

# Creating CNN Models For Sperm Morphology Analysis

Abdurrahman Ebrar Yücel  
Bilgisayar Mühendisliği Bölümü  
Yıldız Teknik Üniversitesi, 34220 İstanbul, Türkiye  
a.ebraryucell@gmail.com

**Özetçe** —Tıbbi görüntü analizi,teknolojinin gelişmesiyle sağlık alanında gittikçe kullanışlı bir hale gelmektedir. Evrışimsel sinir ağları bu amaçla kullanılan yöntemlerden biridir. Sperm morfolojisinin analizi tıbbi görüntü analizinin güncel konularından biridir. Kısırlık tanısı konan çiftler için son derece önemli bir konu olan sperm morfolojisi, kısırlığa neden olan unsurların araştırılması sırasında sperm analizinin yapılmasına ihtiyaç duyar.Bu analiz sırasında spermlerde yapı bozukluğu incelenir.Özellikle spermin başının boyutu ve şekli analizde çok etkilidir. Elde edilen sonuçlar doktorların tedavi ve teşhis seçiminde yüksek önem taşır. Bu projede HuSHeM,SCIAN-Morpho,SMIDS verisetleri kullanılarak sperm görüntülerinin sınıflandırılması için evrışimsel sinir ağı modellerinin oluşturulması amaçlanmaktadır.

**Anahtar Kelimeler**—Evrışimli sinir ağları,sperm morfolojisi,kısırlık,görüntü analizi.

**Abstract**—Medical image analysis is becoming more and more useful in the field of health with the development of technology. Convolutional neural networks are one of the methods used for this purpose. Analysis of sperm morphology is one of the current topics of medical image analysis. Sperm morphology, which is an extremely important issue for couples diagnosed with infertility, requires sperm analysis while investigating the factors that cause infertility. During this analysis, the structural defects in the sperm are examined. Especially the size and shape of the head of the sperm are very effective in the analysis. The results obtained are of great importance in the choice of treatment and diagnosis by doctors. In this project, it is aimed to create convolutional neural network models for classifying sperm images using HuSHeM, SCIAN-Morpho, SMIDS datasets.

**Keywords**—Convolutional neural networks,sperm morphology,infertility,image analysis.

## I. INTRODUCTION

Sperm, in short, is the reproductive cell of male individuals. When the male individuals enter puberty, its production begins in the testicles. The sperm cell is approximately one 250th of a centimeter long and consists of four parts: head, neck, body, and tail. The head part allows the sperm cell to unite with the egg cell, the body part contains the genetic material of the person, the tail part of it provides the mobility of the sperm cell.

Infertility is a health problem that disrupts the function of the reproductive system. About 10 percent of every 100 couples of reproductive age suffer from infertility. 30 percent of this problem is related to men. Infertility can be a cause or have many reasons. The main reasons

are decreased sperm production and sperm motility, duct obstructions, infections, abnormal sperm function, genetic problems, sperm disorder of morphology, alcohol, smoking etc. The main diagnostic method of infertility in men is semen analysis (spermiogram). The analysis is made by examining the sperm count, mobility and structure of sperm samples taken from individuals. One of the most important parts of this examination is to examine the morphology of sperm.

When examining sperm morphology, successful results can be obtained by using convolutional neural networks, which can nowadays be highly successful in analyzing visual information. Convolutional neural networks are one of the sub-branches of deep learning and are now widely used in image classification, medical image analysis, natural language processing and many other fields. Convolutional neural networks underlie the convolution process, so they are neural networks that do convolution in at least one of their layers. Layer types used in convolutional neural networks include; input layer, convolution layer, activation layer, pooling layer, full link layer, normalization layers, regularization layers, classification layer. These layers come together in a specific architecture and form a convolutional neural network model. The aim of this study is to successfully classify data in HuSHeM, SMIDS, SCIAN datasets and to observe the effects of data augmentation on the results of the models created. However, certain difficulties are encountered while creating models. Some of these difficulties are; similar sperm classes, noisy low-sized sperm images, low-sized sperm images, lack of enough data to train complex models,uneven distribution of training data of sperm classes.

Today, many studies have been published to classify sperm. Various methods and various datasets have been used in these studies.In 2017 Chang et al. prepared first free and available gold standard dataset, "SCIAN-MorphoSpermGS",which analyzes and evaluates morphological classification of sperm heads.This dataset uses the categorization of sperm heads for semen analysis in the WHO laboratory manual:"Normal","Tapered","Pyriform","Small","Amorphous". Sperm head images in the dataset were identified by 3 Chilean experts.Chang et al. also suggests CE-SVM which is a two phase analysis pipeline for the morphological classification of sperm heads in the dataset[5].

On the other hand, Shaker et al. produced the HuSHeM (Human Sperm Head Morphology) dataset with an adaptive dictionary learning based approach. In this approach, the dictionaries are trained to detect those sperm head categories using certain square patches from the sperm head images [7].

In another study Riordon et al. developed another method, a VGG-16 architecture for the classification. In this method, VGG was trained on ImageNet, then fine-tuned on SCIAN dataset. The results show that this deep learning method may contribute the seminogram in an effective way [6].

In another study Ilhan et al. applied several preprocessing steps to eliminate the noisy parts in the image patches. After that, they used SURF and MSER descriptor features to feed a non-linear classification schema, SVM with polynomial kernel. In this way, they increased their classification accuracies in SMIDS dataset from 84 percent to 86 percent [8].

## II. MATERIALS AND METHODS

In this study, Matlab Deep Learning Toolbox was used and it was done on Windows 10 platform. Deep Learning Toolbox provides a framework for designing and implementing deep neural networks with algorithms, pretrained models, and apps. The features of the hardware environment on which the study is performed are as follows: Intel® Core™ i7-4720HQ @ 2.6GHz, NVIDIA GeForce GTX960M with 4GB GDDR5 memory.

### A. Dataset Information

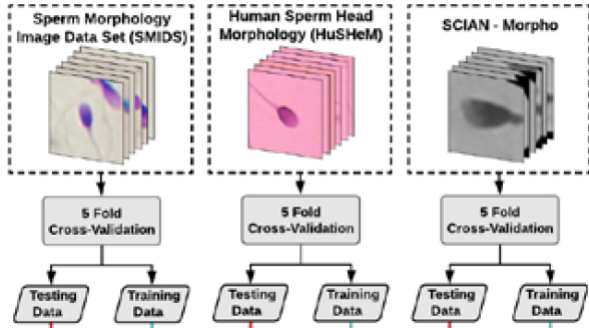


Figure 1 Dataset Structure

While training the created convolutional neural network models, 3 different datasets will be used. These are HuSHeM (Human Sperm Head Morphology), SMIDS (Sperm Morphology Image Data Set), SCIAN-Morpho datasets. Datasets consist of a combination of original datasets and augmented datasets. Each set is divided into 5 for 5-fold cross validation, each folder consists of test and training data. Dataset folder structure is shown in figure 1. The SCIAN dataset consists of pictures of approximately 35x35 sizes. It contains 1132 images from 5 different classes, the classes are: "Normal", "Tapered", "Pyriform", "Amorphous", "Small". There is an unbalanced distribution

of images between classes in the SCIAN dataset. The number of data belonging to the "Amorphous" class is more than other classes. The HuSHeM dataset consists of images of approximately 130x130 sizes. It contains 216 images from 4 different classes, the classes are as follows: "Normal", "Tapered", "Pyriform", "Amorphous". The data are evenly distributed among the classes. SMIDS dataset consists of images ranging from 150-200 pixels. It contains 3000 images from 3 different classes, the classes are: "Normal", "Abnormal", "Non-Sperm". The data are evenly distributed among the classes. The distribution of data in the datasets between classes is shown in figure 2. In order to observe the effects of data augmentation on the results of models, scaling, scrolling, rotating, clipping, etc. were applied on original datasets. It is aimed to increase the performance by applying a certain amount of data augmentation. In the HuSHeM dataset 1x (original), 5x, 10x augmented; in SCIAN dataset 1x, 5x, 9x augmented and balanced; in SMIDS dataset 1x, 5x, 8x augmented data were used for training. For these datasets, different epoch numbers between 10-100 were tested in order to observe the effect of epoch number on performance increase in models and performance and training times have been compared.

| SMIDS     |      | HuSHeM    |     | SCIAN-Morpho |      |
|-----------|------|-----------|-----|--------------|------|
| Labels    | #    | Labels    | #   | Labels       | #    |
| Normal    | 1021 | Normal    | 54  | Normal       | 100  |
| Abnormal  | 1005 | Tapered   | 53  | Tapered      | 228  |
| Non-Sperm | 974  | Pyriform  | 57  | Pyriform     | 76   |
|           |      | Amorphous | 52  | Small        | 72   |
|           |      |           |     | Amorphous    | 656  |
| Total     | 3000 | Total     | 216 | Total        | 1132 |

Figure 2 Distribution of Images

### B. Proposed CNN Architectures

In this study, 2 models were created with different types of network structures in order to achieve high classification performance. In the first model created, model receives data in the size of 131x131x3. The model is based on the VGG-16 architecture. The relu function, which is frequently used in deep learning and reduces the calculation cost of the model, was used as the activation function. All of the convolution layers consist of 3x3 size filters since local features are important in our problem. Compared to VGG models, the number of filters is significantly reduced due to the prevention of over fitting, reduction of computational cost and classification task. The model includes 30 layers in total. 10 convolution layers and 4 maximum pooling layers are used.

The second model created is based on the GoogLeNet architecture. The Model receives data in the size of 150x150x3, includes 5 inception modules and 87 layers in total. Relu is used as the activation function in the model. 3x3 filters are mostly used in convolution and maximum pooling layers as in the first model. This time in order to increase the computation speed of this model consisting of 87 layers, the stride value in the maximum pooling layers outside inception modules was chosen as 3. In addition, the number of filters and padding values in the inception

modules were adjusted to suit the classification task and avoid over fitting.

If the hyperparameters of the models are examined; 10,30,50,60,75,100 epochs were tested in both models. The Epoch number had to be chosen to prevent the models from under fitting in all 3 datasets, so it was decided to be 50 for the first model, which is a smaller model, and 75 for the second model as a result of epoch tests. Learning rate was chosen as 0.0001 for the first model and 0.01 for the second model as a result of systematic trials. The Stochastic Gradient Descent with Momentum algorithm was used in the first model and the Adaptive Momentum algorithm was used in the second model as an optimizer. In the mini-batch size selection, relatively small value 16 was chosen for both models due to slightly negative effect of high mini-batch sizes on the generalization capability of the models. In addition, the hardware environment is insufficient for high mini-batch size. In order not to be affected by the distribution of datasets, shuffling the datasets in every epoch was performed in the training of both models. Hyperparameters of both models showed in figure 3.

|                 | Network 1 | Network 2 |
|-----------------|-----------|-----------|
| Epoch Num.      | 50        | 75        |
| Learning Rate   | 0,0001    | 0,01      |
| Optimizer       | SGDM      | ADAM      |
| Mini Batch Size | 16        | 16        |

**Figure 3** Hyperparameters of Models

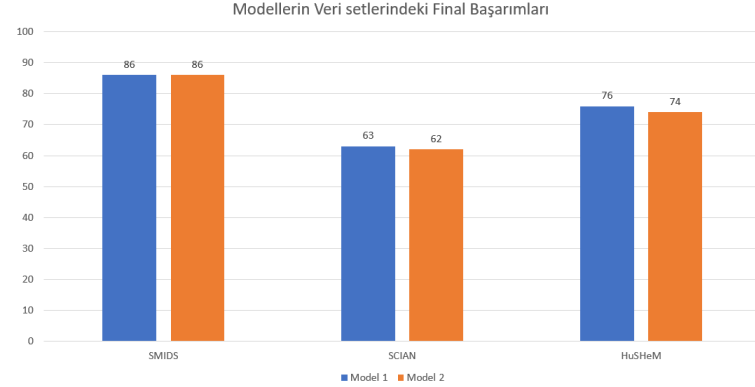
### III. EXPERIMENTAL RESULTS

By testing the models on the original data and the augmented data, the effects of data augmentation on the accuracy of the models and training times were measured. Data augmentation test results showed in figure 4. Average training times for each fold showed in figure 6. Final results of models showed in figure 5.

As a result of the data augmentation tests, in the HuSHeM dataset, the success of data augmentation operations on both models is approximately 10-15 percent due to the low number of samples. However, in SMIDS and SCIAN datasets, success of data augmentation is about 3 percent because of sufficient number of samples. Since the accuracy may vary in the training of the same dataset, the effects of the data augmentation may also vary similarly. When the results are examined on a model basis, networks give similar responses to data augmentation operations. It was observed that the differences between the accuracies were low. With the epoch values determined for the models, the training times of the models were measured with the augmented datasets and the original datasets. Due to the low number of data in HuSHeM dataset, the required number of epochs for networks to learn, It was determined that there is more in the HuSHeM dataset than the other datasets. Thanks to the sufficiency of the number of samples in SCIAN and SMIDS datasets, less epoch values can easily converges higher accuracies.

|           | Network1 |       |       | Network2 |       |       |
|-----------|----------|-------|-------|----------|-------|-------|
| Aug. Rate | HuSHeM   | SMIDS | SCIAN | HuSHeM   | SMIDS | SCIAN |
| 1x        | 65       | 83    | 62    | 58       | 82    | 60    |
| 5x        | 73       | 86    | 61    | 72       | 85    | 62    |
| 8x        | -        | 86    | -     | -        | 86    | -     |
| 9x        | -        | -     | 64    | -        | -     | 63    |
| 10x       | 76       | -     | -     | 74       | -     | -     |

**Figure 4** Data Augmentation Test Results



**Figure 5** Final Test Results

|           | Network 1 |            |             | Network 2 |             |             |
|-----------|-----------|------------|-------------|-----------|-------------|-------------|
| Aug. Rate | HuSHeM    | SMIDS      | SCIAN       | HuSHeM    | SMIDS       | SCIAN       |
| 1x        | 2 min     | 26 min     | 9 min       | 4 min     | 1hr 16 min  | 21 min      |
| 5x        | 9 min     | 2 hr 5 min | 2 hr 21 min | 19 min    | 5 hr 19 min | 4 hr 47 min |
| 8x        | -         | 4 hr 1 min | -           | -         | 8 hr 13 min | -           |
| 9x        | -         | -          | 4 hr 13 min | -         | -           | 8 hr 12 min |
| 10x       | 17 min    | -          | -           | 39 min    | -           | -           |

**Figure 6** Average Training Times For Each Folds

### IV. CONCLUSION

In this study, convolutional neural network models were created in two different architectures. Their accuracies and training times were measured in SCIAN, SMIDS, HuSHeM datasets. It is aimed to prevent models from memorizing their datasets and not being able to learn due to insufficient data by subjecting the datasets to data augmentation at different rates. The effects of data augmentation on the performance of the models and training times were observed. As a result, data augmentation have the same effect on performance in both models. The low number of samples in the HuSHeM dataset increased the effect of data augmentation and it has been found to increase the performance about 10 percent more than other datasets.

The first model was found to be slightly more successful than the second model, especially in tests with the original data. At the same time, the first model has a lower computational cost and training time.

## REFERENCES

- 1- Imran Iqbal 1 , Ghulam Mustafa 2 and Jinwen Ma 1 ;Deep Learning-Based Morphological Classification of Human Sperm Heads;20 May 2020
- 2- Özkan İNİK,\* Erkan ÜLKER;Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri;GAZİOSMANPAŞA BİLİMSEL ARAŞTIRMA DERGİSİ (GBAD) Gaziosmanpasa Journal of Scientific Research
- 3- Omer Lutfu Tortumlu, Hamza Osman Ilhan;The Analysis of Mobile Platform based CNN Networks in the Classification of Sperm Morphology;2020
- 4- Evrişimli Sinir Ağları el kitabı; <https://stanford.edu/shervine/l/tr/teaching/cs-230/cheatsheet-convolutional-neural-networks>
- 5- Chang, V.; Garcia, A.; Hitschfeld, N.; Hartel, S. Gold-standard for computer-assisted morphological sperm analysis. Comput. Biol. Med. 2017, 83, 143–150.
- 6- Riordon, J.; Mccallum, C.; Sinton, D. Deep learning for the classification of human sperm. Comput. Biol. Med. 2019, 111.
- 7- Shaker, F.; Monadjemi, S.A.; Alirezaie, J.; Naghsh-Nilchi, A.R. A dictionary learning approach for human sperm heads classification. Comput. Biol. Med. 2017, 91, 181–190.
- 8- H. O. Ilhan, G. Serbes, and N. Aydin, “Automated sperm morphology analysis approach using a directional masking technique,” Computers in Biology and Medicine, p. 103845, 2020.