Computers in Biology and Medicine 109 (2019) 242–253

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Automatic segmentation of Sperm's parts in microscopic images of human semen smears using concatenated learning approaches

Reza Akbari Movahed, Elnaz Mohammadi, Mahdi Orooji*

Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Jalal Ale Ahmad, P.O. Box 14115-111, Tehran, Iran

ARTICLE INFO

Keywords:
Male infertility
Sperm morphology
Sperm segmentation
Deep learning
Convolutional neural network
Support vector machine

ABSTRACT

Accurate segmentation of the sperms in microscopic semen smear images is a prerequisite step in automatic sperm morphology analysis. It is a challenging task due to the non-uniform distribution of light in semen smear images, low contrast between sperm's tail and its surrounding region, the existence of various artifacts, high concentration of sperms and wide spectrum of the shapes of the sperm's parts. This paper proposes an automatic framework based on concatenated learning approaches to segment the external and internal parts of the sperms. The external parts of the sperms are segmented using two convolutional neural network (CNN) models which produce the probability maps of the head and the axial filament regions. To obtain acrosome and nucleus segments, the K-means clustering approach is applied to the head segments. A Support Vector Machine (SVM) classifier is used to classify each pixel of the axial filament segments to extract tail and mid-piece regions from obtained segments. The proposed method is validated on the images of the Gold-standard dataset. It achieves 0.90, 0.77, 0.77, 0.78, 0.75 and 0.64 of the average of dice similarity coefficient for the head, axial filament, acrosome, nucleus, tail, and mid-piece segments, respectively. Experimental results demonstrate that the proposed method outperforms state-of-the-art algorithms for the head and its internal parts segmentation. It also segments the axial filament region and its internal parts with desirable accuracy. Different from previous works, the proposed method is able to segment all parts of the sperms which enables automatic quantitative analysis of the sperm morphology.

1. Introduction

Infertility is one of the human problems which affects up to 15% of couples in the world [1]. According to the statistical studies, half of the infertility cases are related to the male factor [2–5]. Today, male infertility is considered as a heterogeneous disorder with different impacts on the infertility issue [6]. Semen analysis according to the world health organization (WHO) guidelines is the cornerstone of the evaluation of male infertility [7,8]. During a typical semen analysis, some microscopic parameters such as the sperm concentration, motility, morphology, vitality, and fragmentation of the spermatic DNA are evaluated. The sperm cell consists of two external parts: head and axial filament. The head includes acrosome and nucleus regions and the axial filament is divided into the tail and mid-piece parts. Sperm morphology parameter describes the shape, size, and morphometric characteristics of the sperm's parts. The assessment of this parameter can provide a robust determinant of the state of male infertility or subfertility [9,10]. Also, it is an informative subject for biologists to investigate human sperm physiology. However, the evaluation of the sperm morphology is still a controversial aspect of the semen analysis. The manual assessment of that is a laborious and subjective process due to the direct intervention of the human function [11]. In addition, there is a high inter and intra-laboratory variability between the manual results of human sperm morphology analysis [12]. Even, the state-of-art computer-aided sperm analysis (CASA) systems require human operator skills to conduct morphology assessment which has a negative effect on the accuracy of the diagnosis procedure [13]. Therefore, it is important to improve the automatic evaluation of human sperm morphology. In automatic sperm morphology analysis, sperm segmentation plays a crucial role that has the most impact on the performance of the automatic morphology assessment. However, it remains to be a challenging task because of the non-uniform distribution of light in semen smear images, low contrast between sperms tail and its surrounding region, the existence of various artifacts such as stained spots and other debris, high concentration of the sperms and the wide spectrum of the shapes of the sperms parts. In recent years, researchers have implemented different methods for automatic human sperm segmentation in microscopic images. These approaches are based on thresholding, clustering,

* Corresponding author.
E-mail address: morooji@modares.ac.ir (M. Orooji).

https://doi.org/10.1016/j.compbiomed.2019.04.032
Received 3 January 2019; Received in revised form 22 April 2019; Accepted 23 April 2019

combining various color spaces and active contour methods. The main disadvantage of these methods is that none of them could segment all components of the sperm cells.

In some recent attempts to tackle this problem, Chang et al. presented a framework for the detection and segmentation of the human sperm head, acrosome, and nucleus, using different color spaces and clustering methods [14]. Firstly, the regions of interest for heads are detected using K-means clustering. Then, the candidate's heads are modified using mathematical morphology and histogram statistical analysis techniques. Upon completion of the head segmentation, each detected head is divided into acrosome and nucleus regions using Otsu thresholding and geometric constraints. One of the most useful contributions of Chang et al. work is introducing Gold-standard dataset which provided more feasibility way to validate new methods for segmentation and identification of the sperms. Shaker et al. provided an automatic detection and segmentation method of the sperm head, acrosome, and nucleus in microscopic images of the human semen smears [15]. In this work, an edge-based active contour algorithm extracted the head contours which operates with an initial mask. To obtain the initial head masks, the Otsu thresholding on the hue channel of HSV color space was employed. The segmentation results include heads and some parts of tails and mid-pieces. For eliminating them, a novel tail removing algorithm was designed to modify segment results by locating and eliminating some midpiece parts from the segmented head. Also, the acrosome and nucleus were segmented using k-means clustering and some morphological operations on the head segments. Malte et al. segmented and detected human sperm cells in non-stained microscopic images by implementing some CNN models [16]. However, this method is efficient for sperm localization in motility analysis it cannot be utilized for morphology analysis because the non-stained samples do not have highlighted specific properties for distinguishing the sperm parts. In this paper, we proposed an automatic framework based on deep learning techniques and classical learning methods to accurately segment the human sperm head, axial filament, tail, mid-piece, acrosome and nucleus in the clinical semen smear microscopic images. Different from previous works, the proposed method is able to segment all parts of the sperm structures which provides the first step of the automatic morphology analysis. It begins with a serialized preprocessing method to suppress the unwilling distortions and elevate the appearance of sperm cells compared to the other objects in smear images. Next, the external parts of the sperms are segmented using two CNN models which produce the probability maps of the head and the axial filament regions. To divide head segments into the acrosome and nucleus parts, the K-means clustering approach is applied to the pixels of head segments. Also, an SVM classifier is used to classify each pixel of the axial filament segments to extract tail and mid-piece regions from obtained segments. It is trained by the pixels of dilated axial filament regions from final multi-channel images. The preliminary multi-channel images contain RGB channels of the preprocessed images and the extracted textural features. The best channels for the representation of the internal parts of the axial filament region are selected using the sequential forward feature selection (SFFS) algorithm. Furthermore, some morphological operations such as closing, opening and geometric constraints are utilized to improve all segments. The remainder of the paper is organized as follows: Section 2 describes the proposed method for segmenting the sperm parts. We introduce the evaluation metrics, the database and the experimental setup in section 3. Section 4 presents the experimental results. In section 5, we discuss more details about the obtained results. Finally, the main conclusion of this study is presented in section 6.

2. Proposed method

The proposed method for segmentation of the internal and external parts of the sperms includes four main stages: pre-processing, external parts segmentation, acrosome and nucleus segmentation, tail and mid-

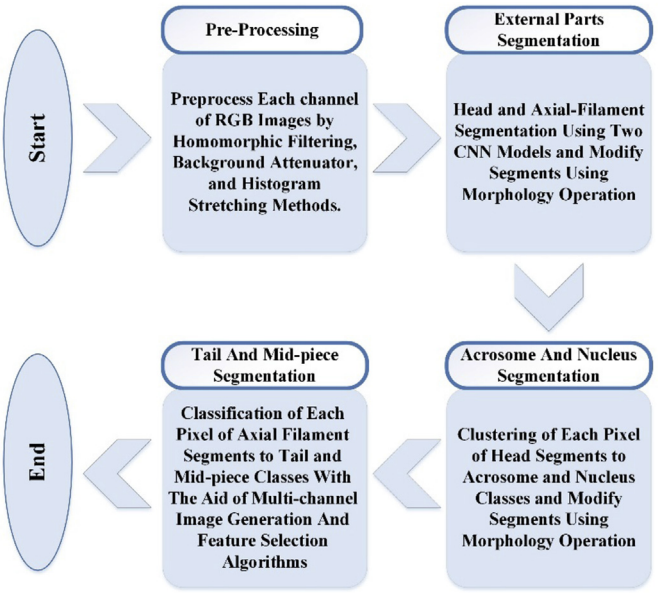

Fig. 1. Overall pipeline of the proposed method for the segmentation of sperm parts.

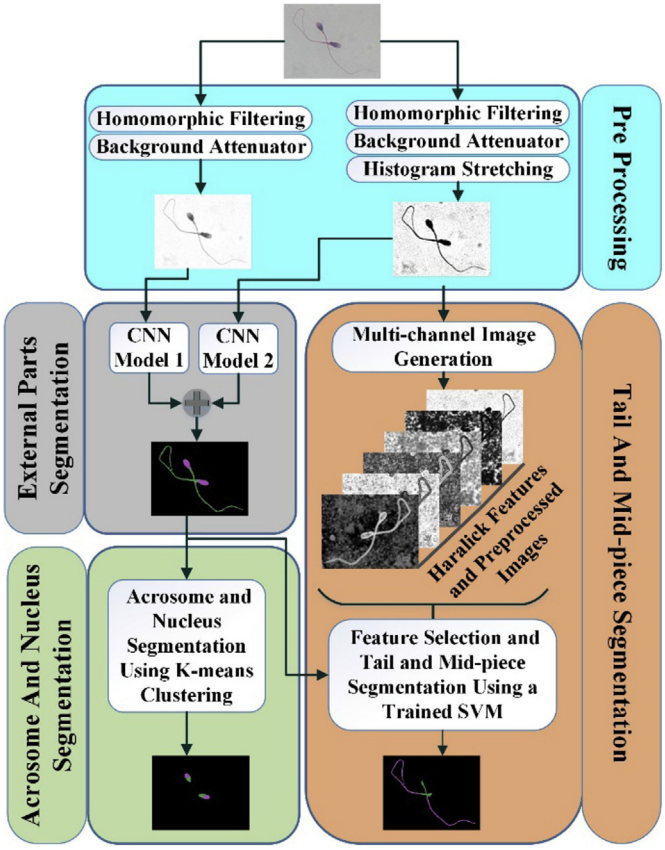

Fig. 2. The graphical diagram of the proposed segmentation method. Firstly, the raw input images are preprocessed using serialized proposed approaches and then the head and axial filament parts are segmented using two different CNN models. The internal parts of the head and axial filament segments are separated using two independent SVM classifier. Especially for internal parts of axial filament segmentation, the multi-channel image generation using textural feature extraction and RGB channels of preprocessed images with the sequential forward selection algorithm is employed.

piece segmentation. Fig. 1 illustrates the overall pipeline of the proposed method for the segmentation of the sperm parts. The details of each stage are presented in the following. Also, Fig. 2 shows the graphical diagram of the proposed method.

2.1. Pre-processing

2.1.1. Homomorphic filtering

Most of the brightfield microscopic images suffer from the non-uniform illumination distribution. It adds a complexity to the histogram of each channel of RGB image and can lead to the creation of some artifacts in image processing purposes. Homomorphic filtering is one of the approaches for correcting the uneven illumination [17]. It compresses the brightness from the light condition while enhancing the contrast from the reflectance properties of the object in the frequency domain. According to this approach, an image can be modeled as the production of the scene illumination and the scene reflectance which is defined as follows:

$$i(x, y) = r(x, y)l(x, y), \tag{1}$$

where i is the image, l represents the scene illumination component, and r is the scene reflectance component. The most frequency components of i are the low-frequency components while r consists of rapid and pronounced variations.

In order to separate out the illumination component from the reflectance component and transform the image into the frequency domain, a Fourier operation followed by a logarithmic function is taken on (1) which results in the following equation:

$$I(u, v) = R(u, v) + L(u, v), \tag{2}$$

where I, R and L are the frequency domain of the logarithmic transforms of image, illumination and reflectance components, respectively. Due to the existence of important information on both high and low frequency components, a modified frequency filter, named high-frequency emphasis filter is used instead of a high-frequency filter, which is the combination of a low and high-frequency filter. It simultaneously amplifies and attenuates high-frequency and low-frequency components of the image respectively. This filter includes an offset value as a low-pass filter and a scaled high-pass filter which is defined as follows:

$$H_e(u, v) = \alpha + \beta H_h(u, v), \tag{3}$$

where Hh is a high-pass filter, α is an offset value, β is a scale factor of the high-pass filter, and He is the high-frequency emphasis filter. In this work, α and β parameters are set to 0.5 and 1.5 values, respectively. This setting leads to amplification of the high frequency components and attenuation of the low-frequency elements. After that, an Inverse Fourier transform operator is employed to obtain the filtered image in the spatial domain. Next, an exponential function is used to neutralize the impact of the logarithmic function. These steps are illustrated in (4)

$$j(x, y) = \exp\{\mathscr{F}^{-1}\{J(u, v)\}\}, \tag{4}$$

where $\mathscr{F}^{-1}$ indicates the inverse Fourier operator, $J(u,v)$ is the output filtering in the frequency domain, and $j(x, y)$ is the inverted log-transform of the output filtering in the spatial domain. Finally, the output values are scaled between 0 and 255. In this paper, the high-pass filter is constructed by subtracting a Gaussian low pass from 1 in the frequency domain which is defined as follows:

$$H_l(u, v) = \exp\left(\frac{(u - u_0)^2 + (v - v_0)^2}{2\sigma^2}\right), \tag{5a}$$

$$H_h(u, v) = 1 - H_l(u, v), \tag{5b}$$

where Hl is the Gaussian low-pass filter, $u0$ and $v0$ is the center of the Gaussian kernel, and σ is the kernel scale. In the experiments, $u0$ and $v0$ are set to the round of $(M + 1 = 2)$ and $(N + 1 = 2)$ where M and N are the height and width of an image, respectively. Also, the kernel scale parameter is set to 10 in the experiment.

2.1.2. Background Attenuator

The background of the semen smear microscopic images contains some inappropriate characteristics for segmentation purposes, e.g. high similarity between sperm's tail and the background which affects negatively the segmentation of the sperm parts. In this paper, a novel algorithm, called Background Attenuator, is proposed to decrease the impact of these characteristics. This approach transforms the background regions to the white color parts approximately while highlights the foreground objects on the image. In this algorithm, the image is convolved many times with a Gaussian spatial filter iteratively which results in an intense blurred image. Next, the input image is divided element-wisely by the blurred image. Finally, the division result is scaled between [0,255] until the result is obtained in the form of the RGB image. The number of times that the image is convolved by the Gaussian filter is 600. Also, the height and width of the filter are equal to half of the image's height.

2.1.3. Histogram stretching

Histogram stretching is an image enhancement technique that stretches the range of intensity values of an image and fits it into the desired range of the intensities. It increases the difference between the maximum intensity value to the minimum one which results in contrast improvement of an image. In this work, each histogram of the channels of the results of the Background Attenuator algorithm in RGB color space is stretched individually and fitted to the [0,255] interval. Next, enhanced channels are categorized together to represent the results in RGB color space. The main reason for using this technique in this work is increasing the contrast between the sperm's tail and the background of the image which facilitates the segmentation of the sperm tails.

2.2. External parts segmentation

The proposed segmentation method of the external parts of the sperms is based on deep learning that it is a newfound branch of machine learning science. Today, deep learning techniques with hierarchical feature learning capability have been applied to many applications such as image classification [18], speech recognition [19], natural language processing [20] and etc. The CNN is one of these techniques which is used in many medical image analysis problems, including segmentation [21], classification [22] and detection [23]. A usual CNN architecture consists of multiple convolutional, pooling, activation and fully-connected layers. It extracts hierarchical features from the input data and generates probability distributions using a softmax function based on the learned features [24]. In this paper, two different CNN architectures are considered to segment the head and axial filament parts.

In these architectures, multiple layers such as convolutional, pooling, activation and fully-connected are stacked such that the output of each layer will be fed to the input of the next layer. A convolutional layer is a set of learnable kernels that takes a feature map as an input and convolve them into these kernels [25] which generate the output feature map as follows:

$$z_{i+1}^{n} = w_{in} * z_i + b_{in}, \forall\ n\varepsilon[1..N], \tag{6}$$

where $z_i \varepsilon \mathbb{R}^{H\times W\times C}$ is the input feature map with W width, H height and C number of channels, win and bin denote the n-th kernel and bias from convolutional layer, respectively, $zin+1$ is the n-th channel of the output feature map $zi+1$, * denotes the convolution operation and N is the number of kernels in convolutional layer. This layer takes account the shape and locally features by sharing the learnable kernels among all units of the input feature map which it is very useful for processing multi-dimensional input that has locally correlated data [24–26]. In these layers, the input feature map is padded with zeros around the border to maintain the height and the width of the input feature map.

Another multi-dimensional layer in CNN is the pooling layer which operates independently on every channel of the input feature map and combines spatially nearby features in that. This combination reduces the number of parameters and computational load in the network. It also controls overfitting by eliminating redundant features. The most common type of this layer is the max pooling layer that picks up the max values over a neighborhood region in each channel of the input feature map [26]. To avoid vanishing/exploding gradient issues in the training procedure and speed up that, rectified linear units (ReLU) is used as the activation layers [18,27] which is defined as follows:

$$f(x) = max(0, x). \tag{7}$$

However, imposing a constant 0 value for the negative inputs can undermine the gradient procedure [28]. To modify this limitation, a leaky rectifier linear unit (LReLU) is employed which considers a small slope for negative inputs. This function is defined as follows:

$$f(x) = max(0, x) + \alpha min(0, x), \tag{8}$$

where α is the slope on the negative part of the function. The stacked convolutional-pooling and activation layers provide local hierarchical features for fully-connected layers that learn higher level feature representations for classification and regression tasks. Fully-connected layers receive flatted input feature map and apply it to the learnable weights and biases. The function of this layer is clearly represented by matrix multiplication and summation which multiplies learnable weights to the flatted input and adds it to the learnable biases. The softmax layer is placed after the last fully-connected layer. The output of the softmax layer indicates the probability of the input data to be the defined classes.

To segment the head and axial filament of the sperms, two CNN models with different architectures are trained. The CNN model 1 and 2 denote CNN architectures for the head and axial filament segmentation, respectively, which are shown in Tables 1 and 2. Both of them are binary output classifiers that take 16 × 16 extracted patches from the preprocessed images. This size of the patches is selected to cover most of the heads, axial-filaments and some artifact objects such as stained spots. For the CNN model 1, the overlapping patches are extracted from enhanced RGB images at the end of Background Attenuator results. While for CNN model 2, the patches are extracted from the RGB image results of Histogram Stretching stage. These models classify each patch to positive regions or not, where the positive regions are head and axial filament parts for CNN model 1 and 2 respectively.

For training these models, the overlapping patches are extracted from the preprocessed images of the training set that each label of them is determined by their central pixel. The central pixel of each training patch belongs to positive regions or not. In order to produce more positive training patches, all positive patches are rotated by 90, 180 and 270° then, they are added to training set patches. This process called data augmentation which increases the number of training sets and reduces overfitting [18]. After obtaining the training patches set, all of the intensity values of the patches are normalized between zero and one. Finally, the learnable parameters of CNN models are trained using stochastic gradient descent optimizer which minimizes the cross-entropy loss function with a constant learning rate. The learning rate and momentum parameters for both models are 0.001 and 0.9, respectively. Also, these models are trained for 70 epochs where each epoch contains several batches. The batch size for both models is 128. Also, the slope of the LReLU activation function is set to 0.333 value for CNN model 2.

Table 1
CNN architecture for head segmentation.

Layer No.	Layer Type	Activation	# kernels	Kernel Size	Output
0	Input	–	–	–	16 × 16 × 3
1	Conv	–	32	3 × 3	16 × 16 × 32
2	Max-pool	ReLU	–	2 × 2	8 × 8 × 32
3	Conv	ReLU	32	3 × 3	8 × 8 × 32
4	Max-pool	–	–	2 × 2	4 × 4 × 32
5	Conv	ReLU	64	3 × 3	4 × 4 × 64
6	Max-pool	–	–	2 × 2	2 × 2 × 64
7	FC	–	–	–	64
8	FC	–	–	–	2
9	Softmax	–	–	–	2

Table 2
CNN Architecture for Axial filament Segmentation.

Layer No.	Layer Type	Activation	# kernels	Kernel Size	Output
0	Input	–	–	–	16 × 16 × 3
1	Conv	LReLU	16	3 × 3	16 × 16 × 16
2	Conv	LReLU	16	3 × 3	16 × 16 × 16
3	Max-pool	–	–	2 × 2	8 × 8 × 16
4	Conv	LReLU	32	3 × 3	8 × 8 × 32
5	Conv	LReLU	32	3 × 3	8 × 8 × 32
6	Max-pool	–	–	2 × 2	4 × 4 × 32
7	FC	–	–	–	64
8	FC	–	–	–	2
9	Softmax	–	–	–	2

During testing procedure, the learned CNN models are applied to the normalized extracted patches from a preprocessed image in the testing set using a pixel-wise sliding window. Before patch extraction using a pixel-wise sliding window technique, the image is padded symmetrically. This would make it possible to extract the patches from the borders of the image. Overall, the learned CNN models with sliding window technique generate two probability maps of the head and axial filament regions with the same size as the image under consideration. Each pixel of these maps is assigned with a probability indicating how likely it belongs to the positive regions.

In order to eliminate the incorrect segments and smooth their boundary, some morphological operations and geometric constraints are utilized to improve the head and axial filament segments. For head segments, the undesired segmented objects with a resulting size below the 102 pixels are removed. Then, a closing morphology process with a disk-shaped structure element of unit radius followed by a filling holes operation is used to smooth the boundary of the head segments. To modify the axial filament segments, a cleaning morphology operation followed by a closing process with a square structure element of 4 pixels is employed. Finally, all objects where their perimeter is smaller than 53, are discarded from the axial filament segments.

2.3. Acrosome and nucleus segmentation

Given the head segments obtained from the previous section, the internal parts of the heads are segmented using a traditional clustering method. Considering the fact that nucleus parts are darker than the acrosome regions due to the staining effect, the K-means clustering approach is applied on the head segments to separate the acrosome and nucleus parts from the obtained head segments. In other words, each pixel of the head segments from preprocessed images is clustered into the two classes using k-means clustering which generates the initial segments of the acrosome and nucleus regions.

To modify the acrosome segments, a four-step postprocessing pipeline is employed. It includes opening, filling holes, closing and removing undesired objects process sequentially. Also, the closing, opening and removing undesired objects process is applied to the initial nucleus segments respectively. It should be mentioned that all the opening and closing operations for modification of acrosome and nucleus segments use a disk-shaped structure element of unit radius. During removing undesired objects process, all segmented clamps with the smaller size than a T1 threshold size are excluded from the acrosome and nucleus segments. The T1 threshold size for acrosome and nucleus segments is 12 and 34, respectively.

2.4. Tail and mid-piece segmentation

The main challenging task of the problem is the tail and mid-piece segmentation, because these parts, especially the tail structure, are very slender. Also, these parts are similar in color pixel values. In section 3, the axial filament segments are obtained which contain the tail and mid-piece parts. To generate more dominant representations that depict the tail, mid-piece parts of the axial filament regions more clearly, thirteen of the Haralick's features are extracted from the green channel of each result of the Background attenuator stage using the gray-level co-occurrence matrix (GLCM) [29]. These features are the angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum variance, sum entropy, entropy, difference variance, difference entropy and information measures of correlation. In this study, the size of the extracted patches for calculation the GLCM matrix is 5×5 and it considers all directions in the calculation. Other configurations of this process such as the distance between the pixel of interest and its neighbor and number of quantized levels are set to 1 and 64 values, respectively. Before feature extraction, the images are padded symmetrically to preserve their dimension. To convert each feature as a gray-scale image, all values of that are scaled between [0,255]. For considering the preprocessed images, all RGB channels of each result of the Background Attenuator and Histogram Stretching are categorized with the obtained features across the third dimension of the array. It causes to build a multi-channel image in which each pixel contains 19 values.

To select a subset of the obtained channels of the multi-channel images that provide the best predictive power between the tail, mid-piece, and border regions, the SFFS algorithm is used that sequentially adds features to an empty candidate set until the addition of further features does not decrease the decision criterion. In this work, the decision criterion is the mean of validation accuracies during the 5-fold cross-validation process. In this process, the training set pixels are randomly partitioned into 5 equal sized subsamples that 4 subsamples are used to train a quadratic discriminant analysis (QDA) model and the remaining subsample is retained as the validation set for testing the model. This process is repeated 5 times that each of the subsamples is used exactly once as the validation data and it produces a validation accuracy for each time. The final subset of the channels which is determined by SFFS contains contrast, sum average, sum average, sum entropy, entropy, red and blue channels of histogram stretching result.

To extract the tail and mid-piece parts from the axial filament segments, an SVM classifier with RBF kernel is trained by each pixel of the axial filament regions that each pixel contains selected features. The pixels of the axial filament regions belong to the tail and mid-piece classes. After training, it classifies each pixel with seven selected features of the axial filament segments into the tail and mid-piece classes which produces tail and mid-piece segments.

For improving the tail segments, a closing morphology operation with a one-pixel structure element is used. Also, a removing undesired objects process followed by a closing operation with a disk-shaped structure element of unit radius is applied to the mid-peice segments. During the process of removing the undesired objects from mid-piece segments, all clamps where their are smaller than 23 are eliminated.

3. Experimental setup

3.1. Dataset

The proposed framework is evaluated on the semen smear images of the public Gold-standard dataset [14]. It includes 20 images with many normal and abnormal sperm cells that the resolution of each image is 780×580. Although the handmade ground truths are publicly available, some of them have defects that should be corrected. Therefore, we annotated all parts of the sperms of each image. To examine the ground truths quality, annotated images are sent to a referent expert of the sperm morphology analysis field. Finally, we refined the annotated images in accord with the comments of the referent expert. Overall, handmade sperm ground-truths are used in all training procedures of the proposed framework. Also, they are used for evaluating the obtained results.

3.2. Implementation details

The experiments of the proposed framework are performed on a machine with 64 GB RAM, 12 processors of Intel Core I7-5930 k CPU (3.50 GHz) and NVIDIA GeForce 960 GTX graphics processing unit (GPU). The proposed CNN models are trained and tested with the PyTorch framework [30] in Python 3.6 on Ubuntu 16.04 operating system. Other parts of the framework such as patch extraction, preprocessing, postprocessing and traditional classification methods are implemented MATLAB R2018a. Also, some numerical packages such as SciPy [31], Numpy, and PyTables [32] are used to transfer the data between Python and MATLAB environments. For the fair comparison with previous works [14,15] and the limited number of the images, the leave-one-out cross validation is used for performance assessment of this research. In the leave-one-out cross validation, the training and testing cycle is repeated 20 times that in each cycle 19 of the 20 images from the Gold-standard dataset are used for the training procedure of the proposed framework and one remaining image is selected for the testing set. The minimum number of patches belongs to the head, axial filament, and background classes during leave-one-out cross validation are 282,996,675,368 and 8,419,917, respectively.

3.3. Evaluation metrics

The segmentation and detection performance of the proposed method is assessed using different performance metrics where some of them consider the object-level (head detection) errors and the others consider the pixel-level (sperm particles segmentation) errors for each segmented image.

The utilized metrics that take the pixel-level errors into account, are Dice similarity coefficient (DC) and Jaccard Index (J) which are defined as follows:

$$J = \frac{|G_i \cap S_j|}{|G_i \cup S_j|}, \qquad DC = \frac{2|G_i \cap S_j|}{|G_i| + |S_j|}, \tag{9}$$

where *Gi* and *Sj* are the sets of true pixels in the ground truth region *i* and the segmented region *j*. Furthermore, some statistical measures based on the confusion matrix are employed to evaluate the segmentation results in a pixel-level manner. These measures are Sensitivity (SE) and False Discovery Rate(FDR) which are defined as follows:

$$SE = \frac{TP}{TP + FN}, \qquad FDR = \frac{FP}{FP + TN}, \tag{10}$$

where *TP*, *TN*, *FN* and *FP* denote the number of true positives, true negatives, false positives and false negatives, respectively. Considering that the ground-truths and the segmentation results are both binary images with true and false values, a true pixel in the ground-truth image is considered as a true positive if it is true on the segmented image. Otherwise, it is regarded as a false negative. A false pixel in the ground-truth image is considered as a true negative if it is false on the segmented image. Otherwise, it is regarded as a false positive.

To evaluate the detection accuracy of the individual sperm's heads, the F1-score (F1) metric is used which is defined as follows:

$$F1 = \frac{2N_{tp}}{2N_{tp} + N_{fp} + N_{fn}}, \tag{11}$$

where *Ntp*, *Nfp* and *Nfn* are the number of true positives, true negatives, false positives and false negatives objects, respectively. A true positive object is a segmented head that covers at least 50% of the corresponded ground truth's head. Otherwise, it is considered as false positive. The

Nfn is calculated as the difference between the number of ground truth heads and the *Nfp*.

4. Results

In this section, the performance of the proposed framework for segmentation of each part of the sperms is evaluated using different comparative methods. These comparative methods consist of state-of-the-art algorithms for the sperm segmentation and traditional learning-based segmentation systems. In order to compare with state-of-the-art algorithms for sperm segmentation approaches, the methods described in Refs. [14,15] are implemented. To evaluate the performance of the proposed CNN models for head and axial filament segmentation, different segmentation systems based on the traditional learning methods are implemented and the obtained results from these systems and the proposed method are compared with each other. These systems consist of three main parts: feature extraction, feature selection, and a traditional classifier. In the first part, the haralick features are generated from the green channel of the Background Attenatur results as new representations for each image. Next, the obtained representations are categorized with RGB channels of the preprocessed images and the best of them are selected by SFFS for the pixel-wise classification goal. Finally, a traditional classifier is trained by multi-channel pixels to segment the external parts of the sperms in the images. The classifiers that are used in these systems are SVM, linear discriminant analysis (LDA), QDA, decision tree (DT), k-nearest neighbor (KNN), random forest (RF), naive Bayes (NB) and multi-layer perceptron (MLP). In the following, the evaluation results for head and axial filament segmentation and their internal parts are presented, respectively. Furthermore, different experiments are conducted for the segmentation of the internal parts of the head and axial filament parts. Fig. 3 shows some samples of the preprocessed images and the corresponded segmentation results using the proposed method.

4.1. Head, acrosome and nucleus segmentation

In the first experiment, the performance of the proposed method for head and its internal parts segmentation is compared with the methods described in Refs. [14,15]. Table 3 summarizes the quantitative results of the proposed method and these approaches for the head, acrosome and nucleus segmentation. It is clear that the experimental results for the head and its internal parts segmentation are significantly improved using the proposed framework.

The main reason for the degradation of the obtained results using the methods described in Refs. [14,15] in the experimental setup of this paper is that they ignored some abnormal sperms in their framework while all the normal and abnormal sperm cells in microscopic images are considered in the experimental setup of this paper. Compared with the [15], the mean of DC metric for the head, acrosome, and nucleus segments is improved by 12.01%, 8.87%, and 4.37%, respectively, using the proposed method. Also, the standard deviation of DC metric for the head, acrosome and nucleus segments is reduced by 61.29%, 45.94% and 63.63% using the proposed method compared with [15]. It means that the segmentation performance of the proposed method is more stable and more accurate than [15]. Compared with [14], the mean of DC metric for the head, acrosome and nucleus segments is improved by 20.53%, 28.61%, and 10.67%, respectively, using the proposed method. In other words, the standard deviation of DC metric for the head, acrosome and nucleus segments is reduced by 61.29%, 45.94%, and 63.63%, respectively, using the proposed method compared with [14]. These results indicate that the proposed method is much better than [14] for the head and its internal parts segmentation. In terms of FDR and SE metrics, the proposed method obtains the higher means of SE metric and the lower means of FDR metric for the head, acrosome, and nucleus segmentation compared with the previous methods [14,15]. It indicates that the head and its internal parts segmentation results using the [14,15] methods suffer from the over-segmentation and under-segmentation issues, while the proposed method leads to a more precise segmentation of the head, acrosome and nucleus parts compared to the previous works. Fig. 4 shows some head and its internal parts segmentation results in which the [14] while [15] methods lead to under segmentation, over-segmentation and missdetection issues and the proposed method obtains more accurate segmentation results compared to the [14] while [15] methods. Also, The mean of the F1 metric for head detection by the methods in Refs. [14,15], and the proposed method are 0.8597, 0.8761 and 0.9450, respectively. It shows that the proposed method has better performance

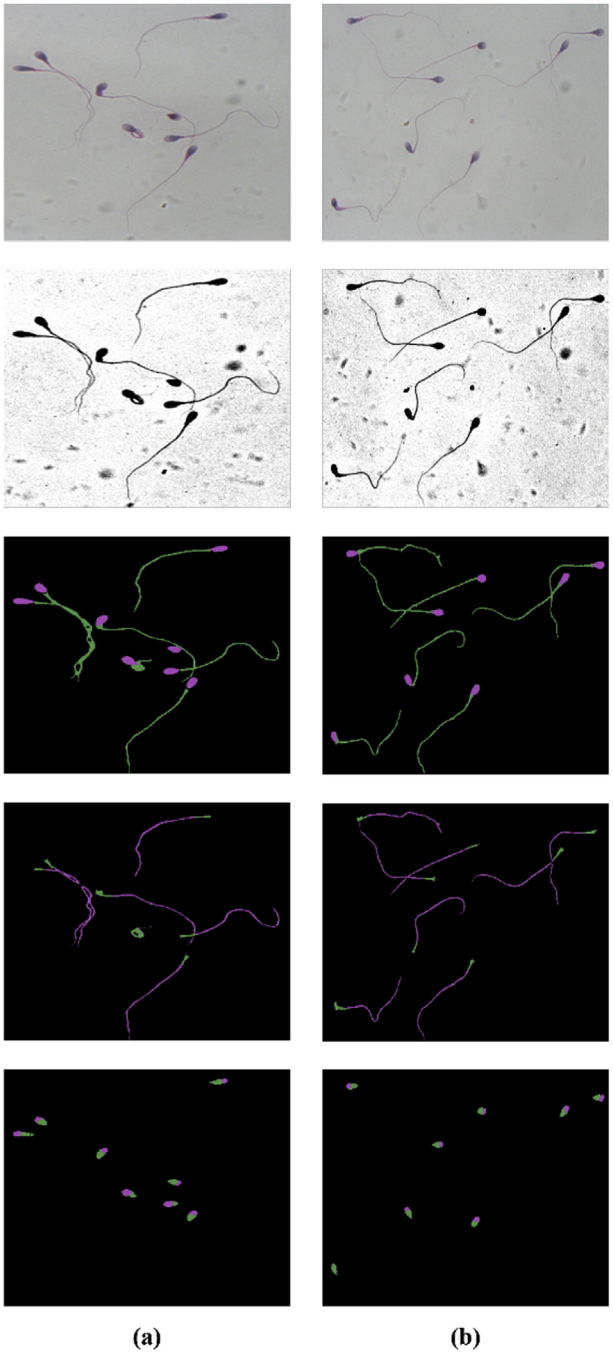

Fig. 3. Examples of some results obtained by using the proposed method. Each column presents an image and corresponded results using the proposed framework. The first, second, third, fourth and fifth row is the original image, the final pre-processed image, final external parts segments, final tail and mid-piece segments, final acrosome and nucleus segments respectively.

Table 3
Performance Comparison of Head, Acrosome and Nucleus Segmentation Methods In Terms of Mean and Standard deviation (STD) of Dice, Jaccard, Sensitivity and FDR Metrics.

Method	Part	DC (Mean ± Std)	J (Mean ± Std)	SE (Mean ± Std)	FDR (Mean ± Std)
Proposed	Head	0.904 ± 0.024	0.826 ± 0.040	0.904 ± 0.038	0.093 ± 0.032
	Acrosome	0.773 ± 0.040	0.631 ± 0.051	0.685 ± 0.054	0.109 ± 0.050
	Nucleus	0.788 ± 0.028	0.651 ± 0.038	0.740 ± 0.039	0.155 ± 0.037
[15]	Head	0.807 ± 0.062	0.680 ± 0.089	0.760 ± 0.096	0.133 ± 0.052
	Acrosome	0.710 ± 0.074	0.555 ± 0.087	0.624 ± 0.097	0.165 ± 0.073
	Nucleus	0.755 ± 0.077	0.612 ± 0.095	0.740 ± 0.096	0.219 ± 0.090
[14]	Head	0.750 ± 0.107	0.610 ± 0.123	0.655 ± 0.137	0.091 ± 0.064
	Acrosome	0.601 ± 0.093	0.436 ± 0.092	0.463 ± 0.098	0.109 ± 0.088
	Nucleus	0.712 ± 0.109	0.562 ± 0.119	0.664 ± 0.142	0.207 ± 0.075

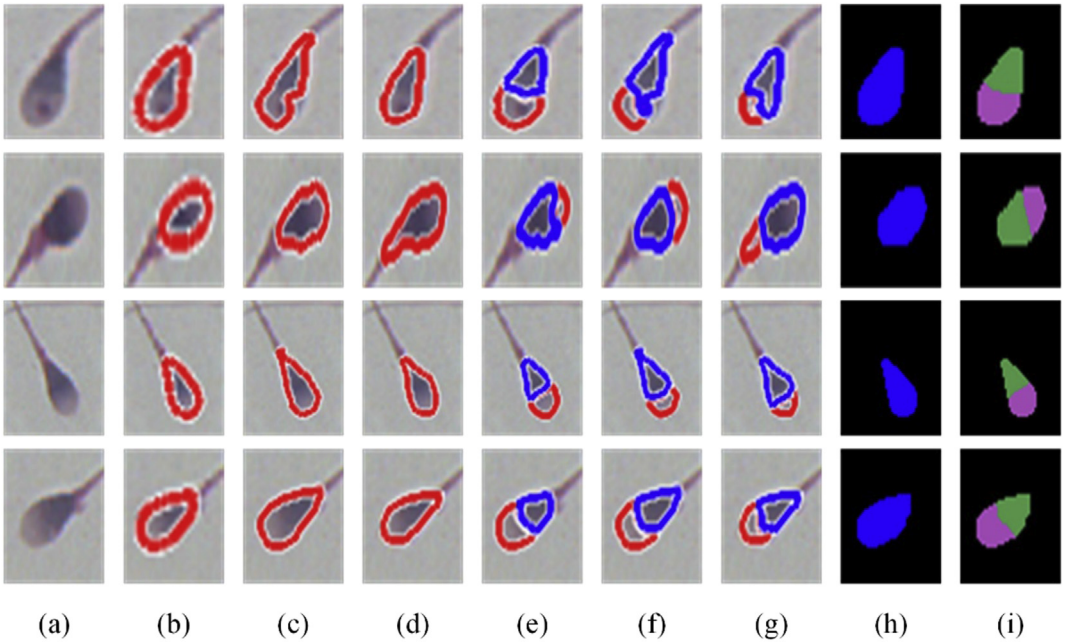

Fig. 4. Some head and its internal parts segmentation results using [14,15] and the proposed method. Each row represent a cropped head and its corresponding segment and ground truth. For acrosome and nucleus segments, blue lines represent nucleus regions and red lines illustrate acrosome regions. The acrosome and nucleus ground truths depict by green and pink colors respectively. (a) Cropped head image, (b) Head segmentation using the proposed method, (c) Head segmentation using Shaker method, (d) Head segmentation using Chang method, (e) Acrosome and Nucleus segmentation using the proposed method, (f) Acrosome and Nucleus segmentation using the Shaker method, (g) Acrosome and Nucleus segmentation using the Chang method, (h) Head ground truth and (i) Acrosome and Nucleus ground truth. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

for detection of sperms head than previous methods [14,15]. Generally, the methods in Refs. [14,15] are not able to detect the challenging heads such as heads with excessive residual cytoplasm and the heads that are overlapped with the other objects of the semen smear images. They are also unable to reject some artifacts like stained spots in semen smear images. However, the proposed method has a strong performance to detect heads with different shapes and sizes. It also rejects the most artifacts which are similar to the head of the sperms. Fig. 5 shows some challenging heads for segmentation goal which are segmented accurately using the proposed method and could not be detected using the previous methods [14,15].

As shown in Table 4, the proposed CNN model produces the best results for the head segmentation among learning-based segmentation methods. In more details, the proposed CNN model achieves a higher and lower mean of DC and FDR metrics, respectively, compared to the traditional learning-based segmentation methods. Specifically, SVM, NB, QDA, and KNN methods lead to the high mean of SE and FDR metrics, indicating the high over-segmentation rate of the head segmentation results of these methods, whereas the CNN achieves the lowest mean of FDR in addition to the high mean of SE. Among the traditional learning-based methods, the LDA, DT, and RF have better performance for the head segmentation. If the MLP model is designed deeper, it may provide better performance than the current MLP model. However, a deeper MLP model has a more computational load and requires much training and testing time cost. The reason why the SVM model provides low accuracy for head segmentation is that this classifier is trained with a sampled training set to possible segmentation process with this classifier.

According to Fig. 6, the obtained DC values using the proposed CNN model are higher than the produced DC values using the traditional learning methods. Also, the height of CNN box for the head segmentation is very less than the height of traditional learning-based boxes. It reveals that most of the head segmentation results which are produced by the CNN model are also close to the average performance.

Also, different supervised and unsupervised classification methods are used to classify each pixel of the head segments to acrosome and nucleus classes. The experimental results of these methods for acrosome and nucleus segmentation are presented in Table 5. Among these methods, K-means clustering achieves the highest mean of DC and J with acceptable standard deviations for all metrics. The Otsu thresholding method ranked in the second place for classification of each pixel of the head segments. Other approaches such as SVM and Gaussian

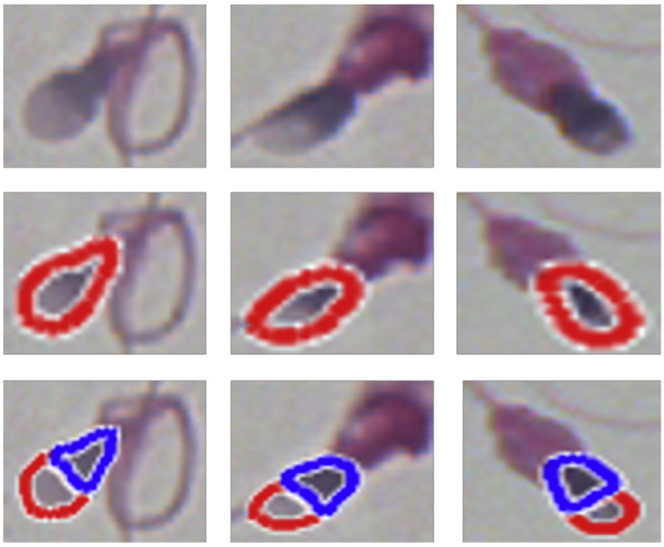

Fig. 5. Some challenging heads for segmentation goal which they are segmented accurately using the proposed method and could not be detected using [14,15]. Each column represents a challenging head and the first, second and third row shows head, final head segment and final internal parts of the head segmentation results respectively.

mixture model (GMM) produce poor performances for classifying head pixels.

4.2. Axial filament, tail and mid-piece segmentation

Fig. 7 shows some examples of the axial filament, tail and mid-piece segmentation results which are obtained using the proposed method. Table 4 lists the quantitative results of the axial filament segmentation results using the proposed CNN model and other segmentation systems based on the traditional learning methods. Furthermore, the box plots of the obtained dice values for the axial filament segments which are produced by CNN and other traditional learning-based segmentation systems are shown in Fig. 6. It is obvious that the proposed CNN model achieves the best experimental results among all learning-based segmentation systems for axial filament segmentation.

According to the results in Table 4, the mean of DC, J, SE and FDR metrics for axial filament segmentation using the proposed method is 0.772, 0.630, 0.763 and 0.208, respectively, which shows the superiority of the proposed CNN model over the traditional learning-based methods for axial filament segmentation. However, the mean of SE metric for the axial filament segmentation using the proposed CNN model is slightly low which indicates the under-segmentation issue of axial filament segments. Qualitatively, the end sections of the axial filament parts in some instances could not be segmented using the proposed method.

It might be the main reason for the low SE metric of axial filament segmentation using the proposed method. Also, the mean of FDR metric for the axial filament segmentation using the proposed CNN model is relatively high which shows some over-segmentation occurrences of axial filament segments. For instance, when some axial filament parts are close to each other, the proposed method considers some parts of the background image as the axial filament segment incorrectly.

On the other hand, most of the learning-based segmentation systems have poor performance for axial filament segmentation. As shown in Table 4, most of these methods such as SVM, KNN, QDA, and NB obtain high averages of the SE and FDR metrics for axial filament segmentation which indicates the over-segmentation issues of these approaches for axial filament segmentation. Also, The LDA, DT, RF, and MLP approaches lead to intense under-segmentation issues because these approaches obtain the low means of SE metric for axial filament segmentation.

Different approaches are carried out to achieve acceptable results for dividing the axial filament segments into tail and mid-piece parts. Some of them are equipped with multichannel image generation and feature selection processes and the others are dealing with RGB pixel values of Background Attenatur results in the preprocessing step. Table 6 summarizes the pixel-wise performance metrics for the tail and mid-piece segmentation using different classification methods. In these tables, all classification methods that are utilized with multi-channel image generation and SFFS algorithm are called with optimized suffix. The experimental results show that the optimized SVM achieves a greater average and lower standard deviation of DC and J metrics among all experimented methods. For mid-piece segments, it achieves a relatively high average of FDR metric.

The main reason for higher FDR metric values for mid-piece segments is related to the twisted tails of the sperms. In other words, most of the pixels of the twisted tails are incorrectly classified as a mid-piece pixel because the color of these regions is very similar to the mid-piece regions. It is observed that the optimized classifiers achieve much better results than directly employing them for classifying each pixel of the

Table 4
Performance Comparison of Learning Methods For Head and Axial filament Segmentation In Terms of Mean and Standard deviation (STD) of Dice, Jaccard, Sensitivity and FDR Metrics.

Method	Part	DC (Mean ± Std)	J (Mean ± Std)	SE (Mean ± Std)	FDR (Mean ± Std)
Proposed CNN Model	Head	0.904 ± 0.024	0.826 ± 0.040	0.904 ± 0.038	0.093 ± 0.032
	Axial Filament	0.772 ± 0.039	0.630 ± 0.050	0.763 ± 0.085	0.208 ± 0.046
DT	Head	0.742 ± 0.167	0.610 ± 0.158	0.787 ± 0.178	0.263 ± 0.105
	Axial Filament	0.419 ± 0.145	0.275 ± 0.114	0.314 ± 0.143	0.259 ± 0.068
SVM	Head	0.608 ± 0.064	0.440 ± 0.065	0.988 ± 0.006	0.556 ± 0.067
	Axial Filament	0.495 ± 0.043	0.330 ± 0.038	0.942 ± 0.063	0.662 ± 0.038
KNN	Head	0.659 ± 0.063	0.495 ± 0.068	0.985 ± 0.011	0.500 ± 0.069
	Axial Filament	0.543 ± 0.047	0.374 ± 0.045	0.848 ± 0.072	0.596 ± 0.049
QDA	Head	0.618 ± 0.054	0.449 ± 0.056	0.965 ± 0.017	0.542 ± 0.059
	Axial Filament	0.538 ± 0.055	0.370 ± 0.049	0.796 ± 0.127	0.589 ± 0.042
LDA	Head	0.765 ± 0.077	0.626 ± 0.095	0.772 ± 0.060	0.234 ± 0.110
	Axial Filament	0.457 ± 0.071	0.298 ± 0.059	0.463 ± 0.112	0.536 ± 0.067
MLP	Head	0.729 ± 0.214	0.605 ± 0.192	0.673 ± 0.216	0.132 ± 0.085
	Axial Filament	0.455 ± 0.144	0.296 ± 0.050	0.355 ± 0.074	0.341 ± 0.067
RF	Head	0.746 ± 0.169	0.616 ± 0.163	0.774 ± 0.175	0.243 ± 0.114
	Axial Filament	0.419 ± 0.144	0.275 ± 0.114	0.314 ± 0.143	0.261 ± 0.069
NB	Head	0.504 ± 0.061	0.339 ± 0.054	0.985 ± 0.010	0.659 ± 0.055
	Axial Filament	0.532 ± 0.046	0.363 ± 0.042	0.856 ± 0.100	0.610 ± 0.043

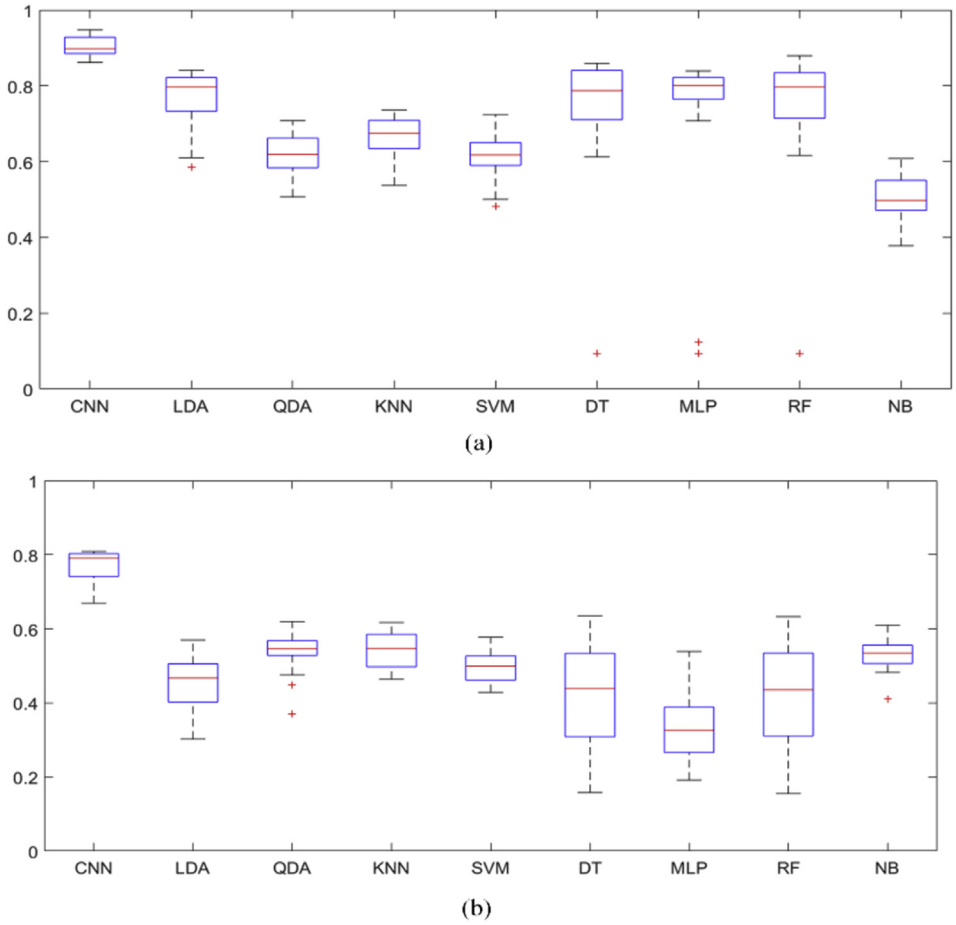

Fig. 6. Box plots of the obtained Dice values for head(a) and axial filament(b) segments produced by different learning-based segmentation systems.

axial filament segments. It demonstrates that the accurate classification of each pixel of axial filament segments might need sophisticated feature design to provide better classification results.

5. Discussion

In this work, a fully automatic framework for accurate sperm segmentation in microscopic images of the human semen smears is presented. Unlike the state-of-art methods that segment only the head and its internal parts, the proposed method is able to segment all the external and internal parts of the sperms. The proposed method uses serialized preprocessing methods to emphasize the appearance of the sperm cells than other objects in the images. It also suppresses undesired characteristics of the image such as the non-uniform distribution of light and the low contrast between the sperm's tails and their surrounding region. The qualitative results indicated that the proposed preprocessing method elevated the appearance of the sperm cells and attenuated unwilling distortion significantly while the state-of-art methods did not present any preprocessing approach to achieve preprocessing goals. For the head and its internal parts segmentation, the proposed method outperforms the state-of-art approaches [14,15] and traditional classification methods in terms of segmentation accuracy and reliability. Due to the automated learned feature representation from the CNN model and the proposed preprocessing pipeline, the reliable probability maps of the sperm's heads could be obtained. It also facilitates dividing these regions into the acrosome and nucleus parts. Furthermore, It is robust against different shapes and sizes of the sperm's head because the proposed CNN model is trained by the patches of various types of sperm's head, while the previous methods could not detect some abnormal heads of the sperms. Another reason for the better performance of the proposed head segmentation method is the precise pixel-wise segmentation of the detected heads. The detected

Table 5
Performance Comparison of Supervised And Unsupervised classification Methods For Acrosome Segmentation In Terms of Mean and Standard deviation (STD) of Dice, Jaccard, Sensitivity and FDR Metrics.

Method	Part	DC (Mean ± Std)	J (Mean ± Std)	SE (Mean ± Std)	FDR (Mean ± Std)
K-means	Acrosome	0.773 ± 0.040	0.631 ± 0.051	0.685 ± 0.054	0.109 ± 0.050
	Nucleus	0.788 ± 0.028	0.651 ± 0.038	0.740 ± 0.039	0.155 ± 0.037
Otsu	Acrosome	0.761 ± 0.043	0.616 ± 0.054	0.666 ± 0.059	0.106 ± 0.047
	Nucleus	0.781 ± 0.026	0.649 ± 0.039	0.738 ± 0.042	0.164 ± 0.036
SVM	Acrosome	0.702 ± 0.083	0.547 ± 0.098	0.580 ± 0.122	0.078 ± 0.055
	Nucleus	0.780 ± 0.030	0.644 ± 0.040	0.736 ± 0.075	0.165 ± 0.054
GMM	Acrosome	0.536 ± 0.259	0.406 ± 0.235	0.546 ± 0.325	0.403 ± 0.154
	Nucleus	0.651 ± 0.042	0.484 ± 0.047	0.574 ± 0.066	0.207 ± 0.166

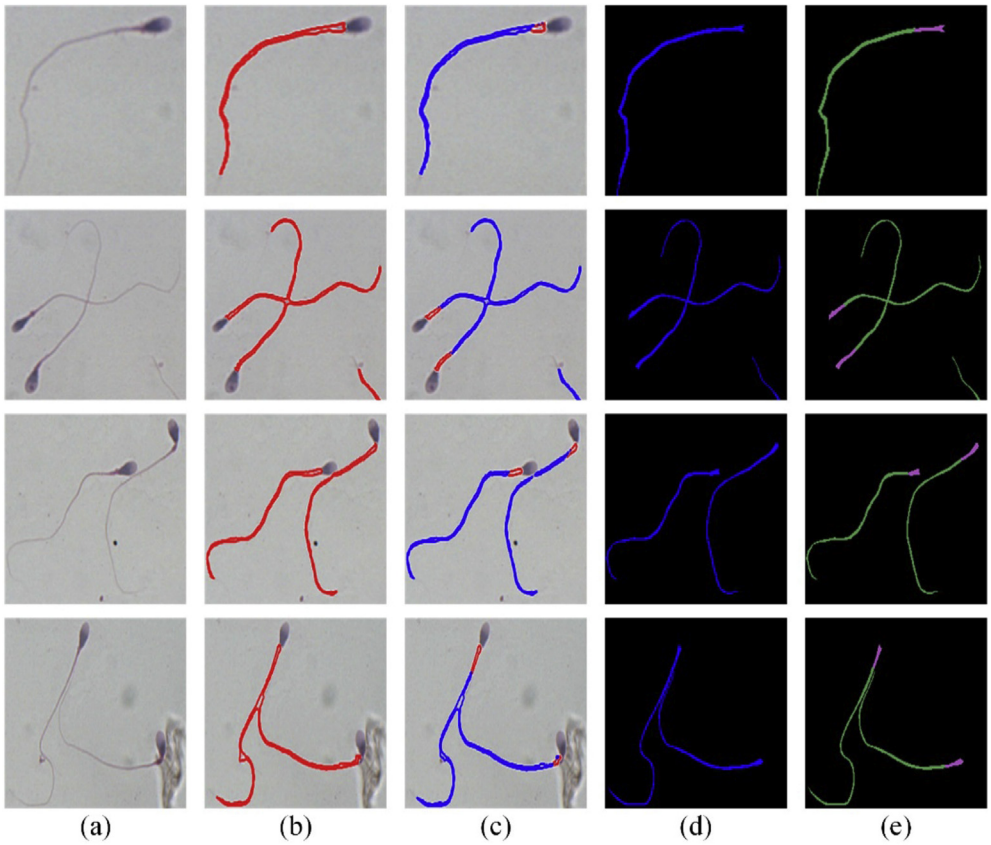

Fig. 7. Some example of axial filament, tail and mid-piece segmentation results obtained by using the proposed method. For tail and mid-piece segments, blue lines represent tail regions and red lines illustrate mid-piece regions. The tail and mid-piece ground truths depict by green and pink colors respectively. Cropped Image of a sperm(a), axial filament segments(b), tail and mid-piece segmentation results(c), axial filament ground truth(d), tail and mid-piece ground truth(e). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

heads using previous methods and traditional learning algorithms suffer from the over-segmentation and under-segmentation issue. For instance, some of these methods in some cases segment the detected heads with some regions of the axial filament parts or they do not segment some of the regions of the acrosome in their head segmentation method. Although the proposed method outperforms the previous works of sperm's head segmentation, it can be improved by training the CNN model with more challenging heads for segmentation. Moreover, applying the multi-channel image generation and SFFS algorithms can improve the separation of the head segments to the acrosome and nucleus regions.

Another major contribution in this paper is the axial filament, tail and mid-piece segmentation which are not addressed in the previous works [14,15]. The experimental results demonstrate the superior performance of the proposed CNN model for axial filament segmentation. However, the proposed approach for axial filament segmentation needs some modifications and refinements to improve its performance. The most important issue of the proposed method for axial filament segmentation is the under-segmentation and over-segmentation problems which occurs in some cases. For example, the proposed approach does not segment the end sections of the axial filament parts in some cases which is very slender. Another major defect of the axial filament segmentation using the proposed method is the over-segmentation of axial filament segments in the presence of several tails near each other. In these cases, in addition to the axial filament parts, some regions of the background are segmented incorrectly. These defects could be modified by developing the architecture of the CNN model and its loss function. For example, Yan et al. presented a CNN model with a novel loss function for retinal blood segmentation [33]. Their proposed loss function emphasizes more on the thickness consistency of thin vessels in the training process. Considering the apparent similarity of the axial filament part of the sperms with the retinal structure, the ideas like this can improve the axial filament segmentation.

The experimental results of the classification of axial filament pixels

Table 6
Performance Comparison of Tail And Mid-piece Segmentation Methods In Terms of Mean and Standard deviation (STD) of Dice, Jaccard, Sensitivity and FDR Metrics.

Method	Part	DC (Mean ± Std)	J (Mean ± Std)	SE (Mean ± Std)	FDR (Mean ± Std)
Optimized SVM	Tail	0.750 ± 0.051	0.602 ± 0.062	0.745 ± 0.090	0.236 ± 0.055
	Mid-piece	0.645 ± 0.101	0.483 ± 0.105	0.612 ± 0.145	0.284 ± 0.096
Non-Optimized SVM	Tail	0.703 ± 0.042	0.543 ± 0.048	0.744 ± 0.088	0.326 ± 0.043
	Mid-piece	0.303 ± 0.103	0.182 ± 0.072	0.209 ± 0.088	0.386 ± 0.137
Optimized KNN	Tail	0.736 ± 0.046	0.585 ± 0.056	0.728 ± 0.085	0.245 ± 0.053
	Mid-piece	0.640 ± 0.086	0.477 ± 0.091	0.603 ± 0.130	0.286 ± 0.099
Non- Optimized KNN	Tail	0.684 ± 0.036	0.521 ± 0.041	0.612 ± 0.067	0.214 ± 0.051
	Mid-piece	0.538 ± 0.099	0.374 ± 0.091	0.437 ± 0.119	0.251 ± 0.097
Optimized DT	Tail	0.738 ± 0.047	0.587 ± 0.056	0.729 ± 0.085	0.243 ± 0.052
	Mid-piece	0.641 ± 0.087	0.478 ± 0.090	0.611 ± 0.134	0.294 ± 0.091
Non-Optimized DT	Tail	0.711 ± 0.043	0.554 ± 0.050	0.718 ± 0.086	0.286 ± 0.053
	Mid-piece	0.531 ± 0.107	0.368 ± 0.098	0.424 ± 0.124	0.238 ± 0.097

from obtained segments show the superiority of the SVM classifier which is optimized with multi-channel image generation and SFFS algorithm over the other classification approaches. These results indicate that the use of the multi-channel image generation and feature selection algorithms improves the classification accuracy of axial filament pixels. Although the proposed method for classification of each pixel of axial filament segments has the desired performance on the normal conditions of the axial filament segments, it leads to failure segmentation on some challenging conditions of the axial filament segments. The most challenging condition that the proposed method does not provide the desired performance is the segmentation of the twisted tail. Generally, the proposed approach classifies the pixels of the detected twisted tails to the mid-piece class, whereas it should classify them to the tail class. Overall, one of the main the main opportunity for extending the scope of this paper is the improvement of the segmentation of the axial filament and its internal parts.

Another issue of the proposed framework is the high computational cost of that. In more details, There are two CNN models, multi-channel image generation process and SFFS algorithm in the proposed framework that each of them, especially CNN models, have a high computational load. Various solutions can be recommended to solve this problem. One recommendation is proposing a three-output CNN model that segments each microscopic semen smear image to the head, axial filament and background classes. This model can be a conventional model of CNN or a sophisticated novel model such as Inception network [34]. Considering the success of Inception network for image classification, the use of it can be a better option for developing the proposed method. On the other hand, the fully convolutional network (FCN) [35] models such as Segnet [36] and Unet [37] could be applied to the microscopic semen smear images for sperm segmentation. Also, instead of using SVM classifier and K-means clustering for dividing head and axial filament segments, the extracted feature for each patch by CNN models could be utilized for the segmentation of tail, mid-piece, acrosome, and nucleus.

One of the major limitations of the sperm segmentation in the microscopic semen smear images is the lack of multiple databases of sperms images with different staining methods. So far, there is only one public database of microscopic semen smear images which is not enough to evaluate sperm segmentation methods. Therefore, The future works can provide new datasets of microscopic human semen smear images with different staining methods to aid the development and benchmarking of the generalized sperm segmentation approaches.

6. Conclusion

In this paper, an automatic framework for human sperm segmentation in the semen smear microscopic images is presented. It includes a serialized preprocessing method, two CNN models, K-means clustering, SVM classifier, multi-channel image generation, and SFFS algorithm. The comparison between the head segmentation result using the proposed method and previous works reveals that the proposed approach outperforms the previous works for the head, acrosome and nucleus segmentation. It also provides the first approach for the axial filament and its internal parts segmentation. Although the proposed method achieves an accurate and reliable performance for the head, acrosome and nucleus segmentation, it needs some modification and improvements for the axial filament and its internal part segmentation. As future work, the enhancement of the axial filament and its internal parts segmentation, reducing computational cost, provide new datasets of microscopic human semen smear images with different staining methods and utilizing FCN models for sperm segmentation can be considered. Overall, the proposed method can enable automatic extraction of the morphometric features for morphology analysis of the sperm cells that is a very important part of semen analysis and infertility treatment process.

References

[1] W. Cui, Mother or Nothing: the Agony of Infertility, (2010).
[2] L.A. Peronace, J. Boivin, L. Schmidt, Patterns of suffering and social interactions in infertile men: 12 months after unsuccessful treatment, J. Psychosom. Obstet. Gynecol. 28 (2) (2007) 105–114.
[3] L. Schmidt, Psychosocial consequences of infertility and treatment, Reproductive Endocrinology and Infertility, Springer, 2010, pp. 93–100.
[4] G. Onat, N.K. Beji, Marital relationship and quality of life among couples with infertility, Sex. Disabil. 30 (1) (2012) 39–52.
[5] N.J. Wiersema, A.J. Drukker, M.B.T. Dung, G.H. Nhu, N.T. Nhu, C.B. Lambalk, Consequences of infertility in developing countries: results of a questionnaire and interview survey in the south of vietnam, J. Transl. Med. 4 (1) (2006) 54.
[6] A. Agarwal, A. Mulgund, A. Hamada, M.R. Chyatte, A unique view on male infertility around the globe, Reprod. Biol. Endocrinol. 13 (1) (2015) 37.
[7] P. C. of the American Society for Reproductive Medicine, et al., Diagnostic evaluation of the infertile male: a committee opinion, Fertil. Steril. 103 (3) (2015) e18–e25.
[8] W. H. Organization, et al., Who Laboratory Manual for the Examination and Processing of Human Semen.
[9] D. Franken, How accurate is sperm morphology as an indicator of sperm function? Andrologia 47 (6) (2015) 720–723.
[10] D.F. KATZ, J.W. OVERSTREET, S.J. SAMUELS, P.W. NISWANDER, T.D. BLOOM, E.L. LEWIS, Morphometric analysis of spermatozoa in the assessment of human male fertility, J. Androl. 7 (4) (1986) 203–210.
[11] C. Brazil, Practical semen analysis: from a to z, Asian J. Androl. 12 (1) (2010) 14.
[12] G. Barroso, R. Mercan, K. Ozgur, M. Morshedi, P. Kolm, K. Coetzee, T. Kruger, S. Oehninger, Intra-and inter-laboratory variability in the assessment of sperm morphology by strict criteria: impact of semen preparation, staining techniques and manual versus computerized analysis, Hum. Reprod. 14 (8) (1999) 2036–2040.
[13] R. Menkveld, C.A. Holleboom, J.P. Rhemrev, Measurement and significance of sperm morphology, Asian J. Androl. 13 (1) (2011) 59.
[14] V. Chang, J.M. Saavedra, V. Castãneda, L. Sarabia, N. Hitschfeld, S. H¨artel, Gold-standard and improved framework for sperm head segmentation, Comput. Methods Progr. Biomed. 117 (2) (2014) 225–237.
[15] F. Shaker, S.A. Monadjemi, A.R. Naghsh-Nilchi, Automatic detection and segmentation of sperm head, acrosome and nucleus in microscopic images of human semen smears, Comput. Methods Progr. Biomed. 132 (2016) 11–20.
[16] M.S. Nissen, O. Krause, K. Almstrup, S. Kjærulff, T.T. Nielsen, M. Nielsen, Convolutional neural networks for segmentation and object detection of human semen, Scandinavian Conference on Image Analysis, Springer, 2017, pp. 397–406.
[17] A. v. Oppenheim, R. Schafer, T. Stockham, Nonlinear filtering of multiplied and convolved signals, IEEE Trans. Audio Electroacoust. 16 (3) (1968) 437–466.
[18] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[19] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath, et al., Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups, IEEE Signal Process. Mag. 29 (6) (2012) 82–97.
[20] R. Collobert, J. Weston, A unified architecture for natural language processing: deep neural networks with multitask learning, Proceedings of the 25th International Conference on Machine Learning, ACM, 2008, pp. 160–167.
[21] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, A. Sethi, A dataset and a technique for generalized nuclear segmentation for computational pathology, IEEE Trans. Med. Imaging 36 (7) (2017) 1550–1560.
[22] L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Trans. Med. Imaging 36 (4) (2017) 994–1004.
[23] A.P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, N. Padoy, Endonet: a deep architecture for recognition tasks on laparoscopic videos, IEEE Trans. Med. Imaging 36 (1) (2017) 86–97.
[24] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[25] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Back-propagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551.
[26] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.
[27] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 807–814.
[28] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, Proc. icml, 30 2013, p. 3.
[29] R.M. Haralick, K. Shanmugam, I. Dinstein, et al., Textural features for image classification, IEEE Trans. Syst. Man Cybern. 3 (6) (1973) 610–621.
[30] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic Differentiation in Pytorch.
[31] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: open source scientific tools for Python, [Online;accessed < today >] (2001-). URL http://www.scipy.org/.
[32] P.D. Team, PyTables: hierarchical datasets in Python (2002-), URL http://www.pytables.org/.
[33] Z. Yan, X. Yang, K.-T. Cheng, Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 65 (9) (2018) 1912–1923.
[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,

V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[35] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[36] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.
[37] O. Ronneberger, P. Fischer, T. Brox, U-net, Convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.