

# Journal Pre-proof

Automated sperm morphology analysis approach using a directional masking technique

Hamza Osman Ilhan, Gorkem Serbes, Nizamettin Aydin



PII: S0010-4825(20)30204-3

DOI: <https://doi.org/10.1016/j.combiomed.2020.103845>

Reference: CBM 103845

To appear in: *Computers in Biology and Medicine*

Received date: 9 April 2020

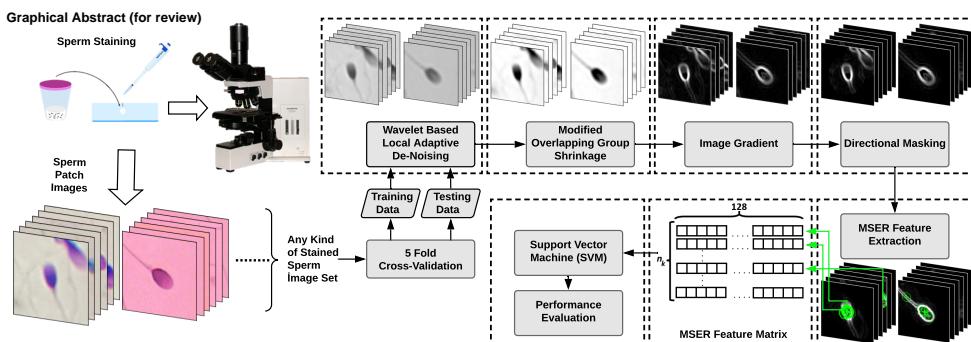
Revised date: 3 June 2020

Accepted date: 3 June 2020

Please cite this article as: H.O. Ilhan, G. Serbes and N. Aydin, Automated sperm morphology analysis approach using a directional masking technique, *Computers in Biology and Medicine* (2020), doi: <https://doi.org/10.1016/j.combiomed.2020.103845>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.



**Highlights (for review)**

- In this study, a computation framework including multi-stage cascade connected preprocessing techniques, region based descriptor features and non-linear kernel SVM based learning is proposed for the classification of any stained sperm images in terms of morphology assessment.
- Wavelet based local adaptive de-noising, modified overlapping group shrinkage, image gradient and automatic directional masking techniques were utilized in the multi-stage preprocessing step of the proposed framework.
- Two descriptors, Speed Up Robust Features (SURF) and Maximally Stable Extremal Regions (MSER), were tested and compared in order to define the most informative feature extraction technique.
- The performance of framework was evaluated on two sperm morphology datasets, Human Sperm Head Morphology dataset (HuSHeM) and Sperm Morphology Image Data Set (SMIDS).
- In the light of obtained results, the proposed framework increased the classification accuracy by 10% and 5% for the HuSHeM and SMIDS, respectively.

# Automated Sperm Morphology Analysis Approach using a Directional Masking Technique

Hamza Osman Ilhan<sup>a,1,\*</sup>, Gorkem Serbes<sup>b,1</sup>, Nizamettin Aydin<sup>a</sup>

<sup>a</sup>*Department of Computer Engineering, Yildiz Technical University, Turkey*

<sup>b</sup>*Department of Biomedical Engineering, Yildiz Technical University, Turkey*

## Abstract

Sperm Morphology is the key step in the assessment of sperm quality. Due to the effect of misleading human factors in manual assessments, computer-based techniques should be employed in the analysis. In this study, a computation framework including multi-stage cascade connected preprocessing techniques, region based descriptor features, and non-linear kernel SVM based learning is proposed for the classification of any stained sperm images for the assessment of the morphology. The proposed framework was evaluated on two sperm morphology datasets: the Human Sperm Head Morphology dataset (HuSHeM) and Sperm Morphology Image Data Set (SMIDS). The results indicate that cascading the preprocessing techniques used in the proposed framework, such as wavelet based local adaptive de-noising, modified overlapping group shrinkage, image gradient, and automatic directional masking, increased the classification accuracy by 10% and 5% for the HuSHeM and SMIDS, respectively. The proposed framework results in better overall accuracy than most state-of-the-art methods, while having significant advantages, such as eliminating the exhaustive manual orientation and cropping operations of the competitors with reasonable rates of consumption of time and source.

**Keywords:** Directional Masking Technique, Sperm Morphology Classification, Descriptor Based Feature Extraction, Wavelet Based Local Adaptive De-Noising, Support Vector Machines, Maximally Stable Extremal Regions

## 1. Introduction

Today, infertility is a disease of global importance. It affects both personal life and social life in the community. 30% of the world's population is faced with this problem, according to the World Health Organization (WHO) [1]. Infertility is defined as the inability to have a child after one year of sexual intercourse without contraception involving a male and female partner [2]. The reasons for infertility can be based on the male, the female, both, or unexpected factors. In male-based infertility, the analysis of sperm morphology has a crucial significance in determining the type of infertility [1, 3, 4].

In clinics, spermogram tests are performed to evaluate the semen specimen. Spermogram tests include a two-step analysis. First, experts evaluate the semen specimen in terms of physical and chemical aspects, such as viscosity, colour, smell, volume, pH, etc. Then, the sample is manually observed under a microscope for concentration, motility and morphological analyses. In this step, the reliability of the diagnosis depends strongly on the expertise of the observer. This problem is known in the literature as observer variability. It can be seen in the different manual observation analyses in diagnostic medicine [5]. Therefore, computer based systems are currently used in the analysis so as to eliminate the human factor. Systems used in spermogram tests are referred to as Computer Aided Sperm Analysis (CASA) systems. However, these systems are expensive, hard to implement in laboratories, require an exhausting parameter tuning process, and observer education [6]. Additionally, the re-

\*Corresponding Author

Email address: hoilhan@yildiz.edu.tr (Hamza Osman Ilhan)

<sup>1</sup>H.O. Ilhan and G. Serbes contributed equally to this work.

sults and the consistency of the CASA systems are questionable [7, 8]. The algorithms employed in CASA systems need to be improved in many aspects, especially for the morphological analysis [9]. Additionally, the standard deviations in CASA results have been reported to be higher than for the results of manual assessment [10]. Therefore, more robust solutions tent to replace CASA systems and the use of low cost automatic systems has become a vital requirement in the analysis of sperm morphology. Many studies from different disciplines have been performed on sperm images to design an automated analysing system, especially for morphology analysis [11, 12, 13, 14, 15, 16, 17, 18]. Despite the high classification accuracy achieved in previous studies, there is still a significant need for a fully automated sperm morphology analysis framework in order to overcome the hampering effect of undesired noise components that reduce the discrimination capability of current approaches.

In the morphological analysis of spermiogram, spermatozoa are evaluated for their head shape, size, acrosome defects, tail length, and mid-piece distortion. In this study, samples were stained with a modified hematoxylin/eosin assay by an expert. According to the fifth edition of the WHO manual, the normal sperm is about 3.7–4.7 micrometers long, 2.5–3.2 micrometers wide, and is characterized by an oval head in a single long tail with a smooth acrosome cover [1, 19]. The head piece should be properly connected with the mid-piece via a neck. Those spermatozoa that have sizes outside the reference scale given by the WHO are classified as abnormal and can be further categorized into sub-classes, such as head defects (tapered, pyriform, or amorphous), vacuolization defects, acrosome proportion defects, and tail defects.

To be able to develop morphological analysis algorithms, three sperm morphological datasets have been presented in the literature [20, 15, 11]. Chang et al. created a dataset including 4 subcategories of abnormal sperm: small, amorphous, tapered, and pyriform. They manually cropped patches of each sperm from the non-stained images of the semen specimen and then rotated them so as to lie along one direction, so as to decrease the noise based classification failures. At first, they used morphological operations to segment the object. Next, they performed a blob analysis to extract spatial features, such as areas, eccentricities, major/minor axes, etc. They classified each patch according to these spatial features by rule based classification tech-

niques. They reported a 56% accuracy for the classification [14]. Shaker et al. performed a dictionary learning to classify the HuSHeM dataset into 4 classes: Normal, Pyriform, Amorphous, and Tapered [15]. They also manually rotated all images to lie along one specific direction, similar to [14], before the classification process, which has a great impact on code-book based feature extraction techniques, such as in dictionary learning or any descriptor-based learning technique. Then, they used dictionary learning based features in SVM classifiers. According to the results, they achieved 92.9% accuracy for their dataset (HuSHeM). Additionally, they tested their dictionary learning based approach on the SCIAN-Morpho dataset. They reported a 62% accuracy in the classification of the SCIAN-Morpho dataset. In the previously proposed classification studies, which were applied to these two databases [14, 15], the sperm images were manually oriented to lie in one direction so as to enhance the robustness of the extracted features. However, this manual operation constitutes a human-based interference, which reduces the objectivity of the whole system. Additionally, even though this manual operation increases the classification accuracy for the proposed pre-processed sample-sets, new image samples that are intended to be analysed with these methods will also need manual orientation in real-world scenarios. The requirement of this human-based orientation increases the required analysis time and imposes a cost. In contrast to previous studies, our present paper proposes a fully automated sperm-tail orienting masking technique, resulting in segmented sperm shapes, including both sperm head, mid-piece, and tail information.

In [11], a sperm morphology dataset called the Sperm Morphology Image Data Set (SMIDS), which includes normal, abnormal and non-sperm classes, was presented. As a preliminary study, in [21], complex wavelets were employed as the feature extractor in the classification of sperm abnormalities for a small part of the SMIDS without any manual orientation step. Accuracy rates of 82.42% were obtained by using dual tree complex wavelet transform (DTCWT) [22, 23, 24, 25] based features due to their being shift invariant and having a better direction selectivity. Then, several pre-processing techniques were tested on the same data without a manual orientation process in another study [26], in which scale and rotation invariant Speed Up Robust Features (SURF) were fed

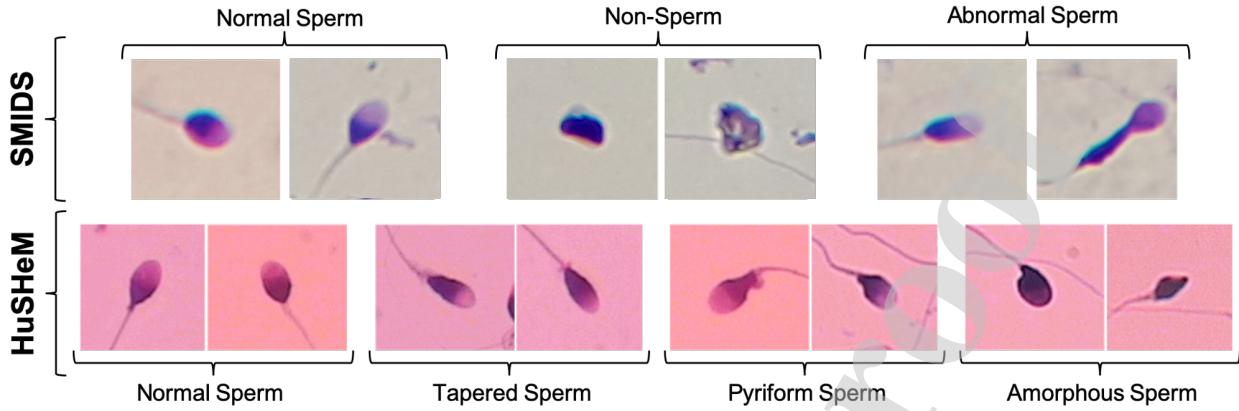


Figure 1: Example images for each class in SMIDS and HuSHeM Datasets

to SVM classifiers. According to the results, denoising processes have an important role in the classification performance, where regular median and adaptive wavelet based de-noising procedures increased the accuracy from 82% to 84% and 85%, respectively. Next, a directional masking technique to be used in sperm morphology analysis was proposed in [16]. The aim of using this technique was to improve the classification performance by only segmenting the sperm zones in the images and eliminating the manual orientation step of [15] and [14]. This directional masking technique was tested on the HuSHeM dataset in a  $k$ -nearest neighbours ( $k$ -NN) classification scenario and yielded an increase in accuracy from 45% to 57% on classification performance. Lastly, a hybrid sperm morphology analysing framework including a comprehensive research study of deep learning and conventional technique was applied to the full version of SMIDS and the results were presented in [11]. In [11], an 87% accuracy rate was achieved with MobileNet on the classification of raw images for the full version of SMIDS.

In our present paper, the drawbacks of the previous studies are considered and novel enhancements are suggested: i) an automated masking technique is suggested to replace the manual orientation steps employed in [14, 15]. Additionally, it is observed that multiple sperm shapes can be located in the images of HuSHeM and SMIDS. Besides, staining blob shapes, which are similar to spermatozoa, can be misclassified by the learning models. The employed masking approach eliminates these residual spermatozoa and sperm-like staining blobs; ii) a wavelet transform based de-noising approach is

applied in order to reduce unwanted noise components caused by an improperly performed staining. To measure the performance of the proposed system, the developed framework is tested on both the HuSHeM and SMIDS datasets, and the effects of the directional masking plus adaptive de-noising approaches are presented by comparing classification metrics obtained with and without them.

## 2. Materials and Methods

### 2.1. Datasets

In this study, several classifiers have been tested on two well-known datasets with/without the proposed directional masking technique. The HuSHeM dataset was first introduced in [15] with a dictionary learning based classification idea. They created an image dataset from stained semen samples by using an embedded camera connected to an Olympos BX50 microscope. They captured the images in 20x zooming. The dataset includes 216 sperm heads under 4 classes: normal (54), tapered (53), pyriform (57), and amorphous (52). The images of the sperm heads are in RGB file format with a size of  $131 \times 131$  pixels.

The images in the SMIDS dataset were collected by a smartphone-based data acquisition approach. This data acquisition approach was first introduced in [27] for motile sperm detection and counting purposes. The proposed smartphone-based approach has been validated in sperm concentration analysis, with a high correlation between the manual counting results and the system's output [28]. The full version of SMIDS was first introduced in [11] with a hybrid morphological analysing framework. In

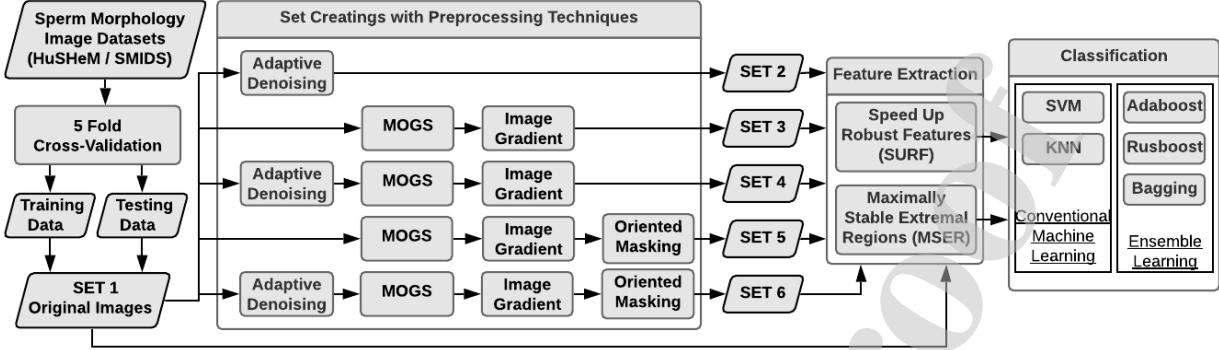


Figure 2: The flowchart of the all tested methods on two main sperm morphology datasets.

that framework, Modified Overlap Group Sparsity (MOGS) [29] was adopted as the preprocessing step and then the Fuzzy C-Means clustering technique was implemented for the segmentation of regions of interest (ROIs) in the stained images. Then, experts labelled each extracted ROI patch. In SMIDS, the abnormal class is not divided into sub-classes as in HuSHeM. The patches were labelled as normal (1021), abnormal (1005), and non-sperm (974). The images are in RGB file format with various sizes. The details of the image datasets are given in Table 1 and two example images for each class are given in Figure 1.

Table 1: The Image Distributions over Classes in Tested Datasets

SMIDS		HuSHeM	
Labels	# of images	Labels	# of images
Normal	1021	Normal	54
Abnormal	1005	Tapered	53
Non-Sperm	974	Pyriform	57
Total	3000	Amorphous	52
		Total	216

## 2.2. Preprocessing Methods

In this section, the preprocessing techniques used, which are designed to enhance the discrimination capability of the extracted features and/or to eliminate irrelevant information (noise components) in the microscopic images, will be described. First, a wavelet transform based adaptive de-noising approach is explained. Second, a Modified Overlap Group Sparsity method, which employs a non-convex regularization term chosen so that the total cost function (consisting of data consistency and regularization terms) is convex, is described [29]. Third, an intensity based image gradient technique that measures the directional change

in a sperm image is presented. Finally, a directional masking technique employing a sperm tail oriented elliptic mask for extracting sperm-only information is considered. The role of these preprocessing steps and the order in which they are applied to create the feature sets are given in Figure 2, which represents all the tested preprocessing, feature extraction, and classification methods.

### 2.2.1. Wavelet Based Local Adaptive De-Noising

One of the main problems of CASA systems is the noise components, which are superimposed on the sperm images, originating due to the lack of adequate light in the low magnification microscope and improperly stained semen smears. De-noising (noise reduction) aims to estimate the true signal (spermatozoa without noise) from noisy observations and increase the classification accuracy in the subsequent steps. Wavelet Based Local Adaptive De-Noising (WBLAD) [30], which is an effective and low complexity algorithm, employs a non-Gaussian bivariate probability distribution function to model the bivariate statistics of the wavelet coefficients of images. The model captures the dependence between a wavelet coefficient and its parent, that is, the coefficient at the same time location but in a coarser scale [31]. In [30, 32] it was proved that the WBLAD performs better than classical image de-noising approaches. In WBLAD, the signal of interest (a sperm image with noise) is decomposed by using the Dual Tree Complex Wavelet Transform (DTCWT) [22, 23], which has near shift-invariance as well as directional selectivity advantages over the ordinary discrete wavelet transform, resulting in complex wavelet coefficients (CWCs). Later, a bivariate non-linear shrinkage function, which is de-

rived by using Bayesian estimation theory and is a generalized version of the classical soft thresholding approach, is applied to the CWCs with the aim of removing the noise components. This non-linear shrinkage function takes into account the statistical dependencies between wavelet sub-bands and also employs simple models by estimating local noise model parameters. For any given wavelet coefficient  $w_1$  and its parent wavelet coefficient  $w_2$ , which is at the same spatial position as  $w_1$  but in the next coarser scale, the formulation of the employed model in WBLAD is

$$y = w + n \quad (1)$$

where  $w$  denotes the noise-free wavelet coefficients,  $y$  denotes the noisy wavelet coefficients, and  $n$  denotes a zero-mean Gaussian noise with variance  $\sigma_n^2$ . In WBLAD, the employed non-Gaussian bivariate probability density function (PDF) is

$$p_w(w) = \frac{3}{2\pi\sigma^2} \cdot e^{-\left(\frac{\sqrt{3}}{\sigma}\sqrt{w_1^2+w_2^2}\right)} \quad (2)$$

where  $\sigma$  denotes the standard deviation of the signal. Additionally, the maximum a posteriori estimator, which is used to estimate the noise free wavelet coefficients from the noisy observations and can be interpreted as a bivariate shrinkage function, is

$$\hat{w}_1 = \frac{\left(\sqrt{y_1^2+y_2^2} - \frac{\sqrt{3}\sigma_n^2}{\sigma}\right)_+ \cdot y_1}{\sqrt{y_1^2+y_2^2}} \quad (3)$$

The '+' sign denotes the positive values of the set and can be formulated as  $(x)_+ = \max(x, 0)$  [33, 34]. Single image samples taken from both the HuSHeM and SMIDS databases with and without WBLAD can be seen in Figure 3.

### 2.2.2. Modified Overlapping Group Shrinkage

After a detailed visual investigation of the sample images, it was observed that in both databases, the morphology of the main sperm bodies are mostly corrupted by the superimposed activity of other spermatozoa or an improperly applied staining process. Additionally, some of the sperm tails are not located in a straight orientation to the sperm body but have dispersed orientations. In order to minimize these corruption effects, a directional masking technique was proposed in [16]. This technique employs two elliptic masks whose orientation was obtained by using the tail direction and the

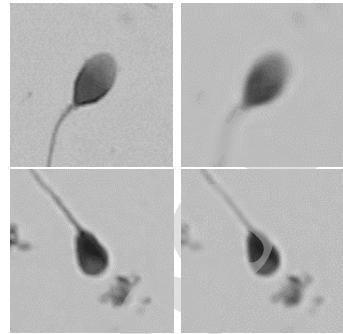


Figure 3: The effects of WBLAD on an example image in both datasets (Top-Left image is the original image from HuSHeM, Top-Right is the WBLAD applied HuSHeM image, Bottom-Left image is the original image from SMIDS, Bottom-Right is the WBLAD applied SMIDS image)

sperm-body. The orientation of the masks are calculated by forming a vector between the centroids of the sperm body and of the border shape of the sperm including the tail. In our study, the Modified Overlapping Group Shrinkage (MOGS) approach was employed to correctly find the centroid of the sperm body. As seen in the sperm images given in Figure 4, acrosomes have brighter intensity when compared to the rest of the sperm body. This bright region disrupts the process of locating the centroid of the sperm body: it might be wrongly shifted due to the pixels forming the tail. Therefore, MOGS [29] was employed with the aim of preserving the main sperm body while decreasing the effect of small particles such as sperm tails that do not form a solid group. MOGS is applied to entire two databases and the visibility of each sperm morphology is enhanced and a cleaner representation of possible blobs (main sperm bodies) in the image samples has been obtained. Later, the centroids of these blobs would be employed in finding the directional masking orientation. An example study describing the Sperm Segmentation performance of the MOGS can be found in [35]. The mathematical formulation of the employed MOGS is given below for a further understanding of the approach.

The observed signal/image that is intended to be processed with MOGS can be modeled as

$$s = x + n \quad (4)$$

where  $s$  is the noisy signal,  $x$  is the clean signal and  $n$  is additive white Gaussian noise. In the MOGS approach,  $x$  is assumed to have a group property in which the large magnitude image components form clusters. Assuming that prior knowledge has al-

ready been gathered about  $x$  (having group sparsity characteristics such as the sperm head and body) and  $n$  (standard deviation of the noise), a sparse component of the clean signal ( $x$ ) can be extracted from the observed signal ( $s$ ) by employing a sparse optimization model:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\{ F(x) = \frac{1}{2} \|s - x\|_2^2 + \lambda R(x) \right\} \quad (5)$$

where  $R(x)$  is a penalty term that can constrain the sparsity of  $x$  and  $\lambda > 0$  is a regularization parameter.  $\frac{1}{2} \|s - x\|_2^2$  is a fidelity term that can constrain the error between the estimated signal  $x$  and the observed signal  $s$ . In MOGS, the signal enhancement problem is taken to be a convex function in which a non-convex regularization term is employed. Therefore, this satisfies the desired group sparsity property [36].

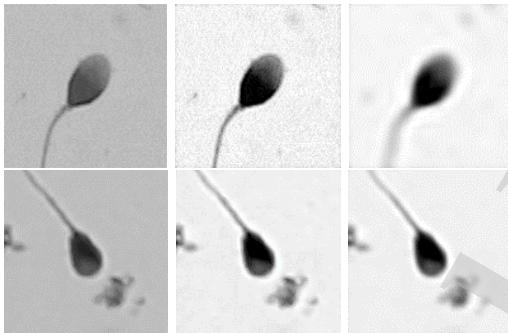


Figure 4: The effects of MOGS on an example image in both datasets with/without WBLAD (Top-Left and Bottom-Left are original images from HuSHeM and SMIDS, respectively. The images in the middle column are the results of implementing MOGS on images without WBLAD. Right column images are the results of implementing MOGS on images with WBLAD)

### 2.2.3. Image Gradient

The gradient of an image refers to the directional changes in the intensity values or colour space. It is one of the fundamental pieces of information used to segment an object in computer based image retrieval applications [37, 38, 39]. As an example, the Canny edge detector uses the image gradient for edge detection. Mathematically, the gradient of an image intensity is a two-variable function at each image point  $(x, y)$  as represented in Eq. 6. The derivatives of the values in the horizontal and vertical directions are stacked in a 2D vector.

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\delta f}{\delta x} \\ \frac{\delta f}{\delta y} \end{bmatrix} \quad (6)$$

where  $\frac{\delta f}{\delta x}$  and  $\frac{\delta f}{\delta y}$  represent the changes in the  $x$ - and  $y$ - directions and calculated by using the spatial filters given in Eqs 7 and 8. At each image point, the gradient vector points in the direction of the largest possible intensity and its length corresponds to the rate of change of the intensity in that direction. However, the calculation of the derivatives of intensity functions requires continuous points, whereas the coordinates are discrete in the image. There are several approximations proposed in the literature in order to convert points to a continuous form; otherwise, the derivatives of the functions cannot be defined. The most common way to approximate the image gradient is to convolve an image with a kernel. The Sobel operator with  $3 \times 3$  window size is selected as the optimum kernel type according to the obtained results in the present study.

$$G_x = \frac{\delta f}{\delta x} = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * Image \quad (7)$$

$$G_y = \frac{\delta f}{\delta y} = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * Image \quad (8)$$

The main aim of using the image gradient in the present study is to find the borders of the sperm shapes including the tail information. Later, the centroid of this border-shape will be calculated in order to find the sperm head-tail orientation. Therefore, only the magnitudes of the gradients are employed with respect to Eq. 9. It is observed from the results that the magnitudes are high at the borders of the sperm shapes while low magnitudes are found inside the sperm body. As expected, information about the sperm tail is also obtained as an output of the image gradient operation: the centroid of this obtained border-shape is shifted towards the sperm tail due to the lack of pixels representing the sperm body. The effect of the image gradient operation on two sample images taken from each database with and without WBLAD can be seen in Figure 5. During the calculations, the Sobel Kernel Operator, which preserves most of the information about the edge and gives the best orientation for a complicated sample sperm patch, was chosen as the main kernel for both databases.

$$G = \sqrt{G_x^2 + G_y^2} \quad (9)$$

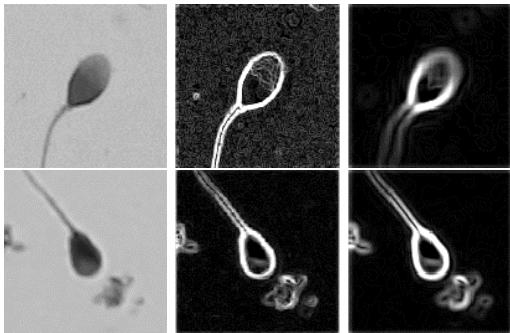


Figure 5: Effects of the image gradient operation on an example image in both datasets with/without WBLAD (Top-left and bottom-left images are the original images from HuSHeM and SMIDS, respectively. Middle column images are the results from implementing the gradient operation on images without WBLAD. Right images are the results of implementing the gradient operation on images after WBLAD)

#### 2.2.4. Automatic Directional Masking

In two previous studies [14, 15], manual orientations for the sperm shapes were provided by employing two steps. First, only one of the sperm shapes in each sample image was cropped manually. Second, each cropped sperm image was rotated manually to reach a fixed orientation. By applying these manual operations, the success of the subsequent classification steps was increased with the contribution of the robust features extracted from the oriented sperm shapes. In contrast to these two previous studies, a fully computerized framework, which can automatically crop and adjust the sperm head-tail orientation, is proposed in our study. In the employed fully automated directional masking technique, two elliptic filters in a merged format are derived to crop the sperm shapes. The merged elliptic masks are applied to the sperm shapes along the sperm head-tail axis, as in Figure 6. The aim of applying these elliptic filters is to remove redundant sperm shapes (as the main aim is to extract features from a single sperm, the additional spermatozoa are referred to as redundant) and noise components (caused by improperly applied staining process and/or the lack of adequate light in the low magnification microscope). The localization and orientation of the elliptic filters employs position vectors formed between the centroids of binary format sperm body shapes (the sperm bodies are obtained by using a blob analysis applied

to the outputs of the MOGS) and the centroid of the gradient magnitudes. The steps of the analysis for the directional masking are given in Figure 6. A more detailed explanation of this method can be found in [16].

#### 2.3. Feature Extraction and Classification

Interest Points (IPs) are the key points of the images where digital image processing algorithms can get discriminating information that could be employed in pattern recognition problems. IPs can be defined as a specific local pattern which differs from its close neighbouring pixels. These local feature extractors have advantages over general feature extractors (for example the Principal Component Analysis [40, 41]): i) they are more invariant under image transformations like translation, rotation and scaling, and ii) more robust to noise components. In the present paper, in order to benefit from these advantages, two well-known descriptors, called SURF and MSER (Maximally Stable Extreme Regions) were employed as the feature extractors. Subsequent to the feature extraction phase, 5 classification models were used for discriminating sperm dataset samples. Support Vector Machines (SVMs) and  $k$ -Nearest Neighbour ( $k$ -NN) were employed as the individual classifiers, while Adaptive Boosting (Adaboost), bagging and Random under sampling boosting (RUSboost) are used as ensemble approaches.

##### 2.3.1. Feature Extraction by SURF and MSER descriptors

A popular key point descriptor, the scale-invariant feature transform (SIFT), has been used as a feature extractor in various image processing and pattern recognition studies [42, 43, 44, 45]. Despite its advantages, such as being invariant to scale and orientation changes, SIFT suffers from imposing a significantly large computational burden during the calculations. An improvement on SIFT, SURF, which has less computational cost compared to SIFT, was proposed in [46]. The SURF descriptor provides the interest points of the images by using the determinant of the pre-calculated integral image based Hessian blob detectors. The sum of the Haar wavelet responses around these detected interest points of the image is used as the feature vector. SURF is also a scale and rotation invariant technique because of its integrating the image and summing the Haar wavelet responses in

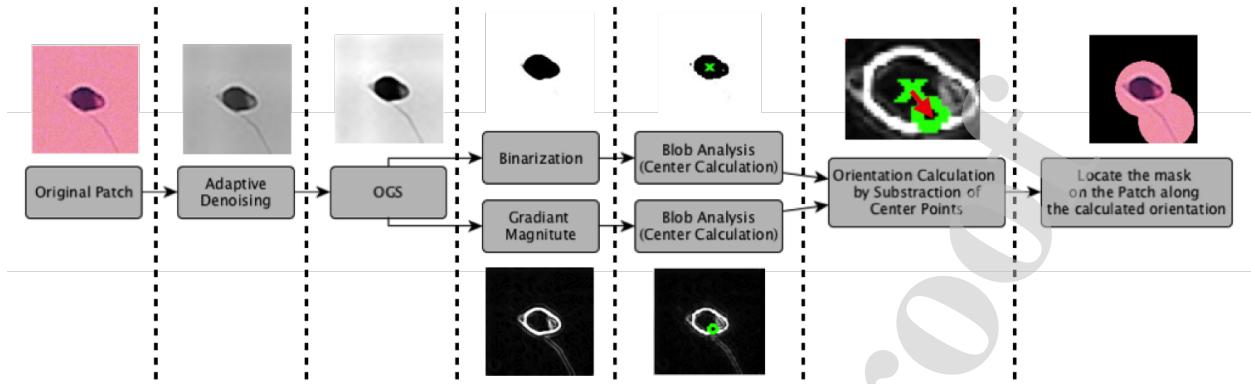


Figure 6: Flowchart of Automatic Directional Masking Technique [16]

both the  $x$ - and  $y$ -directions. The SURF detector is able to detect features from corners and blobs, but it can not detect key points for regions. On the other hand, MSER [47] obtains regions of interest by thresholding the pixel intensities. Extremal regions are defined as those in which all the pixel values are brighter or darker than those on the boundaries. MSER aims to find regions which remain stable (where the intensity values of the pixels are located between the threshold values) over a large range of changes in the threshold [48]. In addition to being invariant under rotation, which is also true for SURF, MSER based features are invariant under affine transformation of the image intensities. Hence, in order to benefit from and compare their complementary properties, SURF and MSER methods were applied to HuSHeM and SMIDS databases with the aim of feature extraction. Two image samples, one taken from HuSHeM and another from SMIDS, that were processed by using SURF and MSER, are represented in Figure 7. In order to highlight the denoising and directional masking effect, the gradient magnitudes resulting from a) raw sperm images, b) sperm images de-noised with wavelet transform and c) de-noised+masked sperm images, are given.

### 2.3.2. Classification Models

Several conventional machine-learning techniques have been employed to compare the performance of the preprocessing and feature extraction methods. SVM classifiers with linear, Gaussian (RBF), polynomial and quadratic kernels have been used as examples of a kernel base classifier. Additionally,  $k$ -NN has been used as a distance based classifier. In order to have some ensemble models, Adaboost, bagging and RUSboost algorithms have

been applied to two databases as the learning models.

In [49], SVM was proposed as a supervised learning technique based on statistical learning theory. SVM has been successfully applied to many medical classification tasks due to its generalization ability [50, 51, 52, 53]. Originally, SVM was proposed for binary classification problems and later it was adapted to multi-class [54] and one-class problems [55]. SVM can employ linear kernels (in linear classification) or non-linear kernels (in non-linear classification) in order to transform the original input space to a higher dimensional feature space in which linear separation can be done by fitting a separation plane [56]. In order to construct an unbiased robust model with SVM, careful tuning of three hyper-parameters should be considered: i) The kernel function which enables the SVM to operate in a high-dimensional space in which the classification turns into a linearly separable problem. ii) The penalty parameter which determines the level of the trade-off between the training error and the model complexity. iii) The parameter gamma for the Radial Basis Function (RBF) kernel, which is the parameter that controls the width of the RBF kernel and is also used to vary the trade-off between the bias and variance of the model.

The  $k$ -nearest neighbour ( $k$ -NN) classification [57] is an instance-based learning algorithm which is suitable for image classification with local features [58, 59]. The  $k$ -NN algorithm searches for the  $k$  nearest feature vectors, which are given in a training database consisting of a set of training feature vectors with known labels, of a query vector whose label is not known. The labels of the  $k$  nearest feature vectors are noted and a majority

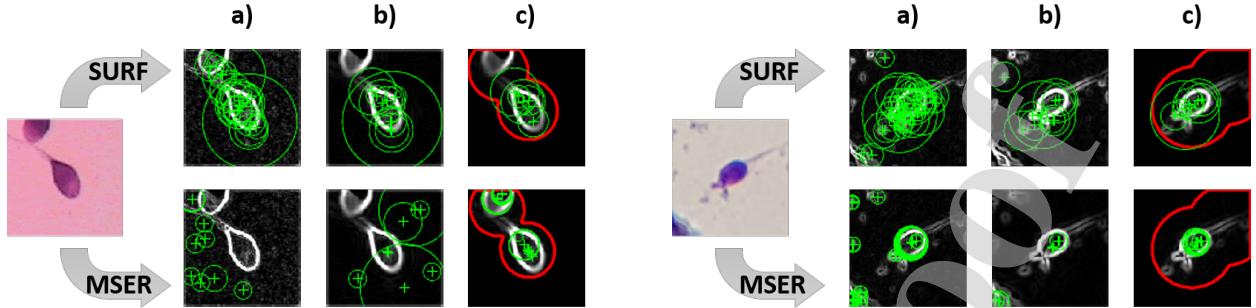


Figure 7: Image gradient magnitudes and interest points for two sample images. Left side group image is taken from HuSHeM while right side image group belongs to SMIDS. Upper row indicates that SURF is applied as the key point descriptor, while lower row images resulted from MSER. For each database: a) column images are taken from Set 3, b) column images are taken from Set 4 and c) column images are taken from Set 6. Green circles are used to represent the key points found, while the red merged elliptic shapes stand for the applied directional masks.

voting strategy is applied to these  $k$  nearest vectors to determine the final label of the query vector (test sample).

Alternatives to individual classifiers include ensemble techniques, such as boosting and bagging, which can combine multiple base models to increase the accuracy beyond that of the single best model. These have become popular in the machine learning literature [60, 61, 62, 63]. Ensemble methods use multiple learning algorithms (also called weak learners) to improve the overall performance by combining the individual decisions of each weak learner. Decision Trees (DT) are commonly employed in ensemble learners due to their fast and unstable characteristics [64]. Decision trees can be unstable because small variations in the input data might cause dramatic changes in the generated tree. However, by using DTs in an ensemble architecture, this drawback can be mitigated and, as a result, superior performance compared to individual classifiers can be obtained. In sperm images, especially in the HuSHeM database, a significant amount of noise components, due to improper staining or lack of adequate light in a low magnification microscope, have been observed. In ensemble learning, as an advantage of employing numerous weak learners, the errors incurred by a single weak learner can be compensated for by the predictions of other learners in the overall combination and a model robust to noise can be created. Therefore, three ensemble learning approaches were tested in the present study with the aim of obtaining robust classifiers even when a significant amount of noise occurs in the sperm patches. As the first ensemble model, bagging employs an ensemble of learning models where each is

given an equally-weighted prediction [65]. In bagging, small sample sets are created by drawing (with replacement) instances from the original dataset. Later, DTs are applied to each new subset (each resampled small sample set differs from the others) and majority voting is applied to the individual outputs of the weak learners. In this process, diversity is obtained with the resampling procedure by the use of different data subsets [66]. As the second ensemble approach, the AdaBoost [67] algorithm, which is an implementation of the boosting [68] approach for the task of classification, was employed. In AdaBoost, an iterative resampling procedure goes on as in the bagging case. However, unlike bagging, after each iteration, it places more focus on difficult instances, with the goal of correctly classifying examples in the next iteration that were incorrectly classified during the current iteration. The degree of focus is adjusted with a weight factor, which is equal for all instances in the beginning and tuned with respect to the outcome of the current iteration. Previous to each resampling operation, the weights of misclassified instances are increased; in contrast, the weights of correctly classified instances are decreased. By doing this, the individual weak predictors are forced to focus more on hardly separable instances, and their combination becomes more powerful [66]. Lastly, as the third ensemble model, RUSboost [69], which is a hybrid sampling/boosting classification model that is particularly well-suited to dealing with the class imbalance problem, was tested on two databases. In RUSboost, instances from the majority class are randomly removed (undersampling) until the desired balance between numbers of samples of the

classes is achieved. Subsequent to this balancing procedure, a boosting strategy is applied to the new subsets [70]. In the present work, even if the number of sperm samples in the used databases are very close, there are minor differences, and RUSboost was applied to both databases in order to understand the effect of these minor variations.

#### 2.4. Experiments and the Proposed Computational Framework

In this study, 6 scenarios have been created to present the performance comparison of the proposed computational framework. Each scenario was derived from a sub-stage of the proposed framework. In this way, it is aimed to present the effects of each preprocessing technique on the classification performances clearly. Each scenario is indicated with the number of its set. Set 1 denotes raw images that have not been subjected to any preprocessing techniques. Wavelet Based Local Adaptive De-Noised Images are gathered in Set 2. Set 3 contains the images preprocessed by two cascade techniques: MOGS and Image Gradient. Set 4 is created to investigate the effect of WBLAD on MOGS and Image Gradient techniques. Set 5 and Set 6 are the image sets created to observe the effects of the previously proposed directional masking technique [16]. In order to show the effect of the WBLAD technique on the directional masking approach, two sets are arranged as with/without WBLAD. Set 5 includes the images resulting from applying directional masking to noisy images, whereas Set 6 contains the images resulting from applying a WBLAD based de-noising. The creation of each set and its description are shown in Figure 2 and Table 2, respectively.

Table 2: Image Set IDs with the corresponding applied preprocessing techniques

ID	Description
Set 1	Original Image Sets (No Preprocessing)
Set 2	WBLAD
Set 3	MOGS + Image Gradient
Set 4	WBLAD + MOGS + Image Gradient
Set 5	MOGS + Image Gradient + Directional Masking
Set 6	WBLAD + MOGS + Image Gradient + Directional Masking

In order to maximize the system performance, the parameters employed in the preprocessing steps were tuned by trial and error. During the entire analysis, the decomposition level was set to 7 by using  $3 \times 3$  convolution filters in the WBLAD preprocessing technique. Therefore, the resolution of each

image in the corresponding sets (Set 2, Set 4 and Set 6) was changed to  $128 \times 128$  to satisfy the needs of the selected decomposition level ( $2^7$ ). MOGS, as an another wavelet based technique, was configured to employ three decomposition levels with a filter size of  $5 \times 5$  for all related sets (Sets 3, 4, 5 and 6). Image gradient basically depends on a convolution process using filters, and the Sobel filter was selected for creating all related sets (Sets 3, 5 and 6). The directional masking technique is a self-adjusting technique which automatically resizes the masking region according to the spatial information of the sperm shape. Therefore, it was directly used in all related sets (Sets 5 and 6).

Subsequent to the arrangement of a set by using preprocessing steps, each obtained set was individually included in the feature-extraction and train/test classification schemes, as illustrated in Figure 2. In the training and testing sample organization,  $K$ -Fold Cross validation was used with  $K = 5$ . In this regard, 171 and 2400 image patches were chosen as training samples, while 45 and 600 image patches were chosen as testing samples for each created set of HuSHeM and SMIDS, respectively. Following the organisation of the samples into training and testing samples, feature matrices were obtained by MSER and SURF descriptors as training and testing feature matrices for each created set. Eight classifiers were trained and tested with these feature matrices. In the testing phase, the decisions of each classifier corresponding to relevant key points (more than one key point may stand for a single image sample) were given to a majority voting scheme with the aim of deciding on the sample class. The flowchart for all tested experiments is presented in Figure 2.

According to the results of all experiments, which will be given and discussed in the next section, we proposed a computational framework, which can be used in classification of any stained sperm images. The proposed approach includes WBLAD to minimize the effect of the non-sperm parts and noise in the patches, MOGS to make the acrosome part of the sperm more visible, Image Gradient to enhance the intensity of the borders of the sperm, and the automatic directional masking technique to cover the non-sperm regions. After these preprocessing steps, MSER was selected as the feature extraction descriptor in the proposed approach. SVM with a non-linear kernel, Quadratic or Polynomial, was picked as the classifier for the extracted features. The flowchart of the proposed computational

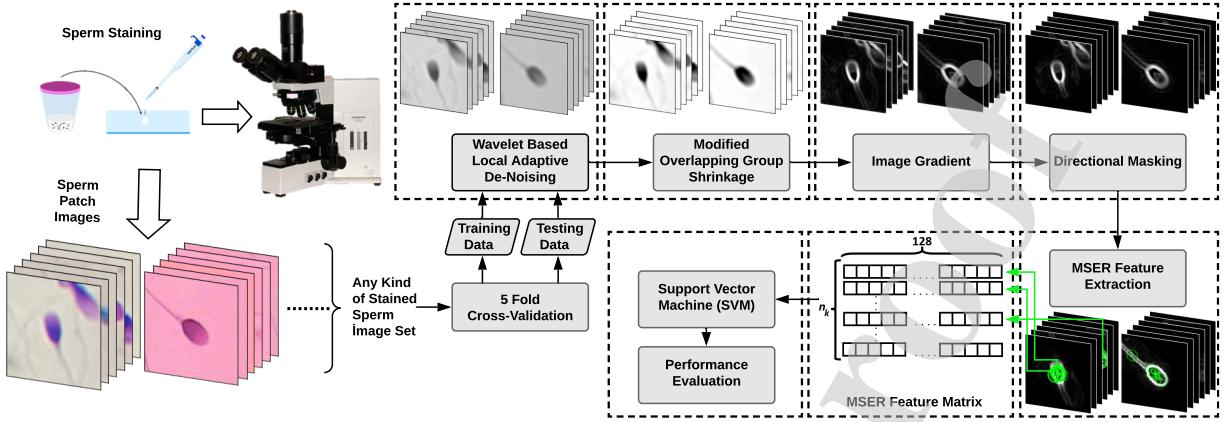


Figure 8: Proposed Computational Framework

framework is given in Figure 8.

### 3. Experimental Results

In this section, the qualitative and quantitative results of the steps of the applied framework, including the preprocessing, SURF and MSER based feature extraction and individual/ensemble classification techniques, are presented. In particular, first, the effect of the preprocessing step will be given and some inferences, which will be supported by the quantitative results in the following parts, will be presented. Second, all the classification results using 6 feature subsets and 5 classifiers will be presented quantitatively and the relation between the subsets including the wavelet based adaptive de-noising and directional masking effects will be highlighted.

#### 3.1. Qualitative Results

Referring to Figure 1, it can be seen that a significant amount of noise components, which are superimposed on the sperm images, are present in both the HuSHeM and SMIDS datasets. These noise components may originate from a lack of adequate light in a low magnification microscope and/or improperly stained semen specimens. Naturally, these noise components will affect the extracted features that characterize the instances for the purpose of classification. Noise may reduce the performance of the system, in terms of classification accuracy, time in building a classifier, and the size of the classifier [71]. In Figure 5, the existence of noise components is represented in a more obvious manner when the

image gradient operator is applied. Even though some machine learning methods have included various mechanisms to enhance their noise immunity, the existence of noise is still a problem, and it should be decreased by using preprocessing methods before the learning model is formed. Therefore, two noise reduction approaches were applied to the sperm samples with the aim of increasing the accuracy. First, WBLAD was employed to reduce texture like noise especially seen in the HuSHeM dataset. In Figure 3 and Figure 5, the impact of WBLAD is presented and it is seen that most of the texture type noise components have been removed from the images. The effect is more impressive in the HuSHeM samples. As a second noise removing approach, the directional masking technique was applied to the sperm images and most of the non-sperm blobs that occur due to improperly applied staining were removed and the individual sperm shapes were extracted. The results of four sperm images, two images from each dataset, are presented in Figure 9 and it is seen that individual sperm shapes were successfully obtained with the directional masking even for the cases where multiple sperm shapes occur in the same image frame.

The performance of the applied noise removal algorithms can be further assessed by visualizing their effect on the extracted features. Figure 7 presents the interest points, which were found by using SURF and MSER descriptors, for sperm images taken from both datasets. It is seen that the number of wrongly detected interest points due to the noise components was highly reduced when the WBLAD and directional masking preprocessing ap-

Table 3: Accuracy values for 6 image sets classified with various classifiers

		HuSHeM Dataset (Set Numbers)						SMIDS Dataset (Set Numbers)							
			1	2	3	4	5	6		1	2	3	4	5	6
Conventional Machine Learning Techniques	SVM	Quad	SURF	68.2	70.4	81.8	83.3	81.7	84.1	77.6	78.2	83.3	84.2	84.9	85.1
		MSER	SURF	78.7	78.5	77.7	84.8	77.9	85.8	80.7	80.9	83.9	84.1	84.3	85.7
		Poly	SURF	65.9	70.9	78.0	83.9	78.6	84.4	72.9	75.2	81.3	82	82.6	82.7
		MSER	SURF	76.6	77.5	77.5	85.7	78.0	86.6	78.9	79.2	81.7	81.9	83.5	84
		RBF	SURF	60.1	63.0	78.0	80.7	80.3	83.3	75	75.5	82.1	82.6	83.3	83.7
		MSER	SURF	73.3	73.6	74.7	83.8	76.1	83.9	80.2	80.9	82.4	83.4	83.7	83.9
		Lin	SURF	49.6	51.5	63.6	67.9	59.6	67.8	65.1	68.8	72.7	75.6	76.1	76.6
		MSER	SURF	66.4	65.0	65.8	80.2	65.6	80.8	73.8	75	75.5	75.8	76.6	76.9
		KNN	SURF	58.4	57.9	50.2	70.7	56.1	72.4	71.9	72.6	78.8	80.5	80.5	80.7
		MSER	SURF	74.0	71.9	72.3	83.0	72.3	83.1	77.7	77.9	81.2	81.4	81.8	82.5
Ensemble Learning Techniques	Adaboost	SURF	42.6	55.2	43.8	54.3	45.7	60.8	64	64.7	60.5	68	62.9	68.9	
	MSER	SURF	70.1	73.4	62.0	73.4	63.4	77.0	74.9	75.2	71.4	71.8	73.5	74.5	
	Bagging	SURF	56.1	60.3	60.2	74.8	56.1	76.2	71.3	71.9	78.2	79.2	80.1	80.2	
	MSER	SURF	73.3	73.8	76.5	80.7	76.9	81.2	77.4	78.8	79.8	79.6	80.9	81.2	
Average Accuracy of Sets	RUSboost	SURF	53.1	63.6	55.0	60.3	57.3	60.8	67.4	67.8	67.6	67.6	68.2	69.1	
	MSER	SURF	61.8	63.2	62.5	66.3	61.3	71.3	70.6	70.7	70.2	70.3	72.6	72.7	
	Overall	SURF	54.7	58.8	61.5	70.0	62.5	71.8	68.4	70.6	74.7	76.4	76.3	77.5	
		MSER	70.8	71.1	70.5	78.5	70.3	79.8	75.1	75.7	77.3	77.6	78.8	79.3	

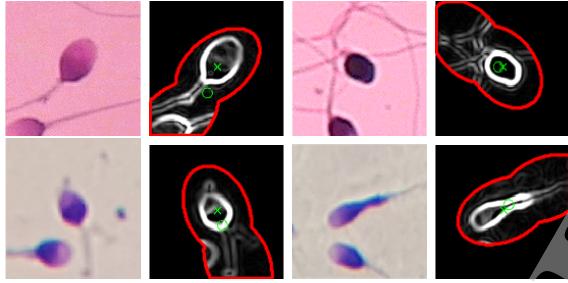


Figure 9: Visual demonstration of automatic directional masking technique on two images from both datasets (Top images from HuSHeM, Bottom images from SMIDS)

proaches were applied.

### 3.2. Quantitative Results

In Table 3, the performances (in terms of accuracy) of all the employed feature sets and classifiers are given in detail. In the bottom three rows of this table, the means of the accuracies are given for both the SURF and MSER features. It is observed that the MSER features outperformed the SURF features, when the average accuracy of all classifiers was considered for Set 6 (the best performing feature set), by approximately 8% and 2% for the HuSHeM and SMIDS datasets, respectively. In addition, when the other 5 feature sets (for both HuSHeM and SMIDS) were considered, it is seen that the classification of MSER based features resulted in having a higher accuracy compared to SURF for all subsets: the increases in the accuracy varied between 1.2% and 16.1%. Further, among all the results, the highest accuracies were obtained with the MSER features: 86.6% and 85.7% for the

HuSHeM and SMIDS datasets, respectively. As can be seen from these facts, the MSER emerges as the best feature extraction method in all of the cases.

Regarding the classification methods, it can be seen from Figure 10 that the performance of an SVM based model, especially with the quadratic kernel, surpassed the other classifiers by having the minimum error rates (for the mean error rate of all feature sets and also for the each individual set error). For the HuSHeM, the lowest error rates were reached by arranging the samples as in Set 6 with a non-linear SVM kernel classification. The polynomial kernel in the SURF based features and the quadratic kernel for the MSER features resulted in the lowest error rates in the SVM classification. Similarly, for the SMIDS dataset, the SVM with a quadratic kernel had the lowest error rates for both the SURF and MSER feature extractors. Another non-parametric and linear classification technique,  $k$ -NN, yielded fewer errors in its classification compared to the SVM with a linear kernel. Mapping, with the non-linear kernels, the features into high-dimensional feature spaces had a great impact on the classification of all the MSER and SURF based feature sets.

It should also be noted that the ensemble methods showed the worst performance except bagging. The weighted re-sampling idea utilized in Adaboost and Rusboost is not effective when the data includes outliers and noises. Therefore the error rate was measured highest especially in HuHSeM dataset. Contrary to the Boosting techniques, bagging employs a random subset creation idea which is more robust to noise and overfitting problem.

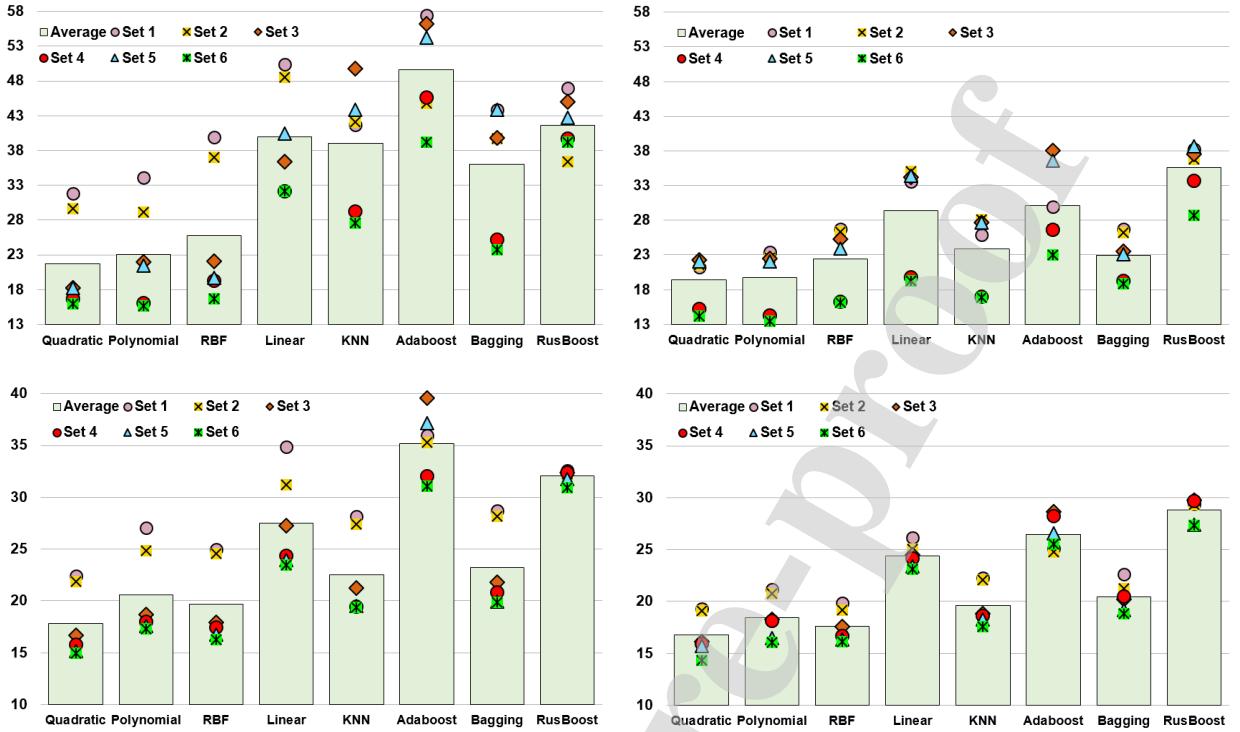


Figure 10: Classifier based classification errors for each created set. Top figures present the SURF (left) and MSER (right) feature classification errors for HuSHeM dataset. Bottom figures present the classification errors for SMIDS dataset in terms of SURF (left) and MSER (right) descriptors

Therefore, the average error rate of bagging was smaller than SVM with linear kernel and boosting techniques for both HuSHeM and SMIDS when SURF and MSER were employed as feature extractor, and it also performed almost same with  $k$ -NN for all cases in terms of mean error rate.

Concerning the performance in terms of the feature subset, Figure 11 shows the error rates obtained with various classifiers (indicated by markers with various colours) for each subset. The average error rates of all classifiers related with a certain subset (given as a bar graph) are also given. For all databases and feature extraction methods, Set 6, in which the WBLAD and directional masking is applied, shines as the best approach for sperm classification. Confirming the inferences obtained from the previous figure, it can be seen from Figure 11 that the SVM classifier with non-linear kernels (RBF, Quadratic and Polynomial) resulted in less error than the averages (average of all classifiers for a specific set, given as bar charts) for all 6 sets derived from both the HuSHeM and SMIDS datasets. It can be seen that bagging and  $k$ -NN are also effective classifiers in MSER based feature

classifications, having less error than the averages of each related set. However, Adaboost and RusBoost, which are used as boosting techniques for weak learners in ensemble schemes, seem inconvenient techniques in the classification of both MSER and SURF based features for all created sets. The error rates of both Adaboost and RusBoost were significantly higher than the average classifier error of each set. Likewise, the SVM with a linear kernel is not effective, due to the non-linear nature of the sperm classification problem. It could classify each created set with better performance than the boosting techniques but less than the other applied non-parametric classification technique,  $k$ -NN. As a result, considering SVM with Quadratic Kernel's consistent performance in all the 6 created sets, it can be concluded that the SVM with Quadratic Kernel is the best classifier for the proposed scheme.

The performance of the proposed automated method can be further assessed based on how the de-noising approaches affect (improve) the classification accuracies. In order to show and emphasize this de-noising effect, Figure 12, which only highlights the performance of WBLAD by using the

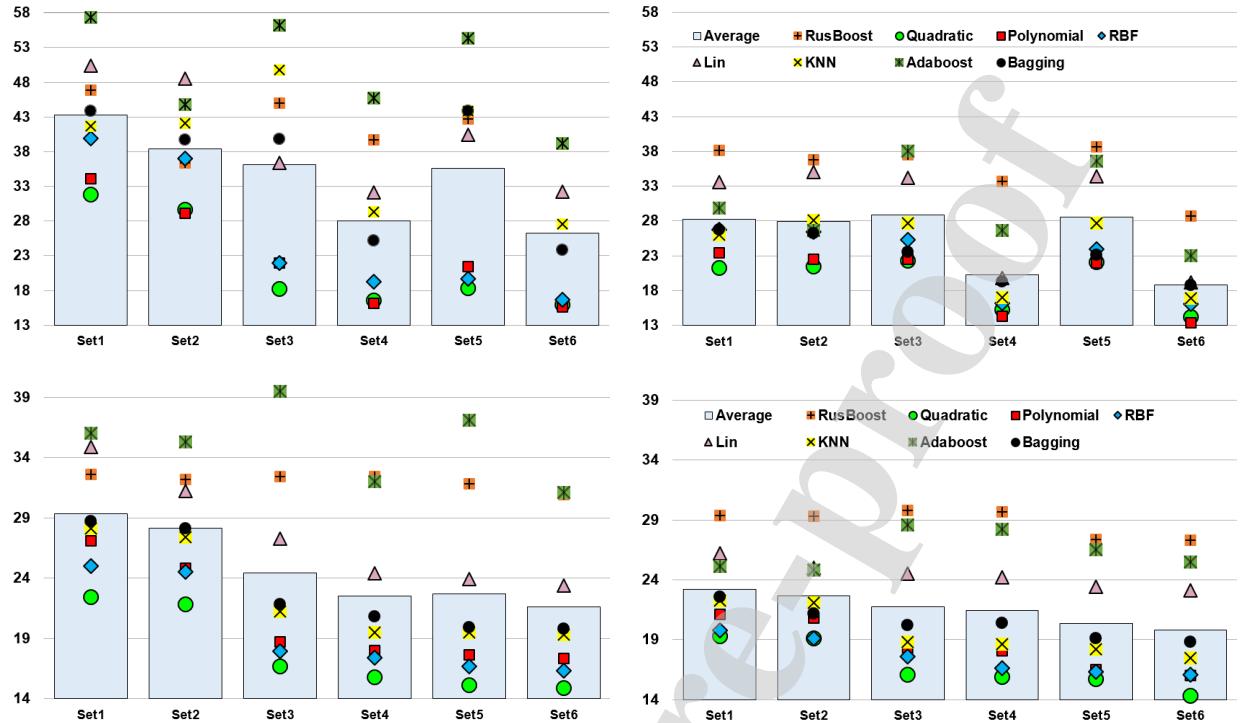


Figure 11: Classification errors for each classifier and each set. Top figures present the SURF (left) and MSER (right) feature classification errors for HuSHeM dataset. Bottom figures present the classification errors for SMIDS dataset in terms of SURF (left) and MSER (right) descriptors

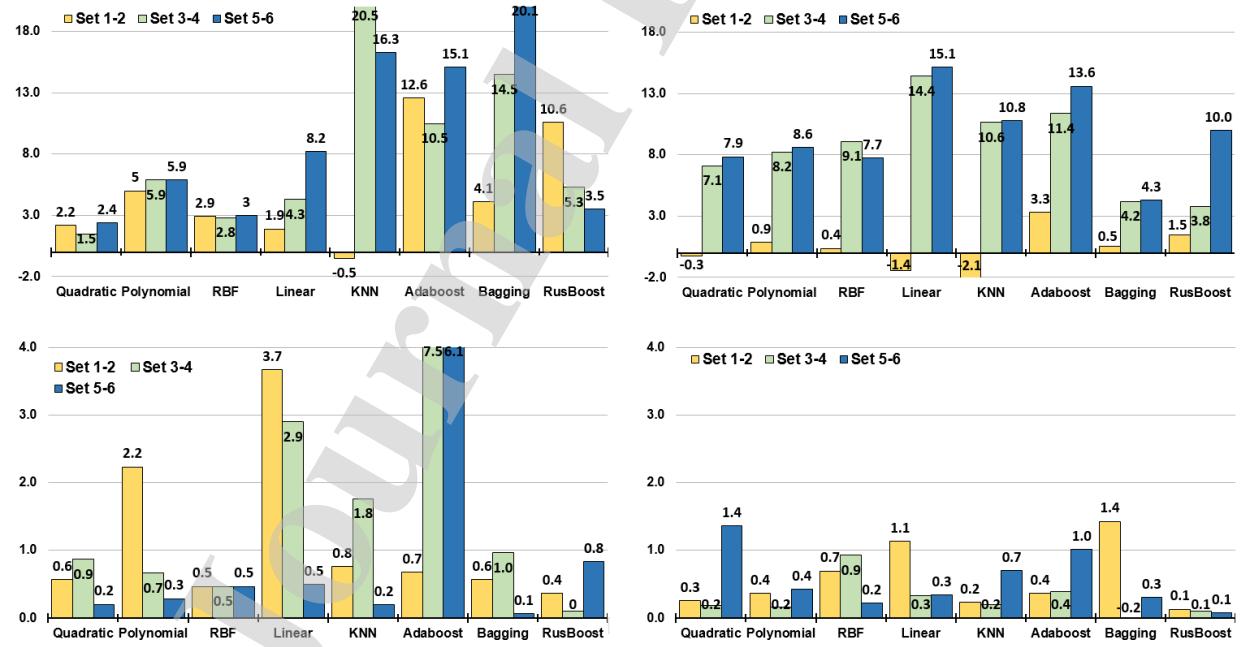


Figure 12: The Effects of Wavelet Based Local Adaptive De-Noising on Classification Accuracy. Top figures present the increment for SURF (left) and MSER (right) feature classification accuracy of HuSHeM. Bottom figures present the increment for SMIDS dataset in terms of SURF (left) and MSER (right) descriptors

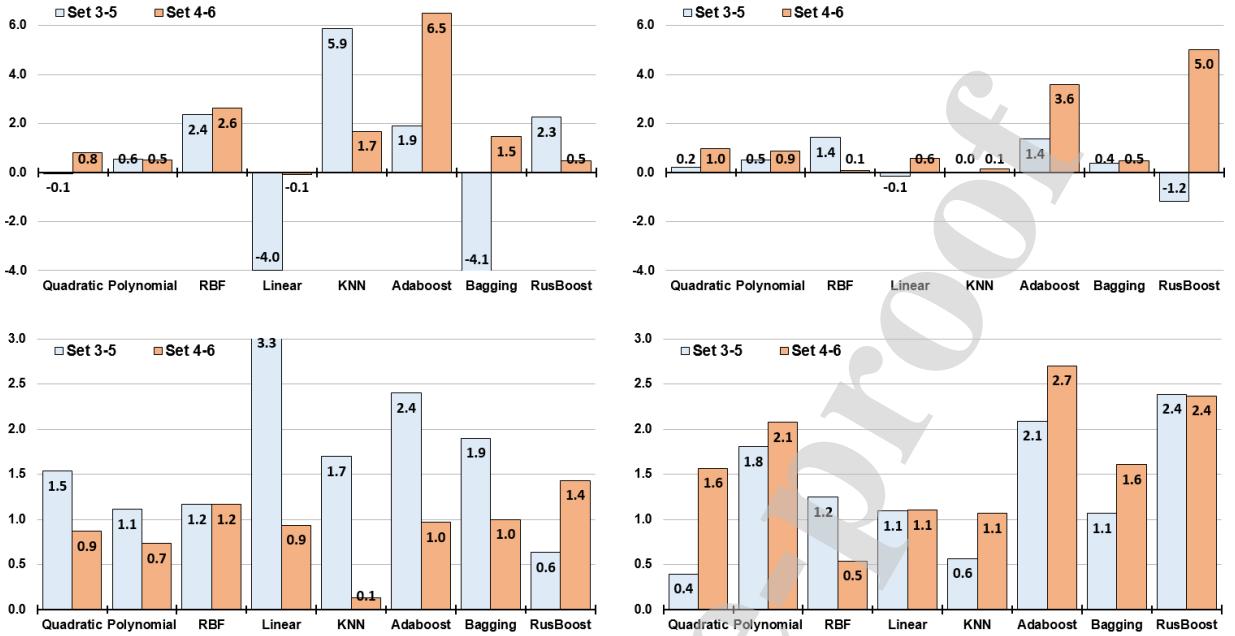


Figure 13: Effects of Automatic Directional Masking technique on Classification Accuracy. Top figures present the increase in the classification accuracy of the SURF (left) and MSER (right) features for HuSHeM. Bottom figures present the contribution to classification accuracy for SMIDS in terms of SURF (left) and MSER (right) descriptors

rates of increase in the accuracy, has been prepared. In this figure, the legend ‘Set 1-2’ labels the increase in classification accuracy between the original raw images and the WBLAD based de-noised images. The legend ‘Set 3-4’ labels the increase in accuracy in the classification of MOGS+Image gradient images with/without WBLAD based de-noising procedure. The legend ‘Set 5-6’ labels the effects on WBLAD when the images to which the directional masking has been applied are de-noised. A close inspection of Figure 12 reveals that WBLAD is a crucial preprocessing step for both datasets and all feature extractors. WBLAD increases the accuracies, especially for SURF based feature classification cases, because of their sensitivity to noise. In addition, it can be seen that the effect of de-noising is more dramatic for HuSHeM, due to its noisiness compared to SMIDS. Regarding the classifiers, the highest impact was measured for Adaboost, which is a noise sensitive ensemble technique. Additionally, the SVM with linear kernel and the other linear based lazy classifier, *k*-NN, is also sensitive to noise due to the implementation of low level linear distance based learning. Therefore, the classification performance of both techniques is also greatly increased by WBLAD. On the other hand, SMIDS is a less noisy dataset compared to HuSHeM, and

MSER has greater noise immunity compared to SURF, which limits the increase when WBLAD is applied to MSER features obtained from SMIDS.

Figure 13 reveals the contribution of the proposed directional masking technique to classification. The results are given in the form of accuracy differences, as in Figure 12. As was also mentioned in the set creation section, Sets 5 and 6 comprise automatically directional masking image sets, while Sets 3 and 4 are created with the same techniques, but without directional masking implementation. The comparison figures emphasize the importance of the implemented directional masking technique by highlighting the obtained improvements in the accuracy for all classifiers in the presence of the masking effect. When the SURF based results are considered, it is seen that only the SVM employing a linear kernel and the bagging classifiers resulted in an accuracy reduction for Sets 3-5, in which the WBLAD is not applied, obtained from HuSHeM. Not only the lack of WBLAD, but also the sensitivity to noise of the SURF technique implemented may be important reasons for this reduction in accuracy. In addition, the characteristics of the applied classifier may cause this reduction. The random subset selection procedure used in bagging may cause a re-sampling of incorrectly masked im-

ages. Similarly, the linear SVM kernel may fail to distinguish the samples by means of a linear hyperplane due to the non-linear nature of the problem. However, it is observed that the automatic directional masking provided significant increases in the accuracies for all other classifiers. The most dramatic increase (6.5%) was for Adaboost, when the de-noised SURF features extracted from HuSHeM were fed into it. Likewise, in the MSER based feature extraction technique applied to the HuSHeM dataset, directional masking also provided a significant accuracy increase in RUSboost. Boosting techniques using non-randomized weighting in the re-sampling process and the directional masking approach may assist the algorithm to correctly boost the ensemble classifiers, by removing non-sperm blobs that may cause label noise, in the re-sampling stage. Besides the mentioned effects of increased accuracy, due to the automatic directional masking approach, seen in the low performance classifiers (ensemble methods,  $k$ -NN, and linear kernel SVM) in the HuSHeM dataset, the naturally noise immune SVM also enjoyed accuracy increases of 1% for each descriptor when the polynomial and quadratic kernels were employed. On the other hand, the SMIDS dataset is less noisy than HuSHeM, and therefore the effect of the directional masking was observed more efficiently in all classifiers due to the correctly masked regions (the centroids that are used to find the orientation vector are found better in the presence of less noise). The directional masking even increased the classification performance of quadratic SVM, which is the most robust classifier, by 1.6%. Similar to the classification of the HuSHeM dataset, directional masking has also increased the accuracies of the boosting techniques in SMIDS.

#### 4. Discussion

One of the intuitive explanations for the presented results is that bagging is the most successful ensemble learner for all datasets and feature extraction methods. This can be explained by the noise immunity of the employed ensemble learning methods. In randomized ensembles, the randomization process is employed to build diverse models that have good generalization capacity, resulting in robust classifiers like bagging [72]. In contrast, adaptive ensemble methods, such as boosting, are dramatically affected by the noise components, causing

incorrect labeling in descriptor based feature extraction. In SURF and MSER, the wrongly found key points cause class-label noise and this type of noise hampers the classification performance [73]. The performance difference between the bagging and boosting based algorithms can be explained by showing how errors are handled during the training phase of these algorithms. In bagging, the randomness injected during the model creation of the ensemble is not correlated with the noise, and this equalizes the influence of different samples during the training process. Unlike in boosting, the weights of the mis-classified samples are increased irrespective of whether they are correctly labeled or not. The increase in the weight of correctly labeled samples that are difficult to discriminate is helpful, because it reduces the classification bias in boosting. However, the focus on wrongly labeled samples tends to mislead the learning process and this misdirected adaptivity renders the boosting algorithms overly susceptible to noise [74].

Regarding the noise immunity of the other classification algorithms, those SVM algorithms which employ non-linear kernels perform the best among all classifiers for all datasets and feature extraction methods. SVM maps the samples to a high dimensional feature space in which the classification problem is linearly separable. Later, it searches for a hyperplane which can correctly separate the training samples and maximizes the minimal distance between the hyperplane and the closest instances. This minimal distance is called the margin of the classifier. A narrow margin implies that the model is sensitive to tiny variations of the samples near the boundary, resulting in a low immunity to noise. On the other hand, large margin classifiers like SVM are more resistant to possible variations and this makes them more robust against noise. Due to this property, SVM outperformed all the other classifiers in this study, as seen in Figure 10. In addition, referring to Figures 12 and 13, the contribution of the WBLAD and directional masking to the ensemble methods and linear SVM is more visible due to the limited generalization capacity of the former, and for the latter, due to the non-linear nature of the problem. However, the effect of the denoising algorithms in the non-linear SVM kernels is relatively low because they have noise immunity in the first place due to their large separation margin. However, it should be noted that even though the contribution to the non-linear kernel SVM is relatively low, up to 9.1% and 2.2% accuracy increases

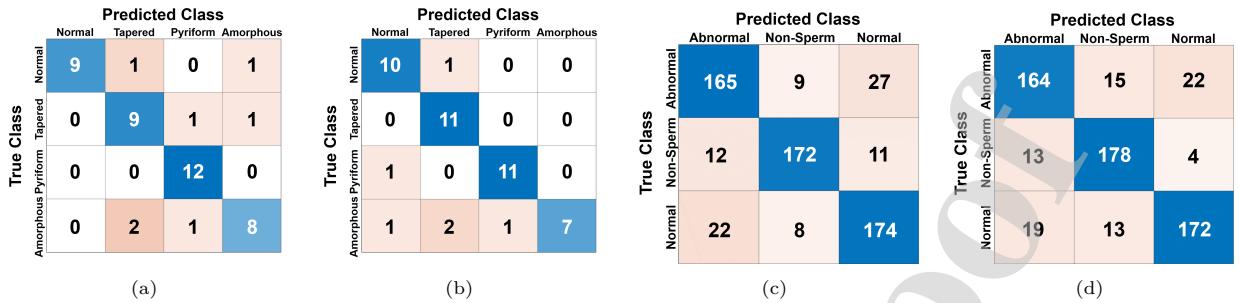


Figure 14: Confusion Matrices for the best obtained accuracies: a) SURF descriptor based classification of HuSHeM Set 6 by SVM with Polynomial Kernel; b) MSER descriptor based classification of HuSHeM Set 6 by SVM with Polynomial Kernel; c) SURF descriptor based classification of SMIDS Set 6 by SVM with Quadratic Kernel; d) MSER descriptor based classification of HuSHeM Set 6 by SVM with Quadratic Kernel

were obtained by using WBLAD in the HuSHeM and SMIDS respectively, while the contribution of directional masking is 2.6% and 2.1% for HuSHeM and SMIDS. This proves that even though the non-linear SVM kernels have natural noise immunity, preprocessing techniques like WBLAD and directional masking still have a critical impact on the classification performance.

The confusion matrix is an informative visualization technique to show the individual classification performance. In other words, it provides detailed information to determine the most challenging and successful classes in classification [75]. Additionally, in some cases, an accuracy metric can be misleading due to having unbalanced labels in the dataset. Therefore, other performance metrics, based on the confusion matrix, such as Precision, Recall, *F*-Score and Kappa, are more informative performance metrics than Accuracy. In order to analyse the best classifiers in these regards, the confusion matrices of the Polynomial and Quadratic SVM for both feature sets (SURF and MSER) of HuSHeM and SMIDS are presented in Figure 14, respectively. In addition to general accuracy values, the comparison metrics based on the confusion matrix, for all the classes, are also given in Table 4 in order to represent the class based performance. The HuSHeM database consists of normal sperm shapes and abnormal spermatozoa having head defects, referred to as tapered, pyriform, and amorphous. However, the SMIDS were formed by using abnormal sperm shapes including head defects, midpiece defects, and tail defects, which provides a greater opportunity to evaluate the success of the recommended automated method. When the precision, recall and *F*-score metrics are considered, further

useful information can be obtained. For instance, high recall rates, which is the ability of a model to find all positive samples, were obtained for tapered spermatozoa by using the MSER features. Tapered spermatozoa have been shown to be related to fertilization rate; therefore, when detected as the predominant abnormal form in sperm samples, tapered head spermatozoa might affect fertilization [76]. This shows that the recommended model has a significant capacity to be employed as an indicator in infertility analysis. Additionally, it is observed from the confusion matrix that the MSER is more successful than SURF in the classification of normal and abnormal spermatozoa, by the resulting higher *F*-Scores. Regarding the Kappa metric, which is a statistical measure of inter-annotator agreement for categorical items by comparing an observed accuracy with an expected accuracy, the highest values were obtained with MSER features: 0.822 and 0.785 for HuSHeM and SMIDS respectively. The possible highest value of Kappa is 1, which is referred to as the perfect classification case. In [77], values of Kappa in the range 0.61–0.80 are called substantial, while values in the range 0.81–1 are called almost perfect classification. This shows that the proposed method is very powerful when the MSER features are employed in the sperm classification problem, for both databases, and the performance of the proposed scheme is promising in terms of accuracy, precision, recall, *F*-Score, and Kappa coefficient.

A minor fact that needs to be considered is that the proposed method has relatively low performance in the classification of the Amorphous type of spermatozoa, indicated as deformations seen in the acrosome walls. This small reduction in the performance of the classification of Amorphous type

Table 4: The evaluation of best classifiers in terms of confusion matrix based performance metrics derived from Figure 14

		HuSHeM Dataset				SMIDS Dataset		
		Normal	Tapered	Pyriform	Amorphous	Normal	Abnormal	Non-Sperm
SURF	Precision	100	75	85.7	80	82.1	82.9	91
	Recall	81.8	81.8	100	72.7	85.3	82.1	88.2
	F-Score	90	78.3	92.3	76.2	83.7	82.5	89.6
	Accuracy			84.4			85.1	
MSER	Kappa			0.792			0.777	
	Precision	83.3	78.6	91.7	100	86.9	83.7	86.4
	Recall	90.9	100	91.7	63.6	84.3	81.6	91.3
	F-Score	86.9	88	91.7	77.8	85.6	82.6	88.8
	Accuracy			86.6			85.7	
	Kappa			0.822			0.785	

sperm is mostly related with the application of MOGS and WBLAD. As can be seen in Figures 3 and 4, the application of these two methods causes a blurring effect in the sperm walls, and this hampers the ability of the employed feature extractors to detect the Amorphous type spermatozoa. The abnormalities seen in the Amorphous type spermatozoa are caused by the deflections of the acrosome wall: some minor deflections that are not noticed due to the blurring may mislead the classification models. However, in [76], it was mentioned that no direct correlation between the high percentage of amorphous heads in sperm samples and fertilization could be proved.

In order to validate the performance of the proposed approach, these results are compared with previous publications concerning sperm morphology analysis. Table 5 provides the accuracy values of various methods that have been used on the HuSHeM and SMIDS datasets. In addition, results referring to another publicly available sperm morphology dataset, SCIAN-Morpho, are also given in Table 5 as alternative studies employing descriptor based feature extraction like the present study. However, the SCIAN-Morpho dataset was not included in the present study as a testing database due to the way it was built. The sperm samples forming this database had already been cropped and oriented. The present study mainly focuses on the analysis of image patches in which single or multiple spermatozoa may be located either separately or superimposed. Besides, the proposed system can handle cases which may include overlapping spermatozoa, obtaining a satisfactory performance. When the SMIDS and HuSHeM datasets are investigated, it was seen that the ratio of separate multiple spermatozoa and overlapped sperm cases are very rare. However, even in these cases,

the proposed algorithm can extract discriminating information from the segmented sperm head and body parts due to the application of the elliptic masks. In the applied masking technique, the orientations of the elliptic masks are derived along the sperm-head and tail axis by using the position vector that goes from the centroid of the binary format sperm body to the centroid of the gradient magnitude shape (the sperm tail shifts the centroid towards itself due to the lack of sperm body in gradient magnitude form). On the other hand, in the cases of separated multiple spermatozoa and of overlapping sperm, the position vector is formed between the two sperm bodies. In this situation, the effect of the sperm tails, which shifts the centroid of the gradient magnitude shape, become insignificant in the presence of two solid blobs, the body/head blob of each sperm. The orientation of the elliptic masks are formed along the centroid of two sperm head/body blobs and this results in omitting the sperm tails in the masked image. In the presented datasets, the sperm defects originate from the head and body parts of the sperm, as seen in Figure 1, which makes the extracted features still robust and powerful. It should be noted that all the accuracy values presented in Table 5 are the best values for the given methods. Regarding the studies based on the SCIAN-Morpho dataset, Chang et al. used an analysis method that employs descriptor based features that are fed to an SVM for classification. In the first approach of [20], a 49% accuracy rate was achieved. Subsequently, fusion of the feature sets extracted by multiple descriptors were fed to a cascade connected SVM structure, resulting in an increase from 49% to 58% in accuracy [14]. As an another enhancement, Shaker et al. increased this accuracy to 62% by using another codebooks based feature extraction technique, dictio-

Table 5: Sperm detection performance of the proposed technique and similar studies in the literature

Data Set Name	Data Set Information	Orientation and Cropping	Paper	Method	Accuracy (%)
SCIAN-Morpho	1132 Sperm Patches 35 x 35 Resolution Not Overlapped Five Classes	Already Oriented and Cropped	Chang et al. 2017 [20]	Feature Extraction by Fourier Descriptor SVM Classification	49
			Chang et al. 2017 [14]	Feature Extraction by Ensemble of Fourier, Zernike and Shape Descriptors Ensemble of SVM Classification	58
			Shaker et al. 2017 [15]	Color Space Converting Dictionary Learning	62
HuSHeM	216 Sperm Patches 131 x 131 Resolution Overlapped Four Classes	Manual	Shaker et al. 2017 [15]	Manually Sperm Head Rotation and cropping into 50 x 76 pixels Color Space Converting Dictionary Learning	92.2
		Automatic	Ilhan et al. 2019 [16]	Automatic Directional Masking k-NN classification	57.4
		Automatic	Proposed Framework	Adaptive De-Noising Overlapping Group Shrinkage Image Gradient Directional Masking Feature Extraction by MSER descriptor SVM Classification (non-linear Kernel)	86.6
SMIDS	536 Sperm Patches 190 x 170 Resolution Overlapped Three Classes	Not Applied	Ilhan et al. 2018 [21]	Dual Tree Complex Wavelet Trans. SVM classification	82.3
		Not Applied	Ilhan et al. 2018 [26]	Adaptive De-Noising Feature Extraction by SURF Descriptor SVM classification	83.4
		Not Applied	Ilhan et al. 2020 [11]	Overlapping Group Shrinkage Fuzzy-C Means Clustering Data Augmentation Classification by MobileNet	87
	3000 Sperm Patches 190 x 170 Resolution Overlapped Three Classes	Automatic	Proposed Framework	Adaptive De-Noising Overlapping Group Shrinkage Image Gradient Directional Masking Feature Extraction by MSER descriptor SVM Classification (non-linear Kernel)	85.7

nary learning [15]. These studies have shown that descriptors can be applied to sperm morphology analysis, although they are performed on a dataset that is already manually manipulated.

With respect to HuSHeM based studies, Shaker et al. used a descriptor based dictionary learning approach in the classification [15]. Codebooks based descriptors such as dictionary learning are highly sensitive to orientation and noise. Therefore, they also manually rotated and cropped the sperm head parts in the HuSHeM images. At this point, it should be noted that our proposed approach is the only fully automatic morphology analysis method among the studies of other groups given in Table 5. It employs a directional masking technique which automatically finds the sperm head-tail orientation and crops the region of interest for further analysis. This directional masking approach was first tested on the HuSHeM dataset by feeding the histogram based features to  $k$ -NN, and it provided a 13% increase in classification accuracy [16]. In [21], our group introduced another sperm morphology database consisting of stained sperm

patches that need additional orientation finding and cropping operations for extracting individual sperm shapes. As a preliminary study, the wavelet and descriptor based features were investigated in the classification performance over a small size version of SMIDS without any directional masking approach [21]. Accuracies of 82.3% and 83.4% were achieved by wavelet and SURF based descriptors, respectively. Furthermore, we expanded our dataset to a sample size of 3000 in order to measure the generalization capacity of our models in a more objective manner [11]. For this enlarged dataset, a fully automated hybrid system based on MobileNet deep learning classification was proposed. Additionally, wavelet transform and descriptors based classification scenarios were also tested on this enlarged SMIDS dataset [11]. The MobileNet deep learning approach, Wavelet and SURF based feature classifications resulted in 87%, 80.07% and 77.6% accuracy rates, respectively. In this study, all the methods were applied to image frames having the original image size, no additional orientation and cropping techniques were employed.

Table 6: The output of the directional masking technique in sperm images of several species

	<b>Human [78]</b>	<b>Stallion [79]</b>	<b>Boar [78]</b>	<b>Bull [80]</b>	<b>Rooster [81]</b>	<b>Mice [82]</b>
Staining Method	SpermBlue	Eosin-nigrosin	SpermBlue	Eosin-nigrosin	Hemacolor	Eosin
Length ( $\mu\text{m}$ )	$4.71 \pm 0.25$	$5.67 \pm 0.36$	$8.96 \pm 0.24$	$10.11 \pm 0.57$	$13.5 \pm 3.5$	$8.05 \pm 0.05$
Width ( $\mu\text{m}$ )	$2.78 \pm 0.23$	$2.85 \pm 0.31$	$4.52 \pm 0.03$	$5.21 \pm 0.43$	$1.3 \pm 1.7$	$3.29 \pm 0.02$
Original Images						
Automated Directional Masked Images						

Even though the MobileNet deep learning approach has shown satisfactory performance, further analysis was required for the following reasons: i) the deep learning based techniques require a high amount of data and system resources in the training phase in order to construct robust models. However, the publicly available sperm morphology databases have a very limited number of samples (HuSHeM has 216 samples) and this number of samples is not enough to train a deep neural network. Therefore, in order to provide a generic classification framework which may be applied to other public databases for validation, a robust system employing a conventional machine learning structure, including preprocessing for image enhancement, feature extraction and train/test of extracted features by conventional machine learning models, is needed. ii) Sperm images inherit noise components due to improperly applied staining process and effects due to the microscope employed, therefore, the performance of the preprocessing approaches must be determined. After establishing enhancements obtained with preprocessing methods in a conventional classification scheme and finding the optimum de-noising parameters, preprocessed images can be given to deep learning models for further analysis. Examining the mentioned needs, an automated system employing de-noising approaches is proposed in the current study. By using the directional masking and WBLAD, we increased the SVM classifier accuracy to 86.6% and 85.7% for HuSHeM and SMIDS (full set), respectively. It has been seen that the proposed automated directional masking approach has competi-

tive performance as well as some additional important advantages, such as eliminating the need for a manual orientation/cropping process, providing faster classification than the Deep learning based approaches and achieving high accuracy rates with limited data size.

An interesting aspect that needs to be discussed is the applicability of the proposed system to databases for other species obtained by using different staining techniques. As has been pointed in [83], the chosen staining method may perform well for a specific species while it is not suitable for other species. Besides, the differences between the morphometrics that are obtained by the use of different staining methods can reach 30%–60% even for the same species [84, 85, 86]. In [87, 88, 89], it is mentioned that the accuracy of a sperm morphology evaluation depends on the care taken in the slide preparation and fixation, and also the choice of staining method. The use of different staining techniques may produce significant variations in the intensity and contrast, and most importantly in the shape and size of the spermatozoa [83, 90]. Therefore, the classical unsupervised sperm detection and clustering approaches using morphometric measurements (such as the length, width, perimeter and area of the sperm body, eccentricity, tail length and total sperm length) are strongly dependent on the shape and size of the sperm. Minor fluctuations occurring in morphometric measures due to the use of various staining methods can be problematic in the unsupervised learning based evaluation of fertility disorders in which the obtained morphometric measures are compared with reference

values for correct assessment. Therefore, in our study, a supervised classification approach employing SURF and MSER features, which are invariant under changes in scale and orientation has been proposed, instead of using an unsupervised approach based on morphometric measures that are highly sensitive to the type of method of staining. Regarding the adaptability of our proposed approach to other sperm shapes obtained from various species using different staining techniques, the most critical issue is the performance of the pre-processing techniques, especially the automatic directional masking technique. If a properly separated sperm head, sperm body and sperm tail can be obtained by using pre-processing steps, the feature extraction and classification substeps would certainly perform well if only the whole framework is fed with the samples that belong to a specific species and staining method from scratch. In order to validate the proposed approach's adaptability, sperm samples from five species (including stallions, boars, bulls, roosters, and mice) that were stained by using various chemicals were pre-processed by employing our method. The results are given in Table 6. In addition, human sperm stained by SpermBlue (other than hematoxylin eosin) was also fed to our pre-processing steps in order to validate the success of the directional masking technique. The name of the species, the averages and standard deviations of the sperm-head widths and heights for the entire collection of sperm samples employed in relevant studies, and the staining methods used, are given in the upper part of Table 6. As can be seen, our framework performs well for all the species, due to the morphological similarities between the human spermatozoa and those of the other species, even if different staining methods were used. However, the evaluation of the proposed method's success on databases for other species is out of the scope of our current study, so we leave this for future research, only showing the applicability of our approach without providing a detailed analysis.

## 5. Conclusion

Fifteen per cent of the world's population is suffering from male factor based infertility. Therefore, sperm morphology analysis, as one of the important stages in the determination of a diagnosis of male based infertility, has critical importance. In manual clinical tests, the interpretation of the analysis results depends greatly on the observer's

experience. In order to eliminate the human factor in these tests, several computerized diagnostic approaches have been proposed in the literature. In this study, we also proposed a computational framework for the morphology analysis of any kind of stained sperm images. A cascade implementation of several preprocessing techniques, namely, WBLAD, MOGS, Image Gradient and automatic directional masking, is the key part of the proposed framework. In experimental tests, differently arranged preprocessing scenarios were tested on two datasets (HuSHeM and SMIDS). The proposed paper has also presented the results of a comprehensive suite of experiments comparing the performance of two feature extraction techniques and five machine learning classification models for dealing with sperm morphology analysis.

According to the results, 86.6% and 85.7% accuracy rates have been achieved by non-linear kernel based SVMs in the classification of MSER features extracted from cascade preprocessing implemented on HuSHeM and SMIDS images, respectively. The preprocessing part of the proposed framework, which focuses on the orientation information of the sperm in the images, increased the accuracy rate by 10% and 5% for HuSHeM and SMIDS compared to the original images. Therefore, our findings reveal that a directional masking technique provides better classification due to eliminating non-sperm regions and noise. The performance of the proposed computational framework has been compared with frameworks presented in the existing literature. It has been found from the results that in addition to demonstrating comparable to or better accuracy than most state-of-the-art methods, the proposed framework has a great advantage in eliminating the exhaustive manual orientation and cropping processes, with a reasonable consumption of time and source.

In future research, we aim to concatenate the extracted features by different descriptors from images on which the directional masking has been implemented, before the classification step, in order to increase the accuracy. In addition to the feature extraction level, the ensemble combination of non-linear kernel based classifiers using different voting ideas will be examined. Besides conventional machine learning models, deep learning based classifiers will be investigated in terms of transfer learning for the sake of avoiding long training times. Finally, the proposed computational framework is planned to be used as a real-time processing device.

This intended device can be employed in a clinical centre in order to evaluate the performance of the system in more realistic scenarios.

### Compliance with ethical standards

**Conflict of Interest:** The authors declare that they have no conflict of interest.

**Ethical approval:** All procedures performed were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards and ethical approval was obtained from the Faculty of Medicine of Istanbul University.

**Informed consent:** Informed consent was obtained from all individual participants included in the study.

### References

- [1] W. H. Organization, et al., WHO laboratory manual for the examination and processing of human semen, Geneva: World Health Organization, 2010.
- [2] B. A. Keel, B. W. Webster, Handbook of the laboratory diagnosis and treatment of infertility, CRC Press, 1990.
- [3] J. Auger, P. Jouannet, F. Eustache, Another look at human sperm morphology, *Human Reproduction* 31 (2016) 10–23.
- [4] T. Sivanarayana, C. R. Krishna, G. J. Prakash, K. M. Krishna, K. Madan, B. S. Rani, G. Sudhakar, G. R. Raju, Casa derived human sperm abnormalities: correlation with chromatin packing and dna fragmentation, *Journal of assisted reproduction and genetics* 29 (2012) 1327–1334.
- [5] Z. B. Popović, J. D. Thomas, Assessing observer variability: a user's guide, *Cardiovascular diagnosis and therapy* 7 (2017) 317.
- [6] J. Lu, Y. Huang, N. Lü, Computer-aided sperm analysis: past, present and future, *Andrologia* 46 (2014) 329–338.
- [7] M. J. Tomlinson, A. Naeem, Casa in the medical laboratory: Casa in diagnostic andrology and assisted conception, *Reproduction, Fertility and Development* 30 (2018) 850–859.
- [8] R. P. Amann, D. Waberski, Computer-assisted sperm analysis (casa): capabilities and potential developments, *Theriogenology* 81 (2014) 5–17.
- [9] S. T. Mortimer, G. van der Horst, D. Mortimer, The future of computer-aided sperm analysis, *Asian journal of andrology* 17 (2015) 545.
- [10] J. Talarczyk-Desole, A. Berger, G. Taszarek-Hauke, J. Hauke, L. Pawelczyk, P. Jedrzejczak, Manual vs. computer-assisted sperm analysis: can casa replace manual assessment of human semen in clinical practice?, *Ginekologia polska* 88 (2017) 56–60.
- [11] H. O. Ilhan, I. O. Sigirci, G. Serbes, N. Aydin, A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods, *Medical & Biological Engineering & Computing* (2020) 1–22.
- [12] A. Bijar, A. Pe, M. Mikaeili, et al., Fully automatic identification and discrimination of sperm's parts in microscopic images of stained human semen smear, *Journal of Biomedical Science and Engineering* 5 (2012) 384–395.
- [13] V. Chang, J. M. Saavedra, V. Castañeda, L. Sarabia, N. Hitschfeld, S. Härtel, Gold-standard and improved framework for sperm head segmentation, *Computer methods and programs in biomedicine* 117 (2014) 225–237.
- [14] V. Chang, L. Heutte, C. Petitjean, S. Härtel, N. Hitschfeld, Automatic classification of human sperm head morphology, *Computers in biology and medicine* 84 (2017) 205–216.
- [15] F. Shaker, S. A. Monadjemi, J. Alirezaie, A. R. Naghsh-Nilchi, A dictionary learning approach for human sperm heads classification, *Computers in biology and medicine* 91 (2017) 181–190.
- [16] H. Ilhan, G. Serbes, N. Aydin, Automatic directional masking technique for better sperm morphology segmentation and classification analysis, *Electronics Letters* 55 (2019) 256–258.
- [17] F. Vásquez, C. Soler, P. Camps, A. Valverde, A. García-Molina, Spermogram and sperm head morphometry assessed by multivariate cluster analysis results during adolescence (12–18 years) and the effect of varicocele, *Asian journal of andrology* 18 (2016) 824.
- [18] J. Santiago-Moreno, M. C. Esteso, S. Villaverde-Morcillo, A. Toledano-Díaz, C. Castaño, R. Velázquez, A. López-Sebastián, A. L. Goya, J. G. Martínez, Recent advances in bird sperm morphometric analysis and its role in male gamete characterization and reproduction technologies, *Asian journal of andrology* 18 (2016) 882.
- [19] G. Bellastella, T. G. Cooper, M. Battaglia, A. Ströse, I. Torres, B. Hellenkemper, C. Soler, A. A. Sinisi, Dimensions of human ejaculated spermatozoa in papanicolaou-stained seminal and swim-up smears obtained from the integrated semen analysis system (isas®), *Asian journal of andrology* 12 (2010) 871.
- [20] V. Chang, A. Garcia, N. Hitschfeld, S. Härtel, Gold-standard for computer-assisted morphological sperm analysis, *Computers in biology and medicine* 83 (2017) 143–150.
- [21] H. O. Ilhan, G. Serbes, N. Aydin, Dual tree complex wavelet transform based sperm abnormality classification, in: 2018 41st International Conference on Telecommunications and Signal Processing (TSP), IEEE, pp. 1–5.
- [22] I. W. Selesnick, R. G. Baraniuk, N. C. Kingsbury, The dual-tree complex wavelet transform, *IEEE signal processing magazine* 22 (2005) 123–151.
- [23] N. Kingsbury, The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement, in: *Signal Processing Conference (EUSIPCO 1998)*, 9th European, IEEE, pp. 1–4.
- [24] G. Serbes, B. E. Sakar, H. O. Gulcur, N. Aydin, An emboli detection system based on dual tree complex wavelet transform and ensemble learning, *Applied Soft Computing* 37 (2015) 87–94.
- [25] G. Serbes, B. Sakar, N. Aydin, H. Gulcur, An emboli detection system based on dual tree complex wavelet transform, in: *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, Springer, pp. 819–822.
- [26] H. O. Ilhan, I. O. Sigirci, G. Serbes, N. Aydin, The ef-

- fect of nonlinear wavelet transform based de-noising in sperm abnormality classification, in: 2018 3rd International Conference on Computer Science and Engineering (UBMK), IEEE, pp. 658–661.
- [27] H. O. Ilhan, N. Aydin, A novel data acquisition and analyzing approach to spermiogram tests, *Biomedical Signal Processing and Control* 41 (2018) 129–139.
- [28] H. O. Ilhan, N. Aydin, Smartphone based sperm counting—an alternative way to the visual assessment technique in sperm concentration analysis, *Multimedia Tools and Applications* (2019) 1–27.
- [29] P.-Y. Chen, I. W. Selesnick, Group-sparse signal denoising: non-convex regularization, convex optimization, arXiv preprint arXiv:1308.5038 (2013).
- [30] L. Sendur, I. W. Selesnick, Bivariate shrinkage with local variance estimation, *IEEE Signal Processing Letters* 9 (2002) 438–441.
- [31] E. Bullmore, J. Fadili, V. Maxim, L. Şendur, B. Whitcher, J. Suckling, M. Brammer, M. Breakspear, Wavelets and functional magnetic resonance imaging of the human brain, *Neuroimage* 23 (2004) S234–S249.
- [32] G. Chen, S.-E. Qian, Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage, *IEEE Transactions on Geoscience and remote sensing* 49 (2011) 973–980.
- [33] L. Sendur, I. W. Selesnick, Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency, *IEEE Transactions on signal processing* 50 (2002) 2744–2756.
- [34] L. Şendur, I. W. Selesnick, A bivariate shrinkage function for wavelet-based denoising, in: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 2, IEEE, pp. II–1261.
- [35] H. O. Ilhan, G. Serbes, N. Aydin, The effects of the modified overlapping group shrinkage technique on the sperm segmentation in the stained images, in: 2018 41st International Conference on Telecommunications and Signal Processing (TSP), pp. 1–4.
- [36] W. Zhang, Y. Ding, X. Yan, M. Jia, Weak multiple fault detection based on weighted morlet wavelet-overlapping group sparse for rolling bearing fault diagnosis, *Applied Sciences* 10 (2020) 2057.
- [37] T. N. Tan, K. D. Baker, Efficient image gradient based vehicle localization, *IEEE Transactions on Image Processing* 9 (2000) 1343–1356.
- [38] F. Matern, C. Riess, M. Stamminger, Gradient-based illumination description for image forgery detection, *IEEE Transactions on Information Forensics and Security* 15 (2019) 1303–1317.
- [39] W. Wu, Y. Zhang, Q. Wang, F. Liu, P. Chen, H. Yu, Low-dose spectral ct reconstruction using image gradient 0-norm and tensor dictionary, *Applied Mathematical Modelling* 63 (2018) 538–557.
- [40] C. E. Thomaz, G. A. Giraldi, A new ranking method for principal components analysis and its application to face image analysis, *Image and Vision Computing* 28 (2010) 902–913.
- [41] C. Zhou, L. Wang, Q. Zhang, X. Wei, Face recognition based on pca image reconstruction and lda, *Optik* 124 (2013) 5599–5603.
- [42] S. Saleem, R. Sablatnig, A robust sift descriptor for multispectral images, *IEEE signal processing letters* 21 (2014) 400–403.
- [43] L. Ghoualmi, A. Draa, S. Chikhi, An ear biometric system based on artificial bees and the scale invariant feature transform, *Expert Systems with Applications* 57 (2016) 49–61.
- [44] X. Wu, Y. Tang, W. Bu, Offline text-independent writer identification based on scale invariant feature transform, *IEEE Transactions on Information Forensics and Security* 9 (2014) 526–536.
- [45] G. A. Montazer, D. Giveki, Content based image retrieval system using clustered scale invariant feature transforms, *Optik* 126 (2015) 1695–1699.
- [46] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (surf), *Computer vision and image understanding* 110 (2008) 346–359.
- [47] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions, *Image and vision computing* 22 (2004) 761–767.
- [48] Y. Li, S. Wang, Q. Tian, X. Ding, A survey of recent advances in visual feature detection, *Neurocomputing* 149 (2015) 736–751.
- [49] C. Cortes, V. Vapnik, Support-vector networks, *Machine learning* 20 (1995) 273–297.
- [50] G. Serbes, C. O. Sakar, Y. P. Kahya, N. Aydin, Pulmonary crackle detection using time-frequency and time-scale analysis, *Digital Signal Processing* 23 (2013) 1012–1021.
- [51] S. Ulukaya, G. Serbes, Y. P. Kahya, Overcomplete discrete wavelet transform based respiratory sound discrimination with feature and decision level fusion, *Biomedical Signal Processing and Control* 38 (2017) 322–336.
- [52] S. Ulukaya, G. Serbes, Y. P. Kahya, Wheeze type classification using non-dyadic wavelet transform based optimal energy ratio technique, *Computers in biology and medicine* 104 (2019) 175–182.
- [53] M. M. Rahman, S. K. Antani, G. R. Thoma, A learning-based similarity fusion and filtering approach for biomedical image retrieval using svm classification and relevance feedback, *IEEE Transactions on Information Technology in Biomedicine* 15 (2011) 640–646.
- [54] J. Weston, C. Watkins, Multi-class support vector machines, Technical Report, Citeseer, 1998.
- [55] S. S. Khan, M. G. Madden, A survey of recent trends in one class classification, in: Irish conference on artificial intelligence and cognitive science, Springer, pp. 188–197.
- [56] L. Wang, *Support vector machines: theory and applications*, volume 177, Springer Science & Business Media, 2005.
- [57] K. Q. Weinberger, J. Blitzer, L. K. Saul, Distance metric learning for large margin nearest neighbor classification, in: *Advances in neural information processing systems*, pp. 1473–1480.
- [58] G. Amato, F. Falchi, knn based image classification relying on local feature similarity, in: *Proceedings of the Third International Conference on SIMilarity Search and APplications*, pp. 101–108.
- [59] M. Mejdoub, C. B. Amar, Classification improvement of local feature vectors over the knn algorithm, *Multimedia Tools and Applications* 64 (2013) 197–218.
- [60] M. H. D. M. Ribeiro, L. dos Santos Coelho, Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series, *Applied Soft Computing* 86 (2020) 105837.
- [61] Y.-C. Chang, K.-H. Chang, G.-J. Wu, Application of extreme gradient boosting trees in the construction of credit risk assessment models for financial institutions,

- Applied Soft Computing 73 (2018) 914–920.
- [62] N. Simidjievski, L. Todorovski, S. Džeroski, Predicting long-term population dynamics with bagging and boosting of process-based models, Expert Systems with Applications 42 (2015) 8484–8496.
- [63] W. W. Ng, X. Zhou, X. Tian, X. Wang, D. S. Yeung, Bagging–boosting-based semi-supervised multi-hashing with query-adaptive re-ranking, Neurocomputing 275 (2018) 916–923.
- [64] S. Mounce, K. Ellis, J. Edwards, V. Speight, N. Jakomis, J. Boxall, Ensemble decision tree models using rusboost for estimating risk of iron failure in drinking water distribution systems, Water Resources Management 31 (2017) 1575–1589.
- [65] L. Breiman, Bagging predictors, Machine learning 24 (1996) 123–140.
- [66] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, F. Herrera, A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42 (2011) 463–484.
- [67] Y. Freund, R. E. Schapire, A desicion-theoretic generalization of on-line learning and an application to boosting, in: European conference on computational learning theory, Springer, pp. 23–37.
- [68] R. E. Schapire, The strength of weak learnability, Machine learning 5 (1990) 197–227.
- [69] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, A. Napolitano, Rusboost: A hybrid approach to alleviating class imbalance, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 40 (2009) 185–197.
- [70] A. R. Hassan, M. I. H. Bhuiyan, Automated identification of sleep states from eeg signals by means of ensemble empirical mode decomposition and random under sampling boosting, Computer methods and programs in biomedicine 140 (2017) 201–210.
- [71] X. Zhu, X. Wu, Class noise vs. attribute noise: A quantitative study, Artificial intelligence review 22 (2004) 177–210.
- [72] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, Journal of artificial intelligence research 11 (1999) 169–198.
- [73] P. Melville, N. Shah, L. Mihalkova, R. J. Mooney, Experiments on ensembles with missing and noisy data, in: International Workshop on Multiple Classifier Systems, Springer, pp. 293–302.
- [74] M. Sabzevari, Ensemble learning in the presence of noise, Ph.D. thesis, Universidad Autónoma de Madrid, Departamento de Ingeniería Informática, <https://repositorio.uam.es/handle/10486/687166>, 2019.
- [75] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Information processing & management 45 (2009) 427–437.
- [76] B. Dariš, A. Goropevšek, N. Hojnik, V. Vlaisavljević, Sperm morphological abnormalities as indicators of dna fragmentation and fertilization in icsi, Archives of gynecology and obstetrics 281 (2010) 363–367.
- [77] J. R. Landis, G. G. Koch, An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers, Biometrics (1977) 363–374.
- [78] G. Van der Horst, L. Maree, Spermbule®: a new universal stain for human and animal sperm which is also amenable to automated sperm morphology analysis, Biotechnic & Histochemistry 84 (2010) 299–308.
- [79] L. F. Brito, Evaluation of stallion sperm morphology, Clinical Techniques in Equine Practice 6 (2007) 249–264.
- [80] S. Kondracki, D. Banaszewska, A. Wysokińska, J. Chomicz, Sperm morphology of cattle and domestic pigs, Reprod Biol 6 (2006) 99–104.
- [81] S. Villaverde-Morcillo, M. Esteso, C. Castaño, A. Toledano Díaz, A. López-Sébastián, J. Campo, J. Santiago-Moreno, Influence of staining method on the values of avian sperm head morphometric variables, Reproduction in domestic animals 50 (2015) 750–755.
- [82] N. Takeda, K. Yoshinaga, K. Furushima, K. Takamune, Z. Li, S.-i. Abe, S.-i. Aizawa, K.-i. Yamamura, Viable offspring obtained from prm1-deficient sperm in mice, Scientific reports 6 (2016) 27409.
- [83] M. Czubaszek, K. Andraszek, D. Banaszewska, R. Walczak-Jedrzejowska, The effect of the staining technique on morphological and morphometric parameters of boar sperm, PLoS one 14 (2019).
- [84] T. Rijselaere, A. Van Soom, G. Hoflack, D. Maes, A. de Kruif, Automated sperm morphometry and morphology analysis of canine semen by the hamilton-thorne analyser, Theriogenology 62 (2004) 1292–1306.
- [85] M. Igner-Ouada, J. Verstegen, Validation of the sperm quality analyzer (sqa) for dog sperm analysis, Theriogenology 55 (2001) 1143–1158.
- [86] T. Rijselaere, D. Maes, G. Hoflack, A. De Kruif, A. Van Soom, Effect of body weight, age and breeding history on canine sperm quality parameters measured by the hamilton-thorne analyser, Reproduction in domestic animals 42 (2007) 143–148.
- [87] D. Banaszewska, K. Andraszek, E. Zdrowowicz, A. Danielewicz, The effect of selected staining techniques on stallion sperm morphometry, Livestock Science 175 (2015) 128–132.
- [88] E. Lukaszewicz, A. Jerysz, A. Partyka, A. Siudzińska, Efficacy of evaluation of rooster sperm morphology using different staining methods, Research in veterinary science 85 (2008) 583–588.
- [89] D. Banaszewska, M. Czubaszek, R. Walczak-Jedrzejowska, K. Andraszek, Morphometric dimensions of the stallion sperm head depending on the staining method used, Bulletin of the Veterinary Institute in Pulawy 59 (2015) 263–270.
- [90] K. Andraszek, D. Banaszewska, B. Biesiada-Drzazga, The use of two staining methods for identification of spermatozoon structure in roosters, Poultry science 97 (2018) 2575–2581.

The authors declare that they have no conflict of interest.