Automatic Assessment of Human Sperm Images with Capsule Networks

Mert Şen[1]* iD **, Hatice Doğan[1]** iD

[1] Department of Electrical and Electronics Engineering, Dokuz Eylul University, Turkey

Author E-mails
mertsen.deu@gmail.com
*Correspondence to: Mert Şen, Department of Electrical and Electronics Engineering, Dokuz Eylul University, Turkey

Abstract

Infertility which is a psychologically threatening and emotionally stressful problem is seen approximately 15% of couples in worldwide. Recent studies have shown that in 40-50% of couples evaluated for infertility, the problem is caused by the male individual. Sperm morphology analysis that provides separation of normal and abnormal sperm is very important in evaluating male infertility and showing the causes. Since manual evaluation of sperm morphology is time consuming and subjective, automatic assessment methods are needed. In this study, Capsule Networks, a special model of Deep Neural Networks (DNN), are used for the classification of human sperm head images. The classification performances of capsule networks are measured using the Modified Human Sperm Morphology Analysis dataset (MHSMA). The results show that the best classification accuracy is achieved as 73%.

Keywords: Capsule networks, infertility, sperm morphology.

1. INTRODUCTION

Sperm morphology analysis, which refers to the analysis of the size and shape of a sperm, is a very important tool in the diagnosis of male infertility. Problems in the maturation phase of the sperm cause abnormalities in sperm morphology and egg fertilization function [6]. In sperm morphology analysis, sperms with normal and abnormal morphology are detected by examining semen samples under a microscope. Normal sperm has a long tail and an oval head as shown in Figure 1. Sperm with abnormal

morphology may have tail defects such as a cracked or double tail, or head defects such as a large or misshapen head [6-8].
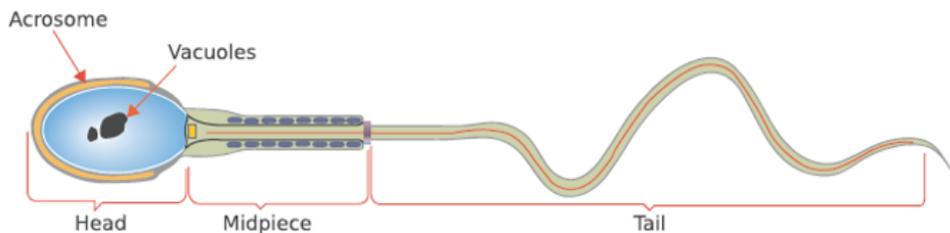
Figure 1. Parts of a human sperm cell [6].

Since manual evaluation of sperm morphology is time-consuming and subjective, different techniques have been developed in the literature for the automatic assessment of human sperm cells. In [1], both normal and abnormal sperm images are converted into grayscale images and noises are removed using filters. Then, edge detection and segmentation algorithms are performed to segment the sperm images into different parts and classify them with statistical measurement methods. In [3], sperm images are segmented using Bayesian classifier and entropy based expectation and maximization algorithm. In [2], digital image processing and learning vector quantization is used to classify the acrosome part of the sperm cell. Chang et al. utilizes statistical histogram analysis and clustering methods in different color ranges for the segmentation of acrosome and nucleus of the sperm cell [4]. In [7], principal component analysis is used to extract the features from the sperm images while K-nearest neighbors technique is used to determine whether the sperm cell is normal or not.

In recent studies, deep learning methods in which features are automatically extracted are used for the sperm assessment [6]. The best known of the deep learning methods are the convolutional neural networks (CNN). Although CNN methods have been proven to perform better than traditional methods in machine learning problems, CNNs have disadvantages, such as inability to tolerate perspective variations and losing spatial information between features. To overcome these drawbacks, Capsule Networks have been proposed by Hinton. In this study, the classification performance of capsule networks are evaluated on the Modified Human Sperm Morphology Analysis (MHSMA) dataset whose images were captured by a microscope occupied with a CCD camera. MHSMA set has several challenging points: the number of total sperm images is not sufficient for training phase, hence image data augmentation techniques must be used, the dataset is imbalanced and images of the sperm cells are unclear and too noisy.

The rest of the paper is organized as follows. In Section 2, capsule networks and MHSMA dataset are introduced, In Section 3, results are presented and in Section 4 the conclusions are given.

2. MATERIALS AND METHODS

2.1. Capsule Networks

Capsule networks have been proposed to overcome the disadvantages of the convolutional neural networks. A capsule is a set of neurons with activity vector that contains the information about an object. The length of the activity vector of a capsule represents the probability of the object's existence while the direction of the vector represents the instantiation parameters such as pose, velocity, and albedo, etc. [10]. The activity vectors of each capsule are marked in red and blue in Figure 2.
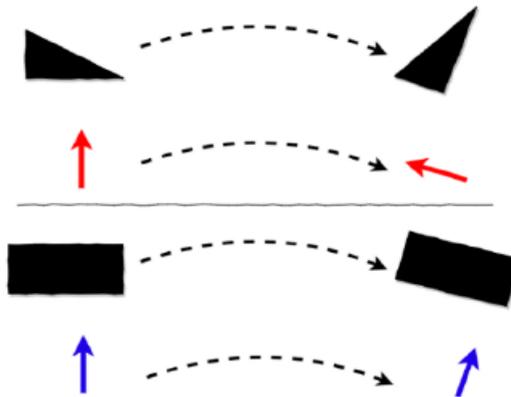
Figure 2. Representation of the activity vector of capsule networks [10].

The differences between a capsule and a traditional neuron are shown in Table 1. One of the crucial differences is that capsule network performs affine transformation before the weighted sum [9].

Table 1. Comparison of a capsule to a traditional neuron [9].

Operation Type	Capsule	Traditional Neuron
Input Shape	$vector(\mathbf{u}_i)$	$scalar(x_i)$
Affine Transform	$\hat{\mathbf{u}}_{j\|i} = \mathbf{W}_{ij}\mathbf{u}_i$	-
Weighting & Sum	$\mathbf{s}_j = \sum_i c_{ij}\hat{\mathbf{u}}_{j\|i}$	$a_j = \sum_i w_i x_i$
Non-linear Activation	$\mathbf{v}_j = \frac{\|\|\mathbf{s}_j\|\|^2}{1+\|\|\mathbf{s}_j\|\|^2}\frac{\mathbf{s}_j}{\|\|\mathbf{s}_j\|\|}$	$h_j = f(a_j)$
Output Shape	$vector(\mathbf{v}_j)$	$scalar(h_j)$

In affine transform step, the vectors **u₁**, **u₂** and **u₃** coming from a lower level layer are considered as input vectors. These vectors are multiplied by a weight matrix denoted as **W** in order to store important spatial and other relationships between lower level features and higher level features. In fact, the estimated position of the high level feature is found [9].

In the weighting and sum step, the lower level capsules send their outputs to the higher level capsules. The critical point is to determine which input of the higher level capsule is connected to the lower level capsule. The weights c_1, c_2 and c_3 determined by using routing algorithm are used for the connection between capsules. Summing step is the same as in a traditional neuron. Namely, it represents the combination of vectors [9].

In squashing (non-linear activation) step, the non-linear activation function takes an input as a vector and squashes it without changing its direction. With this procedure, it is guaranteed that the vectors have a length between 0 and 1 [9].

2.2. Routing-by-agreement Technique

The routing by agreement algorithm described by Hinton is given in Figure 3. The operation of each step in the algorithm can be examined step by step [9].

Step #1 : Algorithm takes the number of routing iteration (r), number of lower layer (l) and the outputs of all capsules in that layer ($\hat{u}$).

Step #2 : At first, b_{ij} is set to 0. The temporary coefficient b_{ij} is iteratively updated until the procedure is completed. Then, it is finally stored in c_{ij}.

Step #3 : Repeat the dynamic routing algorithm with respect to the iteration number.

```
1: procedure ROUTING(û_{j|i}, r, l)
2:     for all capsule i in layer l and capsule j in layer (l + 1): b_{ij} ← 0.
3:     for r iterations do
4:         for all capsule i in layer l: c_i ← softmax(b_i)
5:         for all capsule j in layer (l + 1): s_j ← ∑_i c_{ij} û_{j|i}
6:         for all capsule j in layer (l + 1): v_j ← squash(s_j)
7:         for all capsule i in layer l and capsule j in layer (l + 1): b_{ij} ← b_{ij} + û_{j|i}.v_j
       return v_j
```

Figure 3. Routing algorithm [11].

Step #4 : Compute the values of the vector c_i for all lower level capsules. All coefficients c_{ij} are equal in the first iteration since $b_{ij} = 0$. It means that algorithm shows *maximum confusion and uncertainty.* In other words, the appropriate output of higher level capsule can't be determined by lower level capsule.

Step #5 : After all computations of c_{ij} for lower layer capsules, a linear combination of input vectors that are weighted by c_{ij} is computed for higher level capsules. Then s_j is calculated.

Step #6 : The vector computed in previous step is squashed for non-linearity. The direction of the vector is not changed and its length is scaled. Then, output vector $\mathbf{v_j}$ is produced for all higher level capsules.

Step #7 : This step examines each input and updates the corresponding weight b_{ij} for each higher level capsule j. The similarity between inputs and outputs of the capsule is evaluated by the *dot product*.

2.3. MHSMA Dataset

The MHSMA dataset includes 1540 grayscale human sperm cell images with dimension of 128 x 128 pixels and 64 x 64 pixels. Each image represents a single sperm and sperm is always centered in the image. Data partition is performed as: Randomly-selected 1000 images as training set, 300 images as test set and the remaining 240 images as validation set. All data is shuffled. Each part of the sperm cell (acrosome, vacuole, head and tail-neck) can be examined as a different class. The image distribution of the dataset is given in Table 2 [6]. In this study, *head class* is selected in order to investigate the male infertility problem with capsule networks. Image data augmentation methods such as flipping, shifting, rotating and changing brightness are performed to avoid class imbalance. In Figure 4, several examples of sperm head images in the dataset can be seen.

Table 2. Sample distribution of the dataset

Set	Label	# Positive	# Negative	% Positive
	Acrosome	1,086	454	70.52
	Head	1,122	418	72.86
	Vacuole	1,301	239	84.48
Whole dataset	Tail and neck	1,471	69	95.52
	Acrosome	699	301	69.90
	Head	727	273	72.70
	Vacuole	830	170	83.00
Training set	Tail and neck	954	46	95.40
	Acrosome	174	66	72.50
	Head	176	64	73.33
	Vacuole	209	31	87.08
Validation set	Tail and neck	233	7	97.08
	Acrosome	213	87	71.00
	Head	219	81	73.00
	Vacuole	262	38	87.33
Test set	Tail and neck	284	16	94.67
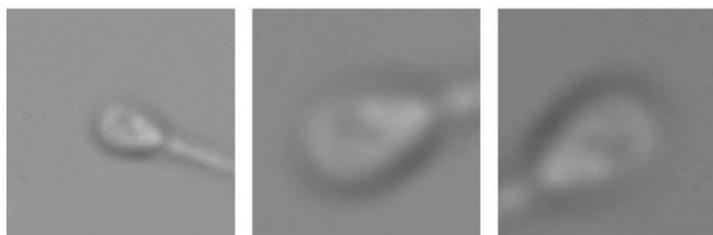
Figure 4. Instances of sperm head images [6].

3. RESULTS AND DISCUSSION

Capsule network used in the study has a convolutional layer with resized 32 x 32 input size for grayscale sperm head images. In this structure, the convolution layer used for feature extraction has 256 filters with stride of 1 and output shape of the convolutional layer is one dimensional. Due to 1D output, there is no orientation. PrimaryCaps layer is another convolution layer that applies squashing. The Caps layer is used for classification and total loss is the sum of individual capsule loss. The critical point is that the routing algorithm is performed between PrimaryCaps and Caps layers. Different simulations are performed on MHSMA dataset. In order to discover the effect of model modifications, the classification performances of capsule networks with different number of convolution layers and primary capsule dimension are evaluated. The results are given in Table 3. All simulations are done with CapsNet software at Google Collaboratory providing Tesla K80 GPU [5].

Table 3. Performance comparison of different models of Capsule Networks on MHSMA.

Class Type	Image Size	Data Augmentation	# of Conv Layer	Dim of Capsule	Test Accuracy (%)
Head	32 × 32	Yes	1	8	73
Head	32 × 32	Yes	2	8	72.32
Head	32 × 32	Yes	2	16	72.34
Head	32 × 32	Yes	1	16	72.67

According to Table 3, capsule networks that have different convolution and primary capsule layers demonstrate similar performances on MHSMA dataset. The capsule network that has one convolutional layer and 8 Dimensional capsule shows the best assessment performance for the evaluation of the male infertility.

The input-output connection of the capsule network that has the best performance on MHSMA dataset can be seen in Table 4.

Table 4. The input-output connection of Capsule Network.

```
_____
Layer (type)                   Output Shape            Param #     Connected to
======================================================================================
input_1 (InputLayer)           (None, 32, 32, 1)        0
_____
conv1 (Conv2D)                 (None, 24, 24, 256)      20992       input_1[0][0]
_____
primarycap_conv2d (Conv2D)     (None, 8, 8, 256)        5308672     conv1[0][0]
_____
primarycap_reshape (Reshape)   (None, 2048, 8)          0           primarycap_conv2d[0][0]
_____
primarycap_squash (Lambda)     (None, 2048, 8)          0           primarycap_reshape[0][0]
_____
caps (CapsuleLayer)            (None, 2, 16)            524288      primarycap_squash[0][0]
_____
input_2 (InputLayer)           (None, 2)                0
_____
mask_1 (Mask)                  (None, 32)               0           caps[0][0]
                                                                    input_2[0][0]
_____
capsnet (Length)               (None, 2)                0           digitcaps[0][0]
_____
decoder (Sequential)           (None, 32, 32, 1)        1591808     mask_1[0][0]
======================================================================================
Total params: 7,445,760
Trainable params: 7,445,760
Non-trainable params: 0
_____
```

The training and validation performance of the simulated networks can be seen in Figure 5.
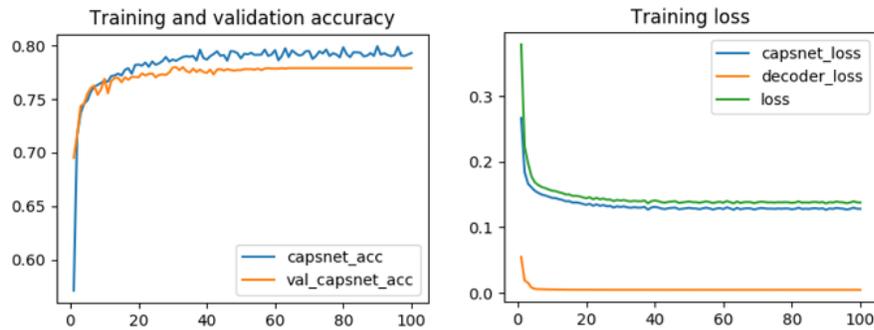
Figure 5. Simulation graphs.

As seen in Figure 5, training and validation accuracy of the network can't exceed 80% due to the similarity of sperm images. The network classifies abnormal sperm with 85% accuracy, but the classification accuracy of normal sperm is less than this ratio.

4. CONCLUSION

In this study, the classification performance of capsule networks is evaluated on a data set containing grayscale human sperm head images and the best classification performance is found as 73%. According to the simulation results, although the performance of the capsule network in the MHSMA dataset is promising, the capsule networks do not exceed the performance of CNN, which was 76.88%. This may be because sperm head images are similar and capsule networks have difficulty distinguishing two objects with similar characteristics.

REFERENCES

1. Abbirany, V., V. Shanti. "Spermatoza segmentation and morphological parameter analysis based detection of teratozoospermia". *Int. J. Comput. Appl.* 3(2010): 19-23.
2. Alegre, E., M. Biehl, N. Petkov, L. Sanchez. "Automatic classification of the acrosome status of boar spermatozoa using digital image processing and LVQ". *Comput. Biol. Med.* 38(2008): 461-468.

3. Bijar, A., A. P Benavent, M. Mikaeili, R. Khayati. "Fully automatic identification and discrimination of sperm's parts in microscopic images of stained human semen smear". *J. Biomed. Sci. Eng.* 5(2012): 384.

4. Chang, V., M. Saavedra, V. Castañeda, L. Sarabia, N. Hitschfeld, S. Härtel. "Gold-standard and improved framework for sperm head segmentation". *Comput. Methods Progr. Biomed.* 117(2014): 225-237.

5. Google. "CapsNet-Keras". Last update 5 August 2009. https:// github.com/XifengGuo/ CapsNet-Keras

6. Javadi, S., S. A. Mirroshandel. "A novel deep learning method for automatic assessment of human sperm images". *Computers in Biology and Medicine* 109(2019): 182-194.

7. Li, J., K. K. Tseng, H. Dong, Y. Li, M. Zhao, M. Ding. "Human sperm health diagnosis with principal component analysis and k-nearest neighbor algorithm". *Medical Biometrics, 2014 International Conference on IEEE*, (2014): 108-113.

8. R. Menkveld, C. A. Holleboom, , J. P. Rhemrev. "Measurement and significance of sperm morphology". *Asian J. Androl.* 13(2011): 59.

9. Google. "Understanding Hinton's capsule networks. part I: intuition". Last update 22 December 2019. https://medium.com/ai%C2%B3-theory-practice-business/under_standing-hintons-capsule-networks-part-i-intuition-b4b559d1159b

10. Google. "Understanding capsule network architecture". Last update 15 December 2019. https://software.intel.com/en-us/articles/understanding-capsule-network-architecture

11. Sabour, S., N. Frosst, G. Hinton. "Dynamic routing between capsules". *31st Conference on Neural Information Processing Systems (NIPS),* (2017): 1-11.