Computers in Biology and Medicine 150 (2022) 106179

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Semantic segmentation of low magnification effusion cytology images: A semi-supervised approach

Shajahan Aboobacker [a,*,1], Deepu Vijayasenan [a,2], Sumam David S. [a,3], Pooja K. Suresh [b,4], Saraswathy Sreeram [b,2]

[a] Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, 575025, Karnataka, India
[b] Department of Pathology, Kasturba Medical College Mangalore, Manipal Academy of Higher Education, Manipal, 575001, Karnataka, India

ARTICLE INFO

Keywords:
Cytology
Deep learning
Digital pathology
Semantic segmentation
Semi-supervised learning

ABSTRACT

Cytopathologists examine microscopic images obtained at various magnifications to identify malignancy in effusions. They locate the malignant cell clusters at a low magnification and then zoom in to investigate cell-level features at a high magnification. This study predicts the malignancy at low magnification levels such as 4X and 10X in effusion cytology images to reduce scanning time. However, the most challenging problem is annotating the low magnification images, particularly the 4X images. This paper extends two semi-supervised learning (SSL) models, MixMatch and FixMatch, for semantic segmentation. The original FixMatch and MixMatch algorithms are designed for classification tasks. While performing image augmentation, the generated pseudo labels are spatially altered. We introduce reverse augmentation to compensate for the effect of the spatial alterations. The extended models are trained using labelled 10X and unlabelled 4X images. The average F-score of benign and malignant pixels on the predictions of 4X images is improved approximately by 9% for both Extended MixMatch and Extended FixMatch respectively compared with the baseline model. In the Extended MixMatch, 62% sub-regions of low magnification images are eliminated from scanning at a higher magnification, thereby saving scanning time.

1. Introduction

Effusion is defined as the abnormal accumulation of fluid in the body cavities such as peritoneal, pleural and pericardial. Effusion might be due to inflammations, pulmonary infections or malignancy. The collected fluid samples are spread on a microscopic slide and stained with Haematoxylin and Eosin stain (H&E stain) or Papanicolaou stain (PAP stain). The prepared slide is examined through a microscope by cytopathologists to detect malignancies. Microscopic examination of the whole slide is tedious, time-consuming and subject to inter-observer variations. This motivates the researchers to focus on automating the detection of malignancy. The time it takes for cytopathologists to screen the entire slide for malignancies is reduced by automation. Typically automation algorithms use only high magnification images [1–7]. On the other hand, cytopathologists take multiple magnifications of cytology images into account when predicting malignancy. In comparison to higher magnification images, lower magnification images cover a larger area in a single frame [8–10]. Cytopathologists search for the malignant regions of interest (ROI) in the lower magnification. They look at the textural and morphological behaviour of the cell clusters [11–14]. The ROI are then zoomed in for a closer examination at cell level [9,10].

Thus the low magnification images can aid in identifying the malignant area. Memory requirements and scanning time can be reduced by scanning only the ROI at a higher magnification. However, identifying malignancy in low magnification ($\lesssim$ 10X) images is challenging due to the blurring of features such as texture and nuclei. This also makes labelling difficult at low magnification levels.

This work aims to use the lower magnification images (4X and 10X) to predict the malignant region compared to the conventional high magnification images (40X). We were able to use a conventional semantic segmentation model for training 10X images. But training 4X images on a semantic model is difficult because labelling itself is challenging at this level.

* Corresponding author.
E-mail address: shajahan.187ec012@nitk.edu.in (S. Aboobacker).
[1] Research Scholar.
[2] Associate Professor.
[3] Professor.
[4] Additional Professor.

https://doi.org/10.1016/j.compbiomed.2022.106179
Received 23 March 2022; Received in revised form 9 September 2022; Accepted 1 October 2022
Available online 4 October 2022

A few studies have used images of multiple magnifications to predict malignancy. One of the objectives of employing multiple magnifications is to improve accuracy. In this approach, the images at multiple magnifications are fed as input to single [15,16] or multiple networks [17–20] of the same or similar architecture such as UNet,VGG or Resnet. These schemes require labels corresponding to each magnification. Each neural network is trained in a supervised way with corresponding labels. Individual outputs are combined either through weighted averaging [19], decision tree [17] or by integrating them directly into the neural network [18,20]. This approach is ineffective in identifying malignant regions from low magnification images.

Another approach that makes use of low magnification images is to identify the ROI [21]. The low and high magnification images used in [21] are obtained from publicly available whole slide image (WSI) datasets. These datasets are already labelled for all the magnifications. However, labelling is challenging at the lowest magnification for a new dataset. Therefore an alternate method is required which does not need labels for the lowest magnification. We propose two semi-supervised semantic segmentation models to adapt the 10X model to 4X data.

The main contributions are: (1) The paper proposes semi-supervised learning models to automatically segment cytology slides at lower magnification levels that are difficult to label or classify. Therefore the approach is promising as current methods mainly focus on higher magnification images. (2) It extends two hybrid semi-supervised classification models to perform semantic segmentation tasks. (3) In this approach, the non-malignant regions identified from low magnification images could be eliminated from high magnification scan, thereby reducing scanning and processing time. (4) It attempts to reduce the labelling load of expert pathologists to label presumably massive datasets required to generate more accurate models.

The remainder of the paper is structured as follows: Section 2 focuses on related works. Section 3 outlines the proposed work. Section 4 explains the experimental setup and results. Section 5 summarizes the work. Section 6 concludes the paper.

2. Related works

The objective of this work is to do semantic segmentation of unlabelled low magnification effusion cytology images. This problem is typically modelled in the literature using weakly supervised or semi-supervised learning algorithms. Literature address this kind of problem either using weakly supervised deep learning models or semi-supervised learning algorithms. The weakly supervised models require prior knowledge of image-level labels and additional information. Girum et al. use a weakly supervised deep learning algorithm for segmentation [22]. The input to the model is the raw image and pseudo-landmarks. The pseudo-landmark consists of four pairs of boundary coordinates and a weak label of whether the object is present or not. This work predicts the exact boundary of the object, but the information is insufficient to segment when different objects are present in a single image. Another class of algorithms uses the weakly supervised method for classification tasks using deep neural networks [23–25]. They are using an image-level label as a weak label. These methods work well when there is a large dataset for training.

Semi-supervised learning (SSL) addresses the classification problem when labelling the entire dataset is either too expensive or not practical. SSL makes use of unlabelled data to supplement labelled data [26]. The majority of SSL algorithms are based on a few essential assumptions, namely manifold, smoothness, and low-density. The manifold assumption asserts that the input space comprises numerous lower-dimensional manifolds on which all samples are situated. The labels of samples on the same manifold are identical. The smoothness assumption implies the same labels for any two adjacent input points. According to the low-density assumption, the decision boundary should not cross through a dense region.

SSL models based on deep neural networks (DNN) employ semi-supervised learning by including a loss term for unlabelled data. The approach is based on entropy minimization [27–29], consistency regularization [30–32] or their combination [33,34]. Entropy minimization ensures low entropy on the classifier's predictions, thereby fulfilling the low-density assumption. Consistency regularization implies the label is unaffected even if the input data is perturbed with noise because the data will still be on the same manifold.

Self-Training methods such as Pseudo-label [28], Noisy Student [29] and Adversarial Training [27] are based on entropy minimization. In the Pseudo-label approach, the goal is to train a model on a batch of labelled and unlabelled images simultaneously. A batch of unlabelled images is predicted on the same model, and the maximum confidence class is selected as pseudo labels. Then the model is trained using a cross-entropy loss on both labelled and unlabelled images. In the Noisy Student approach, there are two models named Teacher and Student. The Teacher model is trained using the labelled images. This trained Teacher model is used to generate pseudo-labels for the unlabelled images. Later, the Student model is trained simultaneously on labelled and unlabelled images. In the Adversarial Training, a generator–discriminator pair is used. The generator will try to predict the output as close to the label so that the discriminator cannot distinguish whether it is a label or a prediction. It trains using a min–max approach for the generator and discriminator. Three loss terms are used for training: generator loss, adversarial loss and discriminator loss.

The consistency regularization approach implies the model predictions should remain consistent even when the data or model is perturbed with noise. Several strategies that make use of consistency regularization are discussed below. The Pi-Model [30] generates two random augmentations of labelled and unlabelled images. These augmented images are fed into a dropout model. Consistency loss is defined as the squared difference between the two predictions. Additionally, a cross-entropy loss is computed for the labelled images. Total loss is calculated as the weighted sum of consistency and cross-entropy losses.

Temporal Ensembling [30] is a variant of the Pi-Model in which the label is an exponential moving average of all previous predictions instead of a single prediction. The consistency loss is calculated as the squared difference between the current prediction and the label.

Mean Teacher [32] comprises two models: "Student" and "Teacher". The student model is a variation of the standard model with dropout. The teacher model is architecturally identical to the student model. Its weights are determined using an exponential moving average of the student model's weights. Two random augmented versions of an image are made for each labelled or unlabelled image. One image is fed into the Student model and another into the Teacher model. The square difference between student and teacher models predictions is employed as a consistency loss. A cross-entropy loss is also calculated for labelled images. The total loss is the weighted sum of these two loss factors.

The central concept in Virtual Adversarial Training [31] is to create an adversarial transformation of an image that modifies the model's prediction. An adversarial variant for every image is created. The first view is the labelled or unlabelled image, while the second view is its adversarial counterpart. Then, the label distributions for both images are predicted using the same model. As a consistency loss, the KL-divergence of these two predictions is employed. Cross-entropy loss is also calculated for labelled images. The total loss is the weighted sum of these two loss factors as in the all above cases.

Hybrid methods such as MixMatch [33] and FixMatch [34] uses the concept of both entropy minimization and consistency regularization. Both of it uses augmentations for consistency regularization and pseudo labelling for entropy minimization.

The algorithms of MixMatch and FixMatch are described in Section 2.1 and 2.2 respectively.

2.1. MixMatch

MixMatch algorithm uses augmentations for consistency regularization and pseudo labelling for entropy minimization. The labelled image is augmented once, and the unlabelled image is augmented K times. These K augmented images are fed to the model, and its predictions are averaged. This averaged image is temperature scaled using Eq. (1) to obtain the pseudo label for all K images.

$$\tilde{y}_{u_i}^{PL} = \tilde{y}_{u_i}^{\frac{1}{T}} \Big/ \sum_{j=1}^{C} \tilde{y}_{u_j}^{\frac{1}{T}} \tag{1}$$

where $\tilde{y}_u^{PL}$ is the pseudo label, $\tilde{y}_u$ is the class probability predictions, C is the total number of classes and T is called as temperature scaling factor. Temperature scaling is a method used to reduce the entropy of the label distribution. Lowering the temperature parameter produces lower entropy predictions. As T tends to zero, the output will approach a one hot distribution.

Additionally, a mix-up is done on both labelled and unlabelled input images and their associated labels. Then the predictions of labelled images are used to calculate the cross-entropy loss with the associated labels. Similarly, the predictions of the unlabelled images are used to calculate the mean squared error with the associated pseudo labels. The final loss is calculated as the weighted sum of labelled and unlabelled losses.

2.2. FixMatch

The FixMatch algorithm uses weak and strong augmentation for SSL. A weak augmentation performs a random augmentation on each image from the vertical flip, horizontal flip and a height and width wise shifting. A strong augmentation performs the following additional augmentations — zoom and crop, contrast and colour variations. The weakly augmented image is fed into the model to get the class probability predictions. The class with highest predicted probability value that exceeds a threshold value is considered as the pseudo label. Then the strongly augmented images are fed into the model to predict classes. These predictions are compared with the pseudo label using cross-entropy loss. The model predictions for the labelled image is compared with the labels also using cross-entropy loss. The overall loss is the weighted sum of these two losses.

FixMatch and MixMatch provide state-of-the-art performance for the classification task. This work extends MixMatch and FixMatch for semantic segmentation tasks.

3. Proposed work

In this work, we perform a semantic segmentation task using semi-supervised learning. A classification algorithm labels an entire image with one of the available classes. But, semantic segmentation is the process of assigning distinct object classes to each pixel in an image. When semi-supervised learning is utilized for a semantic segmentation task, the augmentation of the input image may lead to spatial alteration of the pixel-wise labels. Hence, the averaging of spatially altered labels in MixMatch and comparing spatially altered labels of weak and strong augmented images in FixMatch results in incorrect results. To address this problem, we reverse augment the predicted label. This guarantees the spatial alignment of all predicted labels. The extension of MixMatch and FixMatch utilizing reverse augmentation is described in the following subsections.

3.1. Extended MixMatch

The algorithm generates pseudo labels for unlabelled images, and the SSL model is trained with labelled and unlabelled images using the loss function that is the weighted sum of supervised and unsupervised loss functions. The pseudo label generation and the training process is discussed below.

3.1.1. Pseudo label generation

Each unlabelled input image is augmented K times randomly. These augmentations are chosen at random from a set of available options, including horizontal flip, vertical flip, height–width shift and zoom & crop. Each of these augmented images is fed into the baseline model, which predicts its associated labels. These labels are spatially changed with the input image, as seen in Fig. 1. Since they lack spatial correspondence, the label images cannot be averaged. As a result, before averaging the predicted outputs, we must restore the augmented images to their original size and position. We derive the parameters needed for reverse augmentations from those used for individual random augmentations. For example, the reverse operation of a zoom-in is a zoom-out, and the reverse operation of a right shift is a left shift. Then the reverse augmented predicted outputs are averaged to produce a single label image. A temperature sharpening is applied on the label image to lower the entropy of label distribution as given in Eq. (1). The sharpened image is considered as the pseudo label. The input images are padded with sufficient zeros to avoid boundary pixels being lost. Only the centre portions with the original size are used to calculate the loss to tackle this issue.

3.1.2. Model training

The model is trained with a combined loss ($\mathcal{L}$) function, which is the weighted sum of supervised and unsupervised loss functions as defined in Eq. (2),

$$\mathcal{L} = \mathcal{L}_s + \lambda_u \mathcal{L}_u \tag{2}$$

Consider a set of labelled images (x_l) with labels (y_l) and a set of unlabelled images (x_u). Each unlabelled image in the batch is randomly augmented K times ($x_{u,k}$), as described in the previous section. Let the prediction of the model on labelled image x_l be $\hat{y}_l$ and on unlabelled image $x_{u,k}$ be $\hat{y}_{u,k}$. The supervised loss ($\mathcal{L}_s$), is a pixel-wise cross-entropy loss between the predictions of labelled data ($\hat{y}_l$) and their labels (y_l), is given by

$$\mathcal{L}_s = \frac{1}{B} \sum_{p \in B} H(y_l, P_{model}(\hat{y}_l | \hat{x}_l; \theta)) \tag{3}$$

Pseudo labels ($\tilde{y}_u$) are generated for batches of K augmented images, as explained in the last section. The unsupervised loss ($\mathcal{L}_u$), is the pixel-wise mean square loss between the reverse augmentation of predictions of K augmented images and the pseudo labels, is given by

$$\mathcal{L}_u = \frac{1}{KB} \sum_{p \in KB} \| \tilde{y}_u - P_{model}(\hat{y}_{u,k} | x_{u,k}; \theta) \|_2^2 \tag{4}$$

where λ_u is the hyperparameter that gives weightage for the unsupervised loss component in the total loss. K is another hyperparameter that tells the no. of augmentations in the unlabelled batch, B represents the batch size, and p represents each pixel in each image in the batch. $H(y_l, \hat{y}_l)$ denotes the cross-entropy between y_l and $\hat{y}_l$ distributions and θ is the set of model parameters. The block diagram of Extended MixMatch is shown in Fig. 2. The output from the model is probability values for every pixels. It is shown as an image in Fig. 2 for illustration purposes by giving respective colours for class with the highest probability value in each pixel.

3.2. Extended FixMatch

Extended FixMatch also follows a similar approach to the Extended MixMatch method. Initially, pseudo labels are generated for unlabelled images, and then the model is trained with both labelled and unlabelled images. FixMatch uses two types of augmentations — weak and strong. Weak augmentations are used for pseudo label generation, and strong augmentations are used in the training process. Pseudo label generation and model training is explained in the following sections.

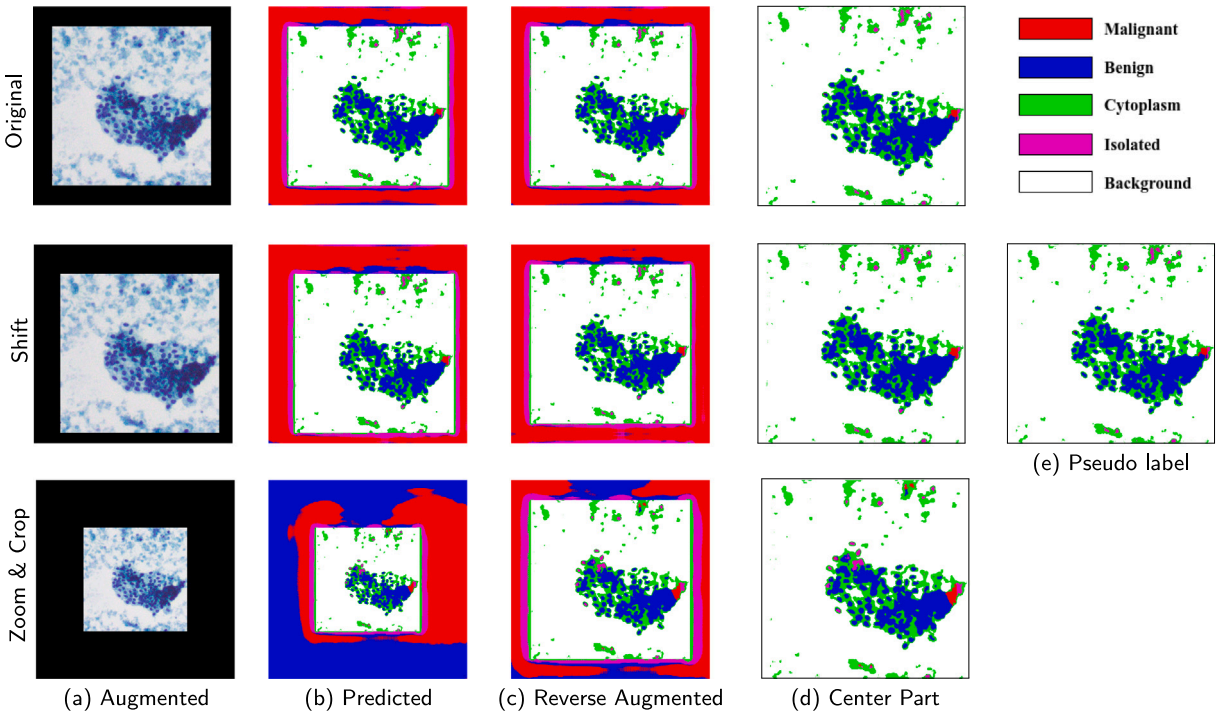

Fig. 1. Different types of augmentation and corresponding reverse augmentation for a sample image.

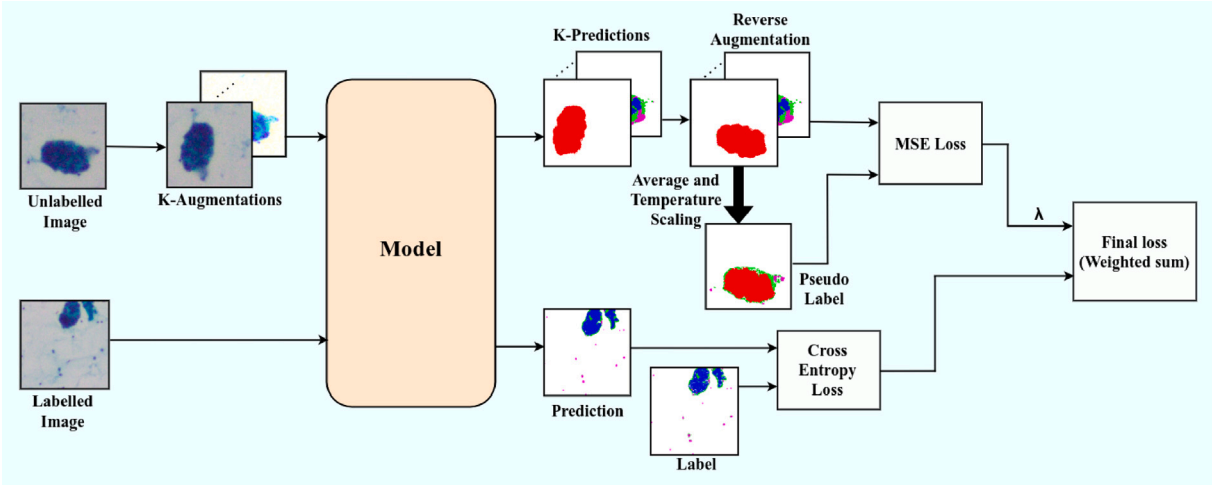

Fig. 2. Block diagram of Extended MixMatch.

3.2.1. Pseudo label generation

Each unlabelled input image is subjected to a weak augmentation. Horizontal flip, vertical flip, and height–width shift are picked at random as weak augmentation. This weak augmented image is fed into the baseline model to get the predictions. The predicted output is reverse augmented, similar to MixMatch, to maintain the spatial alignment while calculating the loss. Only the labels with maximum class probabilities above a predefined threshold are considered while calculating the loss. The remaining labels are masked, as shown in the pseudo-label block of Fig. 3. The maximum class probabilities above the threshold are assigned the value one, and all other class probabilities are assigned as zero to obtain the pseudo label.

3.2.2. Model training

The model is trained with labelled and unlabelled images simultaneously. The labelled images are weakly augmented and fed into the model. Cross entropy loss for labelled image is calculated between the predictions and the labels as defined in Eq. (5). The unlabelled images are strongly augmented and used for training the model. Horizontal flip, vertical flip, height–width shift, zoom & crop, auto-contrast and colour are picked at random as strong augmentation. Pseudo labels ($\tilde{y}_u$) for the unlabelled image is generated as explained in the previous section. Only the class with highest predicted probability value that exceeds a threshold value are used for calculating the loss. Cross entropy loss for unlabelled image is calculated between the predictions ($\hat{y}_u$) and the pseudo labels ($\tilde{y}_u$) as defined in Eq. (7). The weighted sum of these two losses is used as the total loss as defined in Eq. (8).

$$\mathcal{L}_s = \frac{1}{B}\sum_{p\in B} H(y_l, P_{model}(\hat{y}_l|W(x_l);\theta)) \tag{5}$$

$$\tilde{y}_u = P_{model}(\tilde{y}_u|W(x_u);\theta) \tag{6}$$

$$\mathcal{L}_u = \frac{1}{B}\sum_{p\in B} 1(max(\tilde{y}_u) > \tau) H(\tilde{y}_u, P_{model}(\hat{y}_u|S(x_u);\theta)) \tag{7}$$

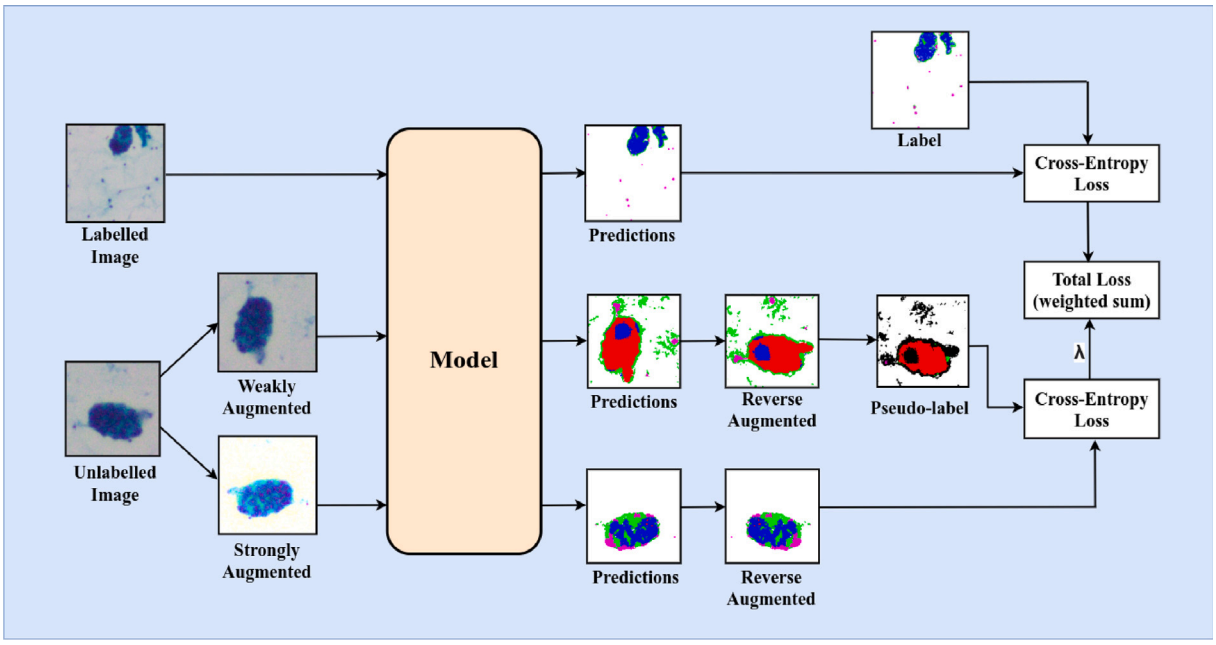

Fig. 3. Block diagram of Extended FixMatch.

Table 1
Distribution of images.

Magnification	Benign	Malignant	Total
40X	127	85	212
10X	38	45	83
4X	17	33	50
Total	182	163	345

$$\mathcal{L} = \mathcal{L}_s + \lambda_u \mathcal{L}_u \tag{8}$$

$W(.)$ and $S(.)$ are the weak and strong augmentations respectively. The hyperparameter λ_u is the weightage of unsupervised loss in the combined loss function and τ is the threshold above which a pseudo label is retained.

4. Experimental setup and results

In this section, we discuss the experiments conducted with two proposed SSL semantic segmentation models. These proposed models are built on two standard baseline models, namely MobileUNet and ResUNet++.

4.1. Dataset

The dataset is collected from Kasturba Medical College Mangalore with clearance (IEC KMC MLR 01-19/34; 16th January 2019) from the KMC Mangalore Institutional Ethics Committee. The dataset is collected from 30 patients that consist of 345 effusion cytology images at a resolution of 1920 X 1440 pixels. Three different magnifications are used to capture the data: 4X, 10X and 40X. The details of the dataset is given in Table 1. This work uses only the low magnification images (4X and 10X) as our aim is to identify the malignant clusters.

The images in the dataset has malignant, benign, cytoplasm, isolated, or background regions. Each pixel in the image is labelled as one of these five classes. The 10X and 4X datasets are annotated and validated by a cytopathologist. This work reuses the 10X model for the 4X data. Hence, the 4X data is upsampled to 10X magnification using bilinear interpolation. Henceforth, this data is named as 4X resampled. So we have two datasets, the 10X data considered as labelled data and 4X resampled data considered as unlabelled data.

All the images are cropped into sub-images of size 384 × 384. Overlapping sub-images along with augmentations are used to increase the size of the training dataset. Non overlapped sub-images are used for validation and testing. The background class pixels are relatively higher (>90%) compared to other classes. Hence, the training sub-images with less than 5% foreground pixels are discarded. A few sample images and their corresponding labels are shown in the first and second rows of Fig. 4.

4.2. Metrics

The number of background pixels is very high compared to foreground pixels. Hence, accuracy is not a good metric to evaluate the semantic segmentation performance in this work. Therefore, we used Recall (R) and Precision (P) as the metrics to assess performance. Recall and Precision are defined as follows.

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$P = \frac{TP}{TP + FP} \tag{10}$$

True Positive (TP) indicates the number of pixels that have been classified correctly, False Positive (FP) represents the number of negative pixels that have been wrongly classified as positive. False Negative (FN) indicates the number of positive pixels that have been incorrectly classified as negative.

To report a single metric, we used the F-score (F), which is the harmonic mean of Recall (R) and Precision as defined in Eq. (11).

$$F = 2 \times \frac{R \times P}{R + P} \tag{11}$$

Benign and malignant are two classes that are important for detecting malignancy. Hence, we calculate the F-scores for benign pixels (Benign F-score) and malignant pixels (Malignant F-score) separately and also their average (Average F-score).

We also calculated IoU (Intersection over Union) for benign and malignant pixels as an additional performance measure. It is the ratio of overlap between predictions and labels to the union of both, as shown in Eq. (12).

$$IoU = \frac{Predictions \cap Labels}{Predictions \cup Labels} \tag{12}$$

Table 2
Results of 5 fold cross validation of 4X resampled dataset on the baseline MobileUNet and ResUNet++ models.

Model	Malignant		Benign		Average	
	F-score	IoU	F-score	IoU	F-score	IoU
MU[a]	0.86	0.76	0.55	0.38	0.71	0.57
RU[b]	0.78	0.63	0.57	0.4	0.68	0.52

[a]MobileUNet.
[b]ResUNet++.

4.3. Baseline

Algorithms on semantic segmentation rely on two pieces of information, global information and local information. The global information conveys the content in the image, and the local information conveys the position of that content in the image. The literature discusses semantic segmentation techniques based on Convolutional Neural Networks [35–41]. Most of the networks require a large dataset to obtain a good result [42]. However, medical images are not available as large datasets. U-Net [40] is one of the semantic segmentation models that work well for such datasets. Here we consider two state-of-the-art models based on U-Net namely MobileUNet [43] and ResUNet++ [44] as our baseline models.

The MobileUNet and ResUNet++ architectures consist of 4 stages in the encoder and decoder sections and a bottleneck. In the MobileUNet, we used 64, 128, 256, and 512 filters in the first, second, third and fourth stages of the encoder section. The exact number of filters in the reverse order for the decoder section and 512 filters in the bottleneck. In the ResUNet++, we used one by fourth of the filters that we have used for the MobileUNet. MobileUNet has around twelve million and ResUNet++ has around four million trainable parameters.

The 10X images are randomly partitioned into training, validation and testing datasets with the ratio of 60:20:20 respectively. The 4X resampled dataset does not have reliable labels, so it cannot be used in supervised training. The baseline models are trained with only the 10X dataset in a supervised manner with a cross-entropy loss function. As the pixels of each class are not equal in numbers, there is a class imbalance problem. The predictions are more biased to the background as the background pixels are relatively higher than other pixels. The class imbalance issue is solved by giving class weights for pixels from each class. The weights of each class is calculated using the inverse frequency method [45] as given in Eq. (13).

$$w_c = \frac{N}{P_c} \tag{13}$$

where N is the total no. of pixels and P_c is the no. of pixels of class c in the training datasets.

We have used RMSProp as the optimizer and a learning rate of 10^{-5} for training the model. The training is continued for 200 epochs, and the best performing model for the validation data was chosen. The test set of 4X resampled dataset is evaluated on the trained baseline models and results are given in Table 2. The MobileUNet model has a better Malignant F-score, and ResUNet++ has a better Benign F-score when compared to each other. The third and fourth rows of Fig. 4 show predictions for a few samples of 4X resampled images using the baseline models MobileUNet and ResUNet++. The MobileUNet predictions are slightly biased towards malignant pixels, and ResUNet++ predictions are biased towards benign pixels.

4.4. Proposed SSL models

The Proposed SSL models Extended MixMatch and Extended FixMatch are trained separately with the 10X labelled dataset and 4X unlabelled dataset. A 5 fold cross validation is performed for the 4X resampled data. The model weights are randomly initialized and trained for 20 epochs for tuning the hyperparameters.

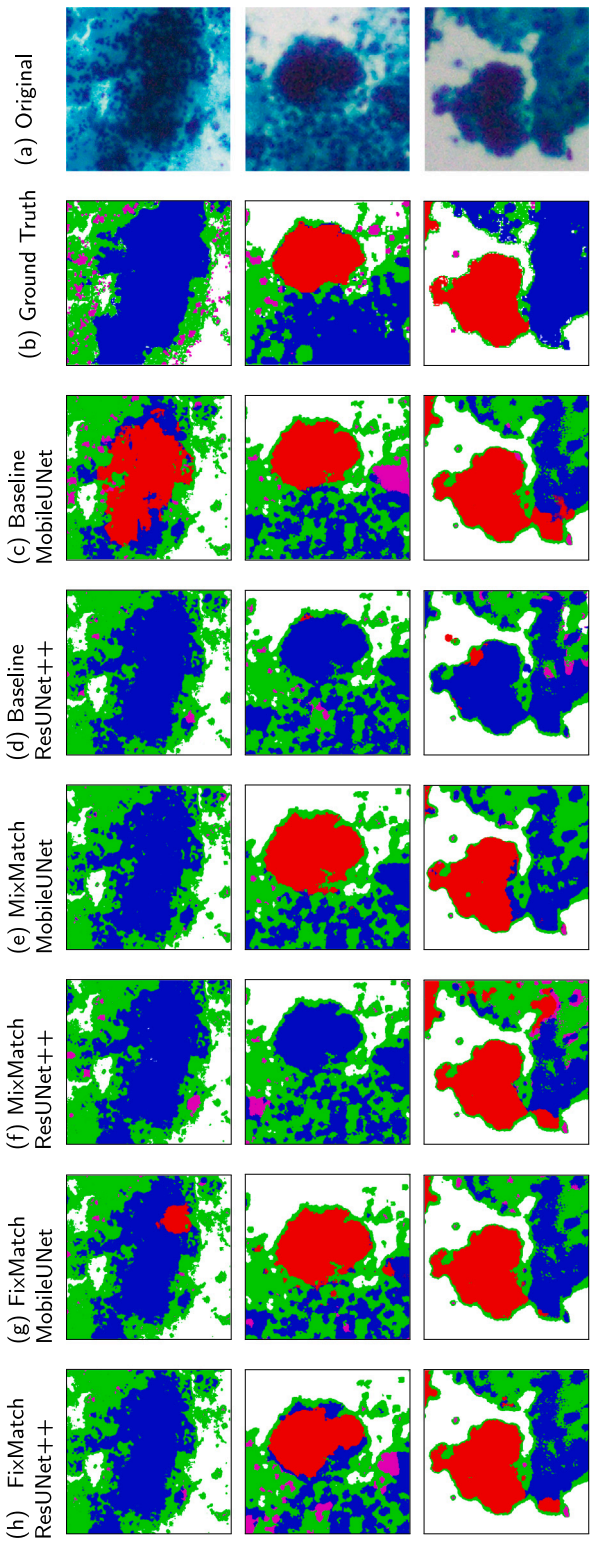

Fig. 4. Predictions of a few samples on the baseline models MobileUNet and ResUNet++ and the proposed models.

4.4.1. Extended MixMatch

The labelled image and the unlabelled images are randomly augmented using the distribution shown in Table 3. The Extended MixMatch model is trained with both the baseline models described in Section 4.3. There are two hyperparameters to be tuned, number of augmentations used for unlabelled data (K) and the weightage of

Table 3
List of augmentations with its probability of selection, used for augmenting the images in the Extended MixMatch model.

Augmentations	Range	Probability
Vertical flip	–	0.5
Horizontal flip	–	0.5
Zoom & Crop	80% to 120%	0.2
Height–width shift	0 to 5 pixels	0.5

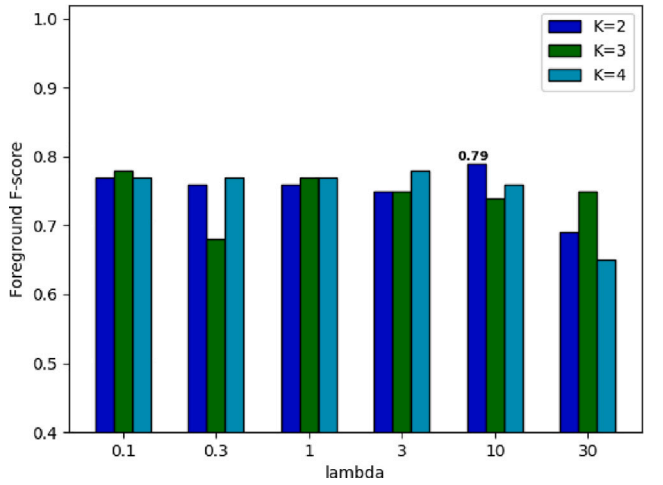

Fig. 5. Foreground F-score of 4X resampled validation data for different K and λ_u values for Extended MixMatch with MobileUNet as baseline.

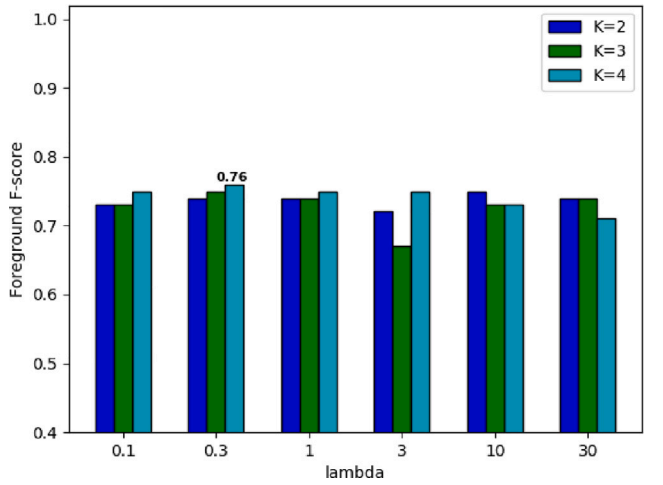

Fig. 6. Foreground F-score of 4X resampled validation data for different K and λ_u values for Extended MixMatch with ResUNet++ as baseline.

unlabelled loss (λ_u) defined in Eq. (2). The foreground F-score of validation set is used to tune the hyperparameters. A grid search is performed for K and λ_u, and the results are shown in Figs. 5 and 6 with the baseline models MobileUNet and ResUNet++, respectively. The Foreground F-score values are not varying much with different K and λ_u values. However, we chose K and λ_u that gives the best Foreground F-score value. The K is selected as 2 and λ_u as 10 for the model with MobileUNet. Similarly, K is selected as 3 and λ_u as 0.3 for the model with ResUNet++.

The model is then trained for 200 epochs with the tuned values of hyperparameters K and λ_u. A 5 fold cross validation of the 4X resampled data is performed on the model. The results are given in Table 4. In the case of the baseline models, MobileUNet has a lower Benign F-score while ResUNet++ has a lower Malignant F-score.

Table 4
Results of 5 fold cross validation of 4X resampled dataset on the Extended MixMatch models with baseline models MobileUNet and ResUNet++.

Model	Malignant		Benign		Average	
	F-score	IoU	F-score	IoU	F-score	IoU
EMM-MU[a]	0.88	0.78	0.57	0.4	0.73	0.59
EMM-RU[b]	0.89	0.8	0.58	0.41	0.74	0.61

[a]Extended MixMatch MobileUNet.

[b]Extended MixMatch ResUNet++.

Table 5
List of augmentations with its probability of selection, used for the weak and strong and augmentations in the Extended FixMatch model. Only first three rows are used in weak augmentation.

Augmentations	Range	Probability
Vertical flip	–	0.5
Horizontal flip	–	0.5
Height–width shift	0 to 5 pixels	0.5
Zoom & Crop	80% to 120%	0.2
Auto-contrast	0.05 to 0.95	0.5
colour	0 to 0.1	0.5

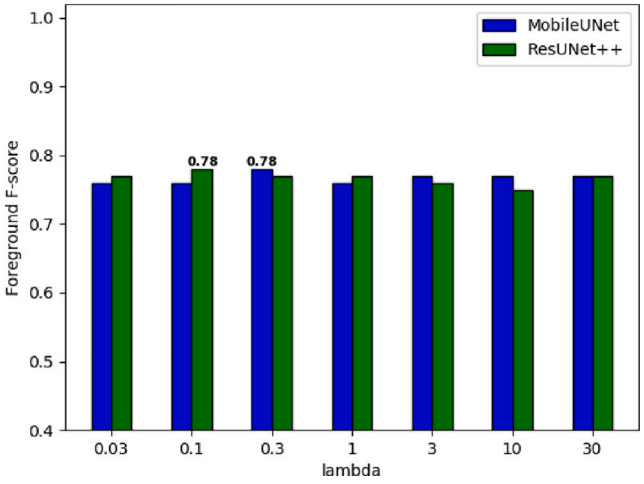

Fig. 7. Foreground F-score of 4X resampled validation data for different λ_u values for Extended FixMatch with MobileUNet and ResUNet++ as baseline models.

Extended MixMatch with MobileUNet improves the Benign F-score, whereas Extended MixMatch with ResUNet++ improves the Malignant F-score. As a result, the SSL model with MobileUNet and ResUNet++ has an improved Average F-score. The fifth and sixth rows of Fig. 4 shows the prediction results of Extended MixMatch with MobileUNet and ResUNet++ respectively. The results of Extended MixMatch with MobileUNet are closer to the ground truth, and the third sample image is improved in the case of Extended MixMatch with ResUNet++.

4.4.2. Extended FixMatch

The weak augmentation required for the generation of pseudo labels is applied randomly using the distributions listed in the first three rows of Table 5. Similarly, the strong augmentation required for training the unlabelled data is applied randomly from the distributions listed in all the rows of Table 5.

There is only one hyperparameter to be tuned in the Extended FixMatch model, which is the weightage of unlabelled loss (λ_u). The foreground F-score value for different λ_u are shown in Fig. 7 for Extended FixMatch with MobileUNet and ResUNet++. λ_u is chosen as 0.3 and 0.1 for the Extended FixMatch with MobileUNet and ResUNet++ models respectively.

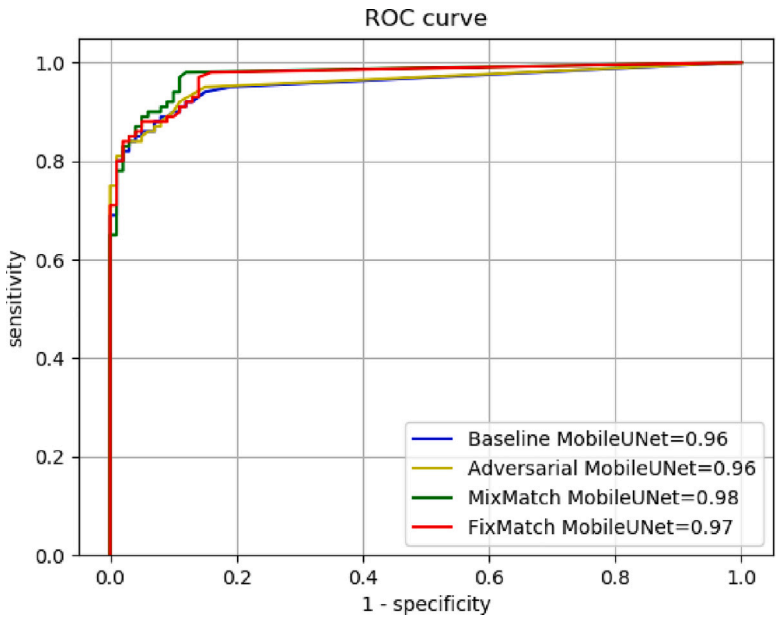

Fig. 8. Receiver operating characteristics curve for the MobileUNet baseline and proposed SSL models with their AUC.

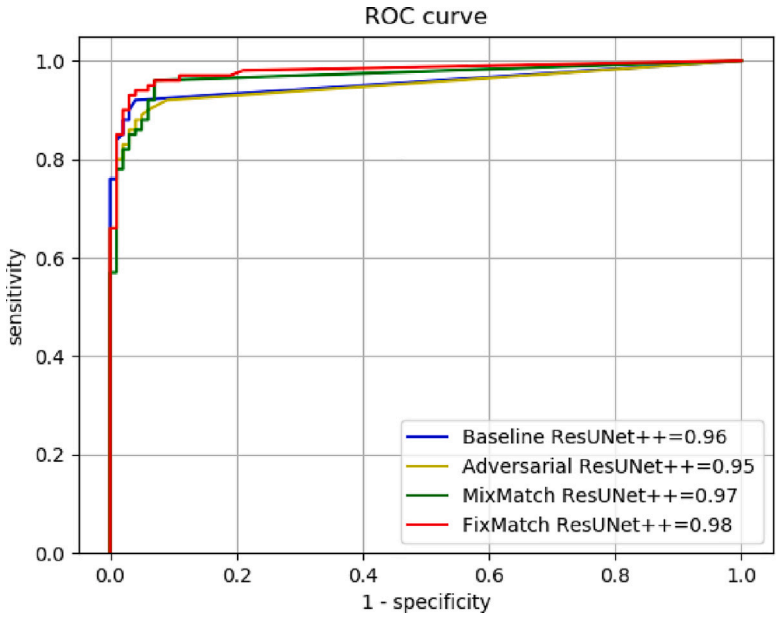

Fig. 9. Receiver operating characteristics curve for the ResUNet++ baseline and proposed SSL models with their AUC.

The model is trained for 200 epochs with the tuned λ_u. A 5 fold cross validation of 4X resampled data is performed on the model. The prediction results for the 4x resampled dataset are given in Table 6. The benign F-score is improved in the case of Extended FixMatch with MobileUNet compared to the baseline MobileUNet, and the malignant F-score is improved for the Extended FixMatch with ResUNet++ as compared with the baseline ResUNet++ model. As a result, the average F-score is improved for both cases. The seventh and eighth rows of Fig. 4 shows the prediction results of Extended FixMatch with MobileUNet and ResUNet++ respectively. The results of Extended FixMatch with MobileUNet and ResUNet++ are closer to the ground truth.

Table 6
Results of 5 fold cross validation of 4X resampled dataset on the Extended FixMatch models with baseline models MobileUNet and ResUNet++.

Model	Malignant		Benign		Average	
	F-score	IoU	F-score	IoU	F-score	IoU
EFM-MU[a]	0.88	0.79	0.59	0.43	0.74	0.61
EFM-RU[b]	0.88	0.79	0.57	0.4	0.73	0.6

[a]Extended FixMatch MobileUNet.
[b]Extended FixMatch ResUNet++.

5. Discussion

We have extended two semi-supervised models for the semantic segmentation task. Both Extended MixMatch and Extended FixMatch outperform the MobileUNet and ResUNet++ baseline models in terms of performance. Both SSL models have similar performance in terms of scores. We trained and evaluated our data on another state-of-the-art SSL model for semantic segmentation based on an adversarial network

Table 7
Results of 5 fold cross validation of 4X resampled dataset on the Adversarial Network with baseline models MobileUNet and ResUNet++.

Model	Malignant		Benign		Average	
	F-score	IoU	F-score	IoU	F-score	IoU
Ad-MU[a]	0.86	0.76	0.59	0.42	0.73	0.58
Ad-RU[b]	0.75	0.61	0.58	0.4	0.67	0.51

[a]Adversarial MobileUNet.
[b]Adversarial ResUNet++.

Table 8
Percentage of area eliminated from scanning at a higher magnification in baseline and proposed SSL methods for various values of sensitivity.

Model	Sensitivity			
	0.98	0.96	0.94	0.92
MobileUNet	0	0	60.7	63.1
ResUNet++	0	0	0	69.7
Adversarial MobileUNet	0	0	59.9	63.1
Adversarial ResUNet++	0	0	0	65.5
MixMatch MobileUNet	62.4	63.6	64.7	65.9
MixMatch ResUNet++	0	65.3	66	67.2
FixMatch MobileUNet	57.7	60.7	60.7	62.3
FixMatch ResUNet++	58.6	67.6	70.6	71.7

for a fair comparison [27]. The results are given in Table 7. The adversarial based network performs better than the baseline models but its performance is lower when compared to our proposed methods.

The objective of this work is to identify non-malignant regions in low magnification images in order to eliminate them in scanning at a higher magnification. We label each sub-image of 4X data as benign or malignant. Any image that has more than a predefined number of malignant pixels is classified as malignant. The Receiver operating characteristics (ROC) curve is plotted for the baseline and proposed SSL models in Figs. 8 and 9. The Extended FixMatch with ResUNet++ and Extended MixMatch with MobileUNet has the highest area under the curve (AUC).

We chose the operating points from the ROC curve at different sensitivity values. For each sensitivity value we determine the threshold and classify the sub-image as malignant or benign. Only the malignant sub-images will be scanned at a higher magnification. The percentage of sub-images eliminated from scanning at a higher magnification for various sensitivity values are given in Table 8. The Extended MixMatch with MobileUNet has eliminated approximately 62% of the area at the sensitivity value of 0.98. Only 38% of the low magnification area has to be scanned at a higher magnification, thereby saving scanning time.

To the best of our knowledge this is the first work which tries to identify non malignant regions from lower resolution effusion cytology images to eliminate those images in the higher resolution scan. This could save scanning and processing time of cytopathologists. This work also uses unlabelled images at low resolution to train the model which again saves lot of time and energy for annotations.

6. Conclusion

Automatic segmentation of low magnification such as 4X and 10X images helps to save scanning and processing time. However, labelling the lower magnification images is challenging. The paper extended two semi-supervised learning techniques, namely MixMatch and FixMatch, for semantic segmentation of low magnification effusion cytology images. This method allows using the 4X images without any labels along with the labelled 10X images to train the semantic segmentation model. The average F-score of benign and malignant pixels on the predictions of 4X images using the Extended FixMatch and Extended MixMatch is improved approximately by 9% compared with the predictions of 4X data on the semantic 10X model. The Extended MixMatch reduces the area to be scanned at a higher magnification by approximately 62%. Only 38% sub-regions of low magnification images have to be scanned at a higher magnification, thereby saving scanning time.

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.

Acknowledgments

The authors would like to acknowledge the Center for Cyber Physical Systems (CCPS), NITK, for providing the computational resources.

References

[1] A. Barwad, P. Dey, S. Susheilia, Artificial neural network in diagnosis of metastatic carcinoma in effusion cytology, Cytom. Part B: Clin. Cytom. 82 (2) (2012) 107–111.
[2] X. Liu, Z. Guo, J. Cao, J. Tang, MDC-net: a new convolutional neural network for nucleus segmentation in histopathology images with distance maps and contour information, Comput. Biol. Med. 135 (2021) 104543.
[3] M.R. Reena, P. Ameer, Localization and recognition of leukocytes in peripheral blood: A deep learning approach, Comput. Biol. Med. 126 (2020) 104034.
[4] R.M. Roy, P. Ameer, Segmentation of leukocyte by semantic segmentation model: A deep learning approach, Biomed. Signal Process. Control 65 (2021) 102385.
[5] A. Teramoto, A. Yamada, Y. Kiriyama, T. Tsukamoto, K. Yan, L. Zhang, K. Imaizumi, K. Saito, H. Fujita, Automated classification of benign and malignant cells from lung cytological images using deep convolutional neural network, Inf. Med. Unlocked 16 (2019) 100205.
[6] K. Win, S. Choomchuay, K. Hamamoto, M. Raveesunthornkiat, Detection and classification of overlapping cell nuclei in cytology effusion images using a double-strategy random forest, Appl. Sci. 8 (9) (2018) 1608.
[7] K.Y. Win, S. Choomchuay, K. Hamamoto, M. Raveesunthornkiat, L. Rangsiratanakul, S. Pongsawat, Computer aided diagnosis system for detection of cancer cells on cytological pleural effusion images, BioMed Res. Int. 2018 (2018).
[8] V. Gupta, A. Bhavsar, Breast cancer histopathological image classification: is magnification important? in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 17–24.
[9] C. Higgins, Applications and challenges of digital pathology and whole slide imaging, Biotech. Histochem. 90 (5) (2015) 341–347.
[10] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (7) (2015) 1455–1462.
[11] A. Belsare, M. Mushrif, Histopathological image analysis using image processing techniques: An overview, Signal Image Proc. 3 (4) (2012) 23.
[12] M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, B. Yener, Histopathological image analysis: A review, IEEE Rev. Biomed. Eng. 2 (2009) 147–171.
[13] C. Loukas, S. Kostopoulos, A. Tanoglidi, D. Glotsos, C. Sfikas, D. Cavouras, Breast cancer characterization based on image classification of tissue sections visualized under low magnification, Comput. Math. Methods Med. 2013 (2013).
[14] V.-T. Ta, O. Lezoray, A. Elmoataz, S. Schüpp, Graph-based tools for microscopic cellular image segmentation, Pattern Recognit. 42 (6) (2009) 1113–1125.
[15] S. Boumaraf, X. Liu, Z. Zheng, X. Ma, C. Ferkous, A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images, Biomed. Signal Process. Control 63 (2021) 102192.
[16] M. Toğaçar, K.B. Özkurt, B. Ergen, Z. Cömert, BreastNet: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer, Physica A 545 (2020) 123592.
[17] Z. Gandomkar, P.C. Brennan, C. Mello-Thoms, MuDeRN: Multi-category classification of breast histopathological image using deep residual networks, Artif. Intell. Med. 88 (2018) 14–24.
[18] D.J. Ho, D.V. Yarlagadda, T.M. D'Alfonso, M.G. Hanna, A. Grabenstetter, P. Ntiamoah, E. Brogi, L.K. Tan, T.J. Fuchs, Deep multi-magnification networks for multi-class breast cancer image segmentation, Comput. Med. Imaging Graph. 88 (2021) 101866.
[19] H. Tokunaga, Y. Teramoto, A. Yoshizawa, R. Bise, Adaptive weighting multi-field-of-view CNN for semantic segmentation in pathology, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12597–12606.
[20] R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, E.A. Janssen, Multiscale deep neural networks for multiclass tissue classification of histological whole-slide images, 2019, arXiv preprint arXiv:1909.01178.

[21] F.A. Zeiser, C.A. da Costa, G. de Oliveira Ramos, H.C. Bohn, I. Santos, A.V. Roehe, DeepBatch: A hybrid deep learning model for interpretable diagnosis of breast cancer in whole-slide images, Expert Syst. Appl. 185 (2021) 115586.
[22] K.B. Girum, G. Créhange, R. Hussain, A. Lalande, Fast interactive medical image segmentation with weakly supervised deep learning method, Int. J. Comput. Assist. Radiol. Surg. 15 (9) (2020) 1437–1444.
[23] G. Campanella, M.G. Hanna, L. Geneslaw, A. Miraflor, V. Werneck Krauss Silva, K.J. Busam, E. Brogi, V.E. Reuter, D.S. Klimstra, T.J. Fuchs, Clinical-grade computational pathology using weakly supervised deep learning on whole slide images, Nat. Med. 25 (8) (2019) 1301–1309.
[24] F. Kanavati, G. Toyokawa, S. Momosaki, M. Rambeau, Y. Kozuma, F. Shoji, K. Yamazaki, S. Takeo, O. Iizuka, M. Tsuneki, Weakly-supervised learning for lung carcinoma classification using deep learning, Sci. Rep. 10 (1) (2020) 1–11.
[25] X. Xie, C.-C. Fu, L. Lv, Q. Ye, Y. Yu, Q. Fang, L. Zhang, L. Hou, C. Wu, Deep convolutional neural network-based classification of cancer cells on cytological pleural effusion images, Mod. Pathol. (2022) 1–6.
[26] J.E. Van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2020) 373–440.
[27] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, M.-H. Yang, Adversarial learning for semi-supervised semantic segmentation, 2018, arXiv preprint arXiv:1802.07934.
[28] D.-H. Lee, et al., Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: Workshop on Challenges in Representation Learning, Vol. 3, ICML, (2) 2013, p. 896.
[29] Q. Xie, M.-T. Luong, E. Hovy, Q.V. Le, Self-training with noisy student improves imagenet classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10687–10698.
[30] S. Laine, T. Aila, Temporal ensembling for semi-supervised learning, 2016, arXiv preprint arXiv:1610.02242.
[31] T. Miyato, S.-i. Maeda, M. Koyama, S. Ishii, Virtual adversarial training: a regularization method for supervised and semi-supervised learning, IEEE Trans. Pattern Anal. Mach. Intell. 41 (8) (2018) 1979–1993.
[32] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, 2017, arXiv preprint arXiv:1703.01780.
[33] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, C. Raffel, Mixmatch: A holistic approach to semi-supervised learning, 2019, arXiv preprint arXiv:1905.02249.
[34] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E.D. Cubuk, A. Kurakin, H. Zhang, C. Raffel, Fixmatch: Simplifying semi-supervised learning with consistency and confidence, 2020, arXiv preprint arXiv:2001.07685.
[35] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2012) 1915–1929.
[36] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[37] J. Dong, Q. Chen, S. Yan, A. Yuille, Towards unified object detection and semantic segmentation, in: European Conference on Computer Vision, Springer, 2014, pp. 299–314.
[38] P. Pinheiro, R. Collobert, Recurrent convolutional neural networks for scene labeling, in: International Conference on Machine Learning, PMLR, 2014, pp. 82–90.
[39] T. Liu, T. Stathaki, Faster R-CNN for robust pedestrian detection using semantic segmentation network, Front. Neurorobotics (2018) 64.
[40] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[41] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 801–818.
[42] Y. Guo, Y. Liu, T. Georgiou, M.S. Lew, A review of semantic segmentation using deep neural networks, Int. J. Multimed. Inf. Retrieval 7 (2) (2018) 87–93.
[43] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017, arXiv preprint arXiv:1704.04861.
[44] D. Jha, P.H. Smedsrud, M.A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, H.D. Johansen, Resunet++: An advanced architecture for medical image segmentation, in: 2019 IEEE International Symposium on Multimedia, ISM, IEEE, 2019, pp. 225–2255.
[45] J. Du, Y. Zhou, P. Liu, C.-M. Vong, T. Wang, Parameter-free loss for class-imbalanced deep learning in image classification, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–7, http://dx.doi.org/10.1109/TNNLS.2021.3110885.

Shajahan Aboobacker is a Research Scholar at National Institute of Technology Karnataka, Surathkal, India, since 2018. He has six years of teaching experience at various engineering colleges. He graduated from National Institute of Technology Karnataka, Surathkal with an M.Tech degree in Communication Engineering. He graduated from Government Engineering College Wayanad, Kerala with a B.Tech degree in Electronics and Communication Engineering. His research interests include bio-medical image processing and machine learning algorithms.

Deepu Vijayasenan received the Ph.D. degree from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, and the M.E. degree from Indian Institute of Science, Bengaluru, Karnataka, India. He was a Postdoctoral Researcher with the Universität Des Saarlandes, from 2011 to 2012. He is currently working as an Associate Professor with the Department of Electronics and Communication Engineering, National Institute of Technology Karnataka. His research interests include image processing and machine learning.

Sumam David S. is a Professor at Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, Mangalore, India. She obtained her B.Tech degree from University of Kerala, India in 1985, M.Tech and Ph.D. degrees from Department of Electrical Engineering, Indian Institute of Technology, Madras, India in 1986 and 1992 respectively. She joined NITK Surathkal as faculty in August 1991. She has served as Head of the Department of Electronics and Communication Engineering and Dean (Academic) of the Institute. Her research interests are in the areas of Signal processing, VLSI architectures for Signal Processing and Engineering Education.

Dr. Pooja K Suresh did her M.B.B.S from Kasturba Medical College Mangalore in the year 2008 and completed her post graduation in pathology from the same institute in the year 2012. She is currently working as an Additional Professor in Department of Pathology, Kasturba Medical College Mangalore, Karnataka. Her research interests include Cytopathology, Gastrointestinal pathology and Digital pathology.

Dr. Saraswathy Sreeram did her M.B.B.S from Kasturba Medical College Mangalore in the year 2013 and completed her post graduation in pathology from the same institute in the year 2016. She is currently working as an Associate Professor in Department of Pathology, Kasturba Medical College Mangalore Karnataka. Her research interests include Genitourinary and Nephropathology, Gastrointestinal pathology, Head and Neck Pathology and Digital Pathology.