary Fig. 2, it can be seen that validation accuracy was reached 99% and training loss were less than 0.05.

We modified original U-Net architecture by adding BN layer after each convolution layer as regularization. To see the effect of adding new layer to network, we trained small number of datasets. The effects of adding BN layer on average metrics is shown in Table 2, and the effects on class metrics is shown Table 3. Results show that adding BN layer increased overall performance of U-Net architecture by increasing performance of all evaluation metrics.

The tests were repeated with U-Net implementation on Pytorch. This

**Table 2**
U-Net segmentation average metrics with BN, and without BN layer.

| | Global Accuracy | Mean Accuracy | Mean IoU | Weighted IoU | Mean-BF Score |
|---|---|---|---|---|---|
| U-Net | 97.0% | 83% | 73% | 94% | 85% |
| U-Net + BN | **97.5%** | **87%** | **78%** | **95%** | **87%** |

**Table 3**
U-Net segmentation class metrics with BN, and without BN layer.

| Class Name | Accuracy | Accuracy(U-Net + BN) | IoU | IoU(U-Net + BN) | MeanBF Score | MeanBFScore(U-Net + BN) |
|---|---|---|---|---|---|---|
| Cell | 0.746 | **0.799** | 0.533 | **0.593** | **0.536** | 0.497 |
| Cluster | 0.575 | **0.656** | 0.524 | **0.632** | 0.262 | **0.396** |
| Bead | 0.909 | **0.941** | 0.701 | **0.760** | 0.775 | **0.820** |
| Micropad | 0.978 | **0.983** | 0.943 | **0.948** | 0.902 | **0.921** |
| Background | **0.986** | 0.979 | 0.979 | **0.979** | 0.948 | **0.951** |

time to evaluate the performance of segmentation algorithm on our dataset, accuracy, precision, recall and f1-score (the same with BF Score) was calculated, which are given by.

$$Precision = \frac{TP}{TP + FN} \tag{5}$$

$$Recall = \frac{TP}{TP + FP} \tag{6}$$

In addition, mean intersection over union value was calculated. The global accuracy was 98.7% and the mean IoU was 86.1%. Results of other matrices on each five classes are shown in Supplementary Table 3.

These results were achieved with initial learning rate 0.00041314, maximum epoch 50, mini batch size 16, and U-net depth 5. As a result, the accuracy for cell, bead, micropad segmentation, and mean IoU for all class segmentation increased because, bigger mini batch size and bigger maximum epoch number was able to train the architectures.

### 4.2. DeepLabV3 and FCN

To compare DeepLabV3 and FCN architectures with proposed U-Net performance, the same evaluation matrices were used, and results of architectures on five classes are given in Supplementary Table 4. Supplementary Table 5. Supplementary Table 6. Supplementary Table 7.

In Supplementary Table 4. Supplementary Table 5. Supplementary Table 6. Supplementary Table 7, results show that, different architectures give different value for performance metric on five classes. However, to see overall performance of architectures on our dataset, we calculated mean accuracy, mean IoU, mean precision, mean recall, and mean F1-score. Overall performance comparison of all five semantic segmentation algorithms is given in Table 4.

As a conclusion, proposed U-Net, DeepLabV3, and FCN semantic segmentation architectures have similar performance on our dataset when we compare overall performance metrics. However, even the difference is small, proposed U-Net architecture gives the best results on performance metrics. In addition, deeper network like DeepLabV3-Resnet101 and FCN-Resnet101 show better performance on accuracy, IoU, and recall values compared to DeepLabV3-Resnet50 and FCN-Resnet50.

### 4.3. Automated quantification after segmentation

In our previous study [9], after identification of each object, we made an automated quantification for each object by using morphological operations and results of the quantification were compared with the manual counting of an expert. In this study, to observe the quantification results, after the segmentation, by using morphological operations, we counted each labeled cells and clusters automatically and we used again the same manual counting for evaluation. In our previous study, beads, micropads and background quantification error were less than 0.02, and also cell and cluster quantification were critical. Therefore, we only focused on quantification results of the cell and cluster. Quantification is applied on test images of dataset and results are given in Table 5. It can be seen in Table 5 that recall level reaches 97% success. This indicates that there is low number of missed detections. In addition, precision level reaches 89% success. This indicates that there is low number of wrong detections (false positives).

In this proposed study, to see the improvement compared to our previous study, quantification results of cells and clusters on test images with previous study and proposed study is compared and results are given in Table 6. Results showed that all evaluation matrices have been improved. Besides, to show statistical relation between expert's manual counting and automated counting, we used Mann-Whitney *U* test [46] because, we had two independent groups with limited amount of data. Our hypothesis was H0: Statistically there is no difference between manual and automated counting, H1: Statistically there are differences between manual and automated counting. When we test four patient's manual and automated counting results, we got 0.64 Asymp. Sig. Value which is bigger than 0.05. This means that, hypothesis H0 was kept, and concluded that there was not statistically difference between the expert's manual counting and the automated counting.

**Table 4**
Overall performance comparison between 5 different semantic segmentation architectures.

| | U-Net + BN | DeepLabV3-Resnet50 | DeepLabV3-Resnet101 | FCN-Resnet50 | FCN-Resnet101 |
|---|---|---|---|---|---|
| Accuracy | **98.7%** | 98.2% | 98.3% | 98.2% | 98.3% |
| IoU | **86.1%** | 85.6% | 86% | 85.8% | 85.8% |
| Precision | **92.2%** | 92% | 92% | 92% | 91.8% |
| Recall | **92.2%** | 92% | 91.6% | 92.2% | 92.4% |
| F1-Score | **92.2%** | 92% | 91.7% | 92.1% | 92% |

## 5. Conclusion

The proposed method is designed to segment objects in the images which is captured from a biochip by bright-field optical microscopy. The images include target B lymphoblast cells, immunomagnetic beads and micropads. Without cell staining, U-Net + BN convolutional neural network architecture with transfers learning from balanced dataset and with hyper-parameter optimization achieved 98.7% global accuracy, 86.1% mean IoU, 92.2% mean precision, 92.2% mean recall and 92.2% mean f-1 score measure on a patient dataset for semantic segmentation. Outcomes of the proposed U-Net is compared with the outcomes of four different semantic segmentation architectures. The results show that all the architectures have similar results on our patient dataset, but U-net gives the best performance. In addition, the results show that even if micropad size and bead density change, segmentation algorithms can detect cells. When determining the performance of the algorithm on test parameters set such as precision, recall, F-measure and others, we used annotated images of test images. In addition, after segmentation with U-Net, for quantification results, we compared the outputs of the algorithm with the output of our manual evaluation of the images.

In the proposed study, annotation of images is done on high-resolution images, and then images and their associated annotations are cropped to smaller patches without overlapping for training and testing. Not using overlapping windows cause some problems. In high-resolution images, if three or more beads bind each other, and after bounding their shape are circular, these bound beads are annotated as cluster. However, when images are cropped to smaller patches, some cluster parts can go to one patch, while other parts can go to different patches, and this time they seem as a bead not a cluster due to number of beads and shape appearance. The proposed method segments these

**Table 5**
Overall performance of quantification of cells and clusters on test images.

| Average TP | Average FP | Average FN | Precision | Recall | F-Measure | Detection Error |
|---|---|---|---|---|---|---|
| 46.5 | 3.6 | 1.2 | 0.89 | 0.97 | 0.93 | 0.06 |

**Table 6**
Comparison of quantification results between our previous study and proposed study on test images. Improvement rate is calculated as (Proposed method's result – the previous method's result) / |ideal result - previous result| * 100.

| | Precision | Recall | Detection error | F-Measure | Max-Precision |
|---|---|---|---|---|---|
| Previous study [9] | 81% | 84% | 16% | 79% | 95% |
| Proposed Method | 89% | 97% | 6% | 93% | 99% |
| Improvement Rate | **9.8%** | **15.4%** | **62.5%** | **17.7%** | **4.2%** |

clusters as bead, and segmentation performance is evaluated based on the annotated images, so evaluation metrics result for cluster are lower than other objects evaluation metrics which is given in Table 3. To increase the performance of cluster segmentation, cropping with overlapping windows will be tested as a future work. Moreover, not to break clusters into pieces, model structure can be changed to analyze high-resolution images by using image pyramid techniques as a future work.

## 6. Discussion

Segmentation of cells from microscopy images were reported in [47] with 96.1% mean accuracy and 92.1% mean IoU on stained images, in [21] with accuracies of 97.45%, 93.34%, and 85.11% for RBCs, WBCs, and platelets, respectively, and mean IoU of 0.31%. However, cells were stained which makes them distinctive. Also, in these reports there was no unbalanced dataset problem. However, in this work, cells are not stained, an image includes other biomaterials, which have more pixel area than cells (unbalanced data). The proposed method achieved the mean accuracy 98.7%, mean IoU 86.1%, and mean BFScore 92%.

In our previous study [9], we used conventional digital image processing methods to detect and quantify cells, clusters, immunomagnetic beads, and micropads which benefit from color, size and shape. The critical task was the quantification of cells and we obtained 85% precision and 86% recall value. However, for each object detection in the images, different methods were implemented, and the parameters such as structuring element of morphological operations were separately optimized. Also, for low-quality images, preprocessing steps were implemented. To increase accuracy for all classes, and to make system fully automated and robust deep learning methods were implemented.

The developed system uses hyper-parameter optimized convolutional neural network architecture U-Net + BN to segment objects in the images and U-Net performance is compared with four different semantic segmentation architectures. Comparison results are given in Table 4. All five architectures result in similar outcomes, but U-Net shows the best performance. Besides, unbalanced dataset dramatically affects performance of the network. In our case, to overcome unbalanced dataset problem, firstly, we trained the network on balanced dataset and transferred initial weights for each class to unbalanced dataset training part. We tried data augmentation technique to increase dataset and so to increase the performance of the network, but it didn't increase the performance of the network, so we didn't use it in the final tests. Our system is fully automated, after the optimization, there was no need to arrange parameters, and the proposed system resulted in segmentation of all objects in one image. Segmentation accuracy for each class was more than 85% and IoU for each class was more than 65%. Moreover, after the segmentation, cells were quantified and results showed that, all evaluation matrices in [9] have been improved with the proposed semantic segmentation methods, and the improvements are given in Table 6. The statistical analysis revealed that deep learning-based cell segmentation and counting can be used as a readout method for the MRD Biochip. The current results are promising and for the future work, we are planning to perform more clinical trials and increase the sample size for testing the whole system. We anticipate that the developments in the microfluidic system and the iterative optimizations would further improve the performance of the deep learning algorithm. We observed that cells from patient samples have more variations in terms of size and appearance under microscope compared to the cells from the cell cultures. In order to improve the reliability of the system, more patient samples should be tested. In addition, we realized that, when we divide the high-resolution images to smaller sizes, some cluster parts are segmented as beads, and this causes less true positive for clusters, and more false positives for beads, and so less precision value for clusters, beads, and overall performance. In order to increase precision, model structure can be changed to analyze high-resolution images by using image pyramid techniques as a future work. Moreover, before segmentation algorithm, some pre-processing methods can be applied to noisy images to increase the precision value.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Acknowledgment*

Authors acknowledge TÜBİTAK (Project No: 115E020) for the financial support and Dr. Ekrem Ünal for collecting the patient samples. All authors declare that there is no conflict of interest.

*Author contributions*

K. I. conceptualized the idea, F.C. implemented the algorithms, all authors designed the algorithms, interpreted the results, and wrote the manuscript.

*Patient Samples*

All patient samples in this study were obtained under the approval of the Clinical Research Ethics Committee of the Erciyes University Faculty of Medicine (Approval date: 09/01/2015, Decision no:2015/21, Kayseri, Turkey). Written informed consent was obtained from all parents/legally authorized representatives of patients. The Declaration of Helsinki was followed throughout the study.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bspc.2022.103783.

## References

[1] G.J. Schuurhuis, M. Heuser, S. Freeman, M.-C. Béné, F. Buccisano, J. Cloos, D. Grimwade, T. Haferlach, R.K. Hills, C.S. Hourigan, J.L. Jorgensen, W. Kern, F. Lacombe, L. Maurillo, C. Preudhomme, B.A. van der Reijden, C. Thiede, A. Venditti, P. Vyas, B.L. Wood, R.B. Walter, K. Döhner, G.J. Roboz, G. J. Ossenkoppele, Minimal/measurable residual disease in AML: a consensus

document from the European LeukemiaNet MRD Working Party, Blood 131 (12) (2018) 1275–1291.
[2] N.J. Short, E. Jabbour, M. Albitar, M. de Lima, L. Gore, J. Jorgensen, A.C. Logan, J. Park, F. Ravandi, B. Shah, J. Radich, H. Kantarjian, Recommendations for the assessment and management of measurable residual disease in adults with acute lymphoblastic leukemia: A consensus of North American experts, Am. J. Hematol. 94 (2) (2019) 257–265.
[3] Z. Wang, M. Guo, Y. Zhang, S. Xu, H. Cheng, J. Wu, W. Zhang, X. Hu, J. Yang, J. Wang, G. Tang, The applicability of multiparameter flow cytometry for the detection of minimal residual disease using different-from-normal panels to predict relapse in patients with acute myeloid leukemia after allogeneic transplantation, Int. J. Lab. Hematol. 41 (5) (2019) 607–614.
[4] I. Della Starza, V. Nunes, F. Lovisa, D. Silvestri, M. Cavalli, A. Garofalo, M. Campeggio, L.A. De Novi, R. Soscia, C. Oggioni, L. Mussolin, A. Biondi, A. Guarini, M.G. Valsecchi, V. Conter, A. Biffi, G. Basso, R. Foà, G. Cazzaniga, Droplet digital PCR improves IG-/TR-based MRD risk definition in childhood B-cell precursor acute lymphoblastic leukemia, HemaSphere 5 (3) (2021) e543.
[5] R. Sánchez, R. Ayala, J. Martínez-López, Minimal residual disease monitoring with next- generation sequencing methodologies in hematological malignancies, Int. J. Mol. Sci. 20 (11) (2019) 2832.
[6] M. Radovich *et al.*, "Association of Circulating Tumor DNA and Circulating Tumor Cells after Neoadjuvant Chemotherapy with Disease Recurrence in Patients with Triple-Negative Breast Cancer: Preplanned Secondary Analysis of the BRE12-158 Randomized Clinical Trial," 2020, doi: 10.1001/jamaoncol.2020.2295.
[7] K. İçöz, Ü. Akar, E. Ünal, Microfluidic Chip based direct triple antibody immunoassay for monitoring patient comparative response to leukemia treatment, Biomed. Microdevices 22 (3) (2020) 48, https://doi.org/10.1007/s10544-020-00503-6.
[8] J.M. Jackson, J.B. Taylor, M.A. Witek, S.A. Hunsucker, J.P. Waugh, Y. Fedoriw, T. C. Shea, S.A. Soper, P.M. Armistead, Microfluidics for the detection of minimal residual disease in acute myeloid leukemia patients using circulating leukemic cells selected from blood, Analyst 141 (2) (2016) 640–651.
[9] F. Uslu, K. Icoz, K. Tasdemir, R.S. Doğan, B. Yilmaz, Image-analysis based readout method for biochip: Automated quantification of immunomagnetic beads, micropads and patient leukemia cell, Micron 133 (2020) 102863.
[10] S. Faley, K. Seale, J. Hughey, D.K. Schaffer, S. VanCompernolle, B. McKinney, F. Baudenbacher, D. Unutmaz, J.P. Wikswo, Microfluidic platform for real-time signaling analysis of multiple single T cells in parallel, Lab Chip 8 (10) (2008) 1700.
[11] F. Uslu, K. Icoz, K. Tasdemir, B. Yilmaz, Automated quantification of immunomagnetic beads and leukemia cells from optical microscope images, Biomed. Signal Process. Control 49 (2019) 473–482.
[12] F. Meyer, Topographic distance and watershed lines, Signal Process. 38 (1) (1994) 113–125, https://doi.org/10.1016/0165-1684(94)90060-4.
[13] M.I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: Overview, challenges and the future, Lect. Notes Comput. Vis. Biomech. 26 (2018) 323–350, https://doi.org/10.1007/978-3-319-65981-7_12.
[14] J.C. Caicedo, J. Roth, A. Goodman, T. Becker, K.W. Karhohs, M. Broisin, C. Molnar, C. McQuin, S. Singh, F.J. Theis, A.E. Carpenter, Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images, Cytom. Part A 95 (9) (2019) 952–965.
[15] O.Z. Kraus, B.T. Grys, J. Ba, Y. Chong, B.J. Frey, C. Boone, B.J. Andrews, Automated analysis of high-content microscopy data with deep learning, Mol. Syst. Biol. 13 (4) (2017) 924.
[16] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, D. Van Valen, Deep learning for cellular image analysis, Nat. Methods 16 (12) (2019) 1233–1246, https://doi.org/10.1038/s41592-019-0403-1.
[17] E. Gómez-de-Mariscal, M. Maška, A. Kotrbová, V. Pospíchalová, P. Matula, A. Muñoz-Barrutia, Deep-Learning-Based Segmentation of Small Extracellular Vesicles in Transmission Electron Microscopy Images, Sci. Rep. 9 (1) (2019) 1–10, https://doi.org/10.1038/s41598-019-49431-3.
[18] Y. Al-Kofahi, A. Zaltsman, R. Graves, W. Marshall, M. Rusu, A deep learning-based algorithm for 2-D cell segmentation in microscopy images, BMC Bioinf. 19 (1) (2018) 1–11, https://doi.org/10.1186/s12859-018-2375-z.
[19] S.K. Sadanandan, P. Ranefall, S. Le Guyader, C. Wählby, Automated training of deep convolutional neural networks for cell segmentation, Sci. Rep. 7 (1) (2017) 1–7, https://doi.org/10.1038/s41598-017-07599-6.
[20] W. Wang, D.A. Taft, Y.-J. Chen, J. Zhang, C.T. Wallace, M. Xu, S.C. Watkins, J. Xing, Learn to segment single cells with deep distance estimator and deep cell detector, Comput. Biol. Med. 108 (2019) 133–141.
[21] M. Shahzad, A.I. Umar, M.A. Khan, S.H. Shirazi, Z. Khan, W. Yousaf, Robust Method for Semantic Segmentation of Whole-Slide Blood Cell Microscopic Images, Comput. Math. Methods Med. 2020 (2020) 1–13.
[22] N. Meng, E.Y. Lam, K.K. Tsia, H.K.H. So, Large-Scale Multi-Class Image-Based Cell Classification with Deep Learning, IEEE J. Biomed. Heal. Informatics 23 (5) (2019) 2091–2098, https://doi.org/10.1109/JBHI.2018.2878878.
[23] B. N. Narayanan, R. A. Ali, and R. C. Hardie, "Performance analysis of machine learning and deep learning architectures for malaria detection on cell images," no. September 2019, p. 29, 2019, doi: 10.1117/12.2524681.
[24] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T.L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, O. Ronneberger, U-Net: deep learning for cell counting, detection, and morphometry, Nat. Methods 16 (1) (2019) 67–70.
[25] A. Bozorgpour, R. Azad, E. Showkatian, and A. Sulaiman, "Multi-scale Regional Attention Deeplab3+: Multiple Myeloma Plasma Cells Segmentation in Microscopic Images," pp. 1–10, 2021, [Online]. Available: http://arxiv.org/abs/2105.06238.
[26] M. Loey, M. Naman, and H. Zayed, "Deep transfer learning in diagnosing leukemia in blood cells," *Computers*, vol. 9, no. 2, 2020, doi: 10.3390/computers9020029.
[27] K.K. Anilkumar, V.J. Manoj, T.M. Sagi, Automated Detection of B Cell and T Cell Acute Lymphoblastic Leukaemia Using Deep Learning, Irbm 1 (2021) 1–9, https://doi.org/10.1016/j.irbm.2021.05.005.
[28] M. Doan, M. Case, D. Masic, H. Hennig, C. McQuin, J. Caicedo, S. Singh, A. Goodman, O. Wolkenhauer, H.D. Summers, D. Jamieson, F.W. Delft, A. Filby, A. E. Carpenter, P. Rees, J. Irving, Label-free leukemia monitoring by computer vision, Cytom. Part A 97 (4) (2020) 407–414.
[29] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.
[30] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," 2017, [Online]. Available: http://arxiv.org/abs/1706.05587.
[31] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015, doi: 10.1109/CVPR.2015.7298965.
[32] O. Mzava, Z. Tas, K. İçöz, Magnetic micro/nanoparticle flocculation-based signal amplification for biosensing, Int. J. Nanomedicine 11 (2016) 2619–2631, https://doi.org/10.2147/IJN.S108692.
[33] S. Ioffe, C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, vol. 37, pp. 448–456, [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html.
[34] C. Chen, W. Bai, R.H. Davies, A.N. Bhuva, C.H. Manisty, J.B. Augusto, J.C. Moon, N. Aung, A.M. Lee, M.M. Sanghvi, K. Fung, J.M. Paiva, S.E. Petersen, E. Lukaschuk, S.K. Piechnik, S. Neubauer, D. Rueckert, Improving the Generalizability of Convolutional Neural Network-Based Segmentation on CMR Images, Front. Cardiovasc. Med. 7 (2020), https://doi.org/10.3389/fcvm.2020.00105.
[35] V. Nair, G. Hinton, Rectified linear units improve restricted boltzmann machines vinod nair, Proceedings of ICML 27 (2010) 807–814.
[36] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, G. Bing, Learning from class-imbalanced data: Review of methods and applications, Expert Syst. Appl. 73 (2017) 220–239, https://doi.org/10.1016/j.eswa.2016.12.035.
[37] N. Praljak, S. Iram, U. Goreke, G. Singh, A. Hill, U.A. Gurkan, M. Hinczewski, F. M. Fadlelmola, Integrating deep learning with microfluidics for biophysical classification of sickle red blood cells adhered to laminin, PLoS Comput. Biol. 17 (11) (2021) e1008946.
[38] F. Niemeyer, F. Galbusera, Y. Tao, A. Kienle, M. Beer, H.J. Wilke, A Deep Learning Model for the Accurate and Reliable Classification of Disc Degeneration Based on MRI Data, Invest. Radiol. 56 (2) (2021) 78–85, https://doi.org/10.1097/RLI.0000000000000709.
[39] A. Rácz, D. Bajusz, K. Héberger, Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification, Molecules 26 (4) (2021) 1–16, https://doi.org/10.3390/molecules26041111.
[40] L. Liu, F.X. Wu, Y.P. Wang, J. Wang, Multi-receptive-field CNN for semantic segmentation of medical images, IEEE J. Biomed. Heal. Informatics 24 (11) (2020) 3215–3225, https://doi.org/10.1109/JBHI.2020.3016306.
[41] J. Snoek, H. Larochelle, R.P. Adams, Practical Bayesian optimization of machine learning algorithms, Adv. Neural Inf. Process. Syst. 4 (2012) 2951–2959.
[42] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. (2015) 1–15.
[43] M.A. Rahman, Y. Wang, Optimizing intersection-over-union in deep neural networks for image segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) vol. 10072 LNCS (2016) 234–244, https://doi.org/10.1007/978-3-319-50835-1_22.
[44] G. Csurka, D. Larlus, F. Perronnin, "What is a good evaluation measure for semantic segmentation?", *BMVC 2013 - Electron*, Proc. Br. Mach. Vis. Conf. 2013 (2013), https://doi.org/10.5244/C.27.32.
[45] T. Tran, O.H. Kwon, K.R. Kwon, S.H. Lee, K.W. Kang, "Blood Cell Images Segmentation using Deep Learning Semantic Segmentation", *2018 IEEE Int*, Conf. Electron. Commun. Eng. ICECE 2018 (2019) 13–16, https://doi.org/10.1109/ICECOME.2018.8644754.
[46] E. Kasuya, Mann-Whitney U test when variances are unequal, Anim. Behav. 61 (6) (2001) 1247–1249, https://doi.org/10.1006/anbe.2001.1691.
[47] R. M.Roy and P. M. Ameer, "Segmentation of leukocyte by semantic segmentation model: A deep learning approach," *Biomed. Signal Process. Control*, vol. 65, no. December 2020, p. 102385, 2021, doi: 10.1016/j.bspc.2020.102385.