Research Article

SWFT: Subbands wavelet for local features transform descriptor for corneal diseases diagnosis

Samer K. AL-SALIHI[1,*]**, Sezgin AYDIN**[2]**, Nebras H. GHAEB**[3]

[1]Electrical-Electronic and Computer Engineering Department, Aksaray University, Aksaray, Turkey
[2]Department of Natural and Mathematical Sciences, Tarsus Unversity, Mersin, Turkey
[3]Biomedical Engineering Department, Al-Khawarizmi College of Engineering, Baghdad University, Baghdad, Iraq

Abstract: Human cornea is the front see-through shield of the eye. It refracts light onto the retina to induce vision. Therefore, any defect in the cornea may lead to vision disturbance. This deficiency is estimated by sets of topographical images measured, and assessed by an ophthalmologist. Consequently, an important priority is the early and accurate diagnosis of diseases that may affect corneal integrity through the use of machine learning algorithms. Images produced by a Pentacam device can be subjected to rotation or some distortion during acquisition; therefore, accurate diagnosis requires the use of local features in the image. Accordingly, a new algorithm called subbands wavelet for local features transform (SWFT) which is mainly based on the algorithm of a scale-invariant feature transform (SIFT) has been developed. This algorithm uses wavelets as a multiresolution analysis to produce images with different scales instead of using the difference of Gaussians as in the SIFT algorithm. The experimental results on the corneal topography dataset indicate that the proposed SWFT outperforms the baseline SIFT algorithm.

Key words: Computer-aided diagnosis, feature extraction, machine learning, support vector machines, wavelet transforms

1. Introduction

Currently, computer-aided diagnosis (CAD) is closely related to the area of medical imaging [1], which is a sophisticated procedure in medicine that assists physicians by using computer algorithms to aid the image interpretation process. Machine learning algorithms have become one of the most prevalent disease diagnostic methods [2]. Obviously, the structure of any CAD consists of the preprocessing phase, segmentation phase, feature extraction phase, and finally, classification phase, as exhibited in Figure 1, which usually involves a machine-learning algorithm. Such algorithms could be used to diagnose human eye diseases, and the most sensitive, aggressive, and dangerous eye diseases are corneal diseases. The cornea is the front part of the eye which controls visual focus, carries more than two-thirds of the visual power, and protects the eye from external foreign elements with its clear and smooth structure [3].

Corneal diseases are serious conditions that cause distortions, clouding, and eventually blindness. There are numerous kinds of corneal diseases. To diagnose them, tests are carried out under ophthalmologists' supervision through sophisticated tools such as corneal topography, which can be obtained by using a device called Pentacam [4]. Pentacam scans the eye and provides the corneal topography as an output that contains

---

*Correspondence: samer.kais@uor.edu.krd

```
┌─────────────────────────┐
│      Preprocessing      │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│  Segmentation/ Detection│
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│    Feature Extraction   │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│      Classification     │
└─────────────────────────┘
```
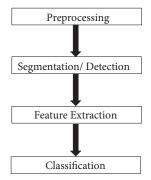
Figure 1. A CAD system.

many topographical parameters as well as four maps (images) that give information about the front and back elevation, axial-based curvature, and the thinness of the cornea, utilized by specialists to make an optimum diagnosis of the cornea condition. Despite the abundance of information provided by Pentacam, it also represents a challenge, an obstacle, and adds difficulty for diagnosing the cornea conditions, where reaching clinical decisions regarding cornea requires reviewing huge numbers of parameters and maps, and this revision needs to be done by ophthalmologists. On the other hand, a lot of human factors are considered in the decision, especially the experience of the ophthalmologist. Moreover, the biggest challenge is the huge topographical parameters which cannot be memorized by surgeons or ophthalmologists while reaching clinical decisions is more related to the borderline of values for the parameters and maps [5]. Therefore, in this work, we propose to take advantage of machine learning techniques to diagnose corneal diseases by extracting the important texture characteristics from medical corneal images and passing them to the support vector machine (SVM) [6] to diagnose the corneal condition. The efficiency of the diagnostic performance may be negatively affected by poor medical images that may be produced for some reason or other such as the noise from the device itself, or the movements of the patient during capturing the image causing a slightly rotating of the image, diverse distances of the sensor, or a different point of view [7]. Therefore, corneal disease diagnosis requires extracting local features that are not much affected by such factors; for instance, the features applied in plenty of real-world implementations [8], like object recognition [9, 10], object tracking [11], and image retrieval [12, 13]. Consequently, for feature extraction, we propose an improved algorithm for extracting the local features based on the SIFT algorithm [14], called subbands wavelet for local features transform (SWFT). As known, SIFT works in two successive phases: the detector and the descriptor phases [15]. This work focuses on developing the first phase by constructing various scales of the image σ by decomposing the image into different subbands using wavelets transform. In the literature, some researchers have improved the SIFT algorithm, which was introduced by Loewe [14]. In 2004, Ke and Sukthankar [16] treated the normalized gradient subimage by using the principal component analysis (PCA) technique. They showed that using such descriptors for image retrieval yields an increase in accuracy. Mikolajczyk and Schmid [17] introduced a gradient location gradient graph (GLOH) and used the SIFT algorithm to extract features from eight angle directions producing 272 histogram bins that were reduced using PCA. Based on the strength of the available detectors and descriptors in 2006 and earlier, Herbert et al. [18] presented a useful scale- and rotation-invariant method for many computer vision tasks that depend on the integral filter for image convolutions. Morel and Yu [19] suggested a strategy based on what was ignored by the SIFT algorithm, which simulated all image views that may occur by shifting latitude and longitude angles of

the camera axis. They attempted to mathematically prove that their algorithm is fully affine-invariant. Tang and Tang [20] proposed an approach to improve the representation of the local image descriptor by augmenting the histogram, which led to high performance in matching real-world images after many levels of geometric and photometric transformations.

The contributions of this work can be briefed as:

1. An unsupervised clinical decision regarding cornea is reached by exploiting machine-learning techniques.

2. A subbands wavelet for local features transform (SWFT) algorithm dealing with topographical maps and their board lines and using wavelet transform to construct different scale space of images was proposed.

3. The SWFT is capable of local feature extraction from the medical images and thus improves system performance under different conditions and using various amounts of training and testing set.

4. To investigate the contributions, a new large dataset is collected, reviewed, and labeled by ophthalmologists, encompassing different cornea conditions.

This paper is organized as follows. Section 2 explains the SIFT and wavelet transform algorithms. Section 3 displays the technique of the proposed method. Section 4 presents the assessment strategies. Section 5 is about the dataset description. Section 6 shows the practical experiments. Section 7 presents the results and discussions, and Section 8 presents the conclusions.

2. Methods

2.1. The SIFT algorithm

SIFT stands for scale-invariant feature transform [14]. This particular feature detector addresses the problem of matching features with changing scale and rotation [21]. It is primarily used for recognizing objects [22]; the recognition system is trained with images of the object and tries to recognize it from new images given in different scenes and with different sides of view and scales. This algorithm, which is one of the best approaches for feature matching, has been tried by a number of researchers and has yielded very successful results [23, 24]. SIFT was initially developed by Lowe in 2004, there have been many variants since that time. The approach of this transform is first to create a scale space of images using the Gaussian function with increasing values of Sigma, as in (1). It creates a set of progressively blurred Gaussian images and then takes the differences of each pair of images in the scale to get a difference of Gaussian pyramid [25], as expressed in (2), which could be an exceptionally great approximation to the function LoG.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y). \tag{1}$$

$$DOG(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma). \tag{2}$$

where L could be a blurred image, G is the Gaussian blur operator, I is an image, and σ could be the "scale" parameter. The next step is to find local extrema in this scale space, which helps to remove the scale uncertainty and find key points or features reliably. Once these key points are found, it is straightforward to calculate the vector by creating a histogram of gradient directions from the 16×16 neighbors around those key points, as shown in Figure 2 which shows the scale space of images that apply a Gaussian, thereafter, computing the difference of Gaussian array via the difference between every pair of images in the stack.
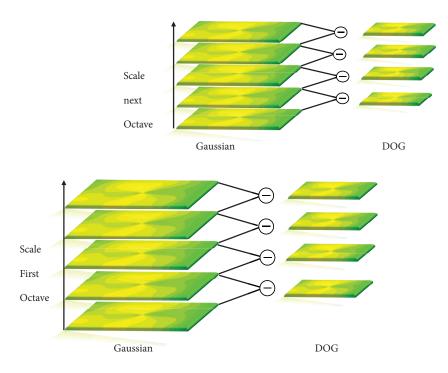
Figure 2. Scale space of an image in different octave used to calculate the difference of Gaussians (DOG).

Ultimately, the keypoints (minima or maximum peak) can be localized more precisely within the image by fitting a quadratic surface to the values in the neighborhoods, as expressed in (3), and the peak of that surface that gives a subpixel precision of the location of the peak is found [26]. The offset of extremum uses finite differences for derivatives, as expressed in (4).

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} X^T \frac{\partial^2 D}{\partial x^2} \tag{3}$$

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} * \frac{\partial D}{\partial x} \tag{4}$$

For selecting canonical orientation, the features are found around all key points and then a histogram of the local gradient directions is computed at the selected scale to disclose the peak (highest pin in the histogram), it is assigned to be the angle of that whole patch, useful to remove the rotational uncertainty or degree of freedom of these points, see Figure 3. At each pixel, A_{ij}, the image gradient magnitude, M_{ij}, and orientation, R_{ij} are computed using pixel differences [14] as in 5 and 6.

$$M_{ij} = \sqrt{(A_{ij} - A_{i+1,j})^2 + (A_{ij} - A_{i,j+1})^2} \tag{5}$$

$$R_{ij} = atan2(A_{ij} - A_{i+1,j}, A_{i,j+1} - A_{ij}) \tag{6}$$

where A_{ij} and $A_{i+1,j}$ are two consecutive pixels in the x-direction, and A_{ij} and $A_{i,j+1}$ are two consecutive pixels in the y-direction.
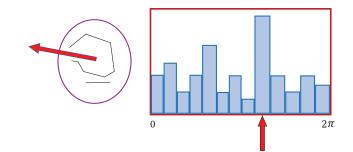
Figure 3. The orientations histogram of an image region used to determine its main orientation which is the highest bin (peak).

Finally, each key point specifies the stable 2D coordinate location, scale, and orientation (X, Y, scale, orientation), which represent the key point features that have been extracted from the image where the length of the vector corresponds to the scale and the angle of the vector, thereafter applying a threshold on the image gradients and sampled over 16×16 array of locations in the scale space. The feature vector is constructed by selecting the 16×16 array around the key point and the gradient directions of all those points are computed and quantized into eight different directions in a 4×4 array, producing 128 elements of a very rich features vector [27], see Figure 4.

8x8 set samples
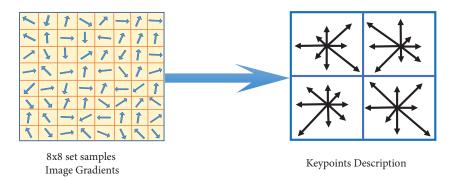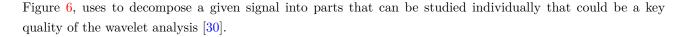Image Gradients

Keypoints Description

Figure 4. The scale and orientation with only eight directions in total.

2.2. Wavelet transform

An image is a group of pixels whose values represent the intensities. These values carry information about the extent of illumination (energy) that exists in pixel. Because the image contains an object (s), changes in intensity occur due to the difference in the amount of illumination between the object and the background of the image which represents an edge. Edge information is important in that it indicates a change in image content. Therefore, representing the image in the frequency domain by scanning a horizontal line from the image will contain different oscillations that change suddenly due to the shift of the scan from the smooth areas to the texture areas in the image or vice versa, as given in Figure 5. Hence, to analyze images precisely, we need to exploit wavelet transform techniques that have characteristics of localizing in time and frequency [28].

Wavelet is a finite duration wave-like oscillation that starts from zero, increases and decreases by the same amount and thus returns to zero. Therefore, it has zero mean, and has diverse forms and sizes [29], see
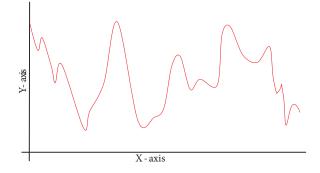
Figure 6, uses to decompose a given signal into parts that can be studied individually that could be a key quality of the wavelet analysis [30].
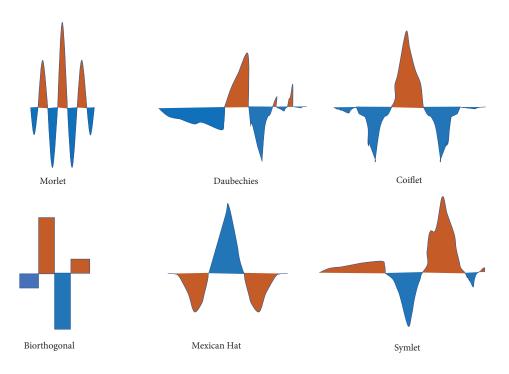

Figure 5. Signals frequently exhibited.

Figure 6. Different sizes and shapes of wavelet.

Set of coefficients are composed by applying the wavelet, which represents the components of signal analysis in different scales and translation, where scale refers to either increasing or decreasing the wavelet pitch, as expressed in 7.

$$\Psi\left(\frac{t}{s}\right) \quad s > 0 \tag{7}$$

where S is a positive value representing the scaling factor, which can be controlled through stretching or shrinking the wavelet [31]; the wavelet is stretched out (high scale) for low frequency and shrunk (low scale) for high frequency. Figure 7 illustrates the inverse relationship between the frequency and scale, which is called the center frequency of the wavelet [32].
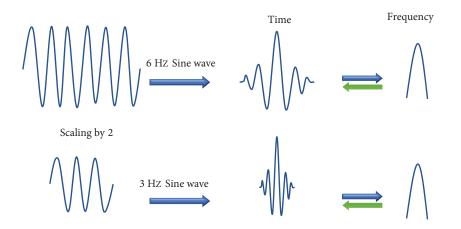
Figure 7. Constant of proportionality of scale and the frequency.

Translation indicates shifting the wavelet forward in time on the entire signal, i.e. shifting and centering at K for acquiring the sought information [33], as expressed in 8.

$$Translation = \varnothing(t - k) \tag{8}$$

The transforms contrast with the way of translation and scaling such as discrete and continuous wavelet transforms.

2.2.1. Continuous wavelet transforms (CWT)

CWT is a mathematical function that expresses the time-frequency analysis, where translation and scale factors are changing continuously [34], as illustrated in Figure 8. The idea is the wavelet constructs a less resolution in time within the condition of low frequencies whereas more resolution in high frequencies [35].
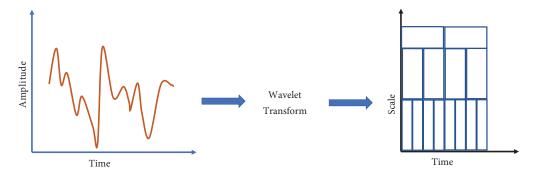

Figure 8. Function of scale or frequency and time.

The wavelet moves along the signal and is repeated over and over again with different scales to compose various coefficients [36].

2.2.2. Discrete wavelet transform (DWT)

It is a mathematic function that helps to represent images with a little number of coefficients that can be dealt with separately, which makes it a suitable tool for image compression [37]. It is established on multiresolution

analysis, where the scale is generated easily by increasing the value of the base scale which is represented by 2^j [38], as in 9. The wavelet is applied over the entire signal, mathematically, and the translation is represented as an integer value, as expressed in 10.

$$2^j \qquad (j = 1, 2, 3,)$$ (9)

$$2^j m \qquad (m = 1, 2, 3...)$$ (10)

The DWT process is a multiresolution signal analysis. Conceptually, to apply 2D-DWT on a given image, first it should pass through a low-pass filter and high-pass filter along on the columns that yield $h\varphi(-n)$ and $h\psi(-n)$ subbands which represent the columns filtering phase (CFP), where n represent the columns of the given image. The bandwidth of the signal essentially gets halved in each of these subbands; the original signal is split up into two subbands, which yields redundant samples. Therefore, a subsampling is applied by a factor of two where out of two consecutive samples only one is picked up while the other is discarded. For the next-level, the same technique is applied along the rows. Therefore, it splits each subband again into two other subbands, $h\varphi(-m)$ and $h\psi(-m)$ subbands, which represent the rows filtering phase (RFP), where m represents the rows of the given image, as shown in Figure 9.

Consequently, each of these subbands will contain half of the samples, so subsample by a factor of two will be getting four outputs $W\varphi(j)$, $W_\psi^H(j)$, $W_\psi^V(j)$, and $W_\psi^D(j)$, referring to horizontal extracted edges, which represent the low-pass and high-pass subbands in CFP and RFP phases, respectively. The V superscript refers to vertical extracted edges that represent the high-pass and the low-pass subbands in the CFP and RFP phases, respectively whereas D superscript refers to diagonal extracted edges that represent the high-pass and the high-pass subbands in the CFP and RFP phases, respectively [39].
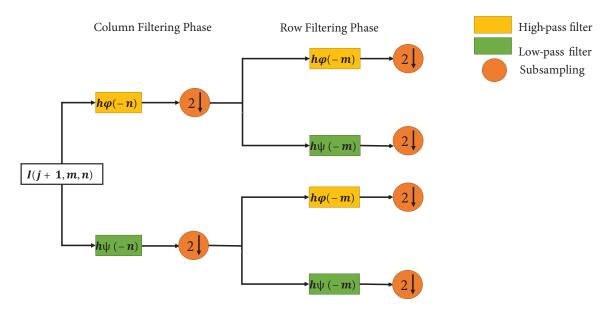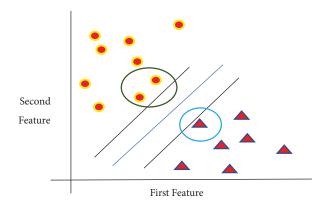

Figure 9. Wavelet filter bank for one-level image decomposition.

2.2.3. Support vector machine

The SVM is an algorithm [40] that draws a decision boundary near the extreme points in the dataset known as a hyperplane, where the features should be chosen to decide where to draw a good decision boundary [41]. The line might draw over different locations and directions. Therefore, it could incorrectly classify classes since there is no optimal decision boundary. Thus, data points help the margin push up against all points that are close to the opposing class known as support vectors [42], see Figure 10.

The support vectors are the most important aspect in SVM, whereas training examples are ignorable; for instance, samples of the first class look like second class or vice versa. Such an issue requires a classifier that looks for extreme points and sets margins based on these support vectors. The shortest distance to the closest positive point called D+, shortest distance to the closest negative point called D-, and separating hyperplane that is the D+ plus the D- called margin [42], see Figure 11.

The line (decision boundary) that segregates the two classes is commonly referred to as a hyperplane because SVM can be used in multidimensional datasets. The data points are referred to as vectors since they have coordinates within the space of data [43]. This type is known as linear support vector machines (LSVM) because the classes are linearly separable. On the other hand, the data might look impossible to separate two classes by using a single line. Consequently, to draw a hyperplane easily, the data need to transform into a high-dimensional space. For example, by using a simple polynomial function, a parabola can be obtained from one-dimensional space. [44], see Figure 12.
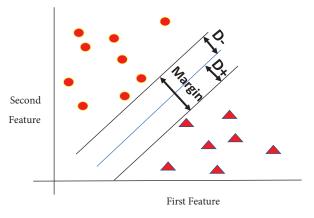

Figure 10. The support vector points samples from two classes.

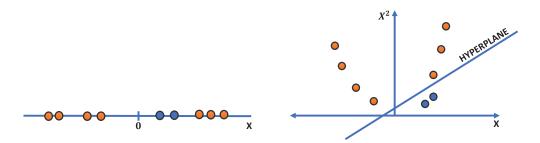**Figure 11**. The margin and D-, D+ with samples from two classes.

Figure 12. Transform nonlinear separable data into 2-dimensional data.

The dataset with two dimension is simply treated with the same method by using a function to transform data from two-dimensional to three-dimensional feature space, where the only problem with transformation is that it is computationally expensive, see Figure 13. This issue is handled by using a kernel trick to reduce the computational costs [45].
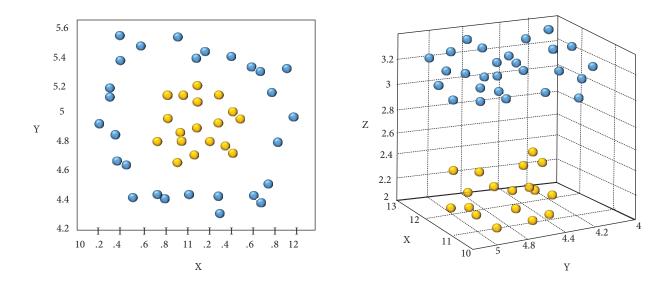

Figure 13. Transform 2-dimensional data into 3-dimensional data.

Kernel trick is a function that takes inputs vectors in the original space and returns the dot product of the vectors in the feature space. Kernel function is also referred to as kernel trick. It finally returns a similarity score between any two points, which can be considered a metric of closeness [46, 47].

There are many SVM kernels, such as linear kernel which is a straight line hyperplane as mentioned in Figure 11. In addition to linear kernel there are other kernels which are summarized below.

1. Radial basis function (RBF) kernel: RBF helps to avoid the expense of additional features and it can also build and plot very complex decision boundaries as a curve shape. The curve shape comes out when the RBF function moves away from the support-vector (center). It provides an opportunity to draw a hyperplane that aptly separates points [48]. It can be calculated by equation 11

$$K(X,Y) = exp(-\gamma|X - Y|2)\gamma > 0 \tag{11}$$

where X and Y are vectors in the input space, and γ is the gamma parameter.

2. The polynomial kernel: Polynomial kernel has the ability to reach the interaction characteristics derived from combining features used to learn patterns. It is worth noting that polynomial kernel with degree one is a linear kernel [49]. It can be applied by the mathematical function expressed in equation 12.

$$K(X,Y) = (X \cdot Y + C)^d \gamma > 0 \tag{12}$$

where X and Y are input space vectors, γ is the gamma parameter, C is the penalty parameter, and d is the polynomial degree.

3. Sigmoid kernel and pairwise metrics: The sigmoid kernel is originated from the neural networks field which is called multilayer perceptron (MLP) kernel. That is why it is very popular and interesting when utilized in SVM, which is considered to be the same as two-layer perceptron neural network. It is also known as hyperbolic tangent [50], see equation 13.

$$K(X,Y) = tanh(\gamma X \cdot Y + C) \tag{13}$$

In addition to the kernel, there are some important optimization parameters such as;

1. The penalty parameter: It is usually symbolized with letter C. It is used to hegemonize systematic outliers and to control errors, i.e. it informs the algorithm of the amount of attentiveness about misclassified points. Accordingly, the large value of the penalty parameter requires high attention to correctly classify the training set instead of leaving plenty of room for the test set [51].

2. The gamma parameter: It is a hyperparameter that should be set prior to training the model. As previously mentioned, the RBF gets weakened when moved away from the center causing the curvature of boundaries. Gamma is the parameter that sets the desired amount of curvature and dissipation. Gamma is the parameter that determines the dissipation swift and the quantity required for the curvature. Accordingly, a larger gama drops the influence of any individual center, and causes more curvature as well [52]. Gamma is one over the number of features in the dataset as shown in equation 14.

$$\gamma(gamma) = 1/N \tag{14}$$

where N is the number of features.

3. Proposed method

The proposed subbands wavelet for local features transform (SWFT) algorithm, which is based on the standard SIFT algorithm, accepts the same inputs as the original algorithm. The essential function for the standard SIFT algorithm is to obtain the difference of Gaussians (DoG) images in which edges and blobs can be found. The image needs to be compared at different scales (image pyramid) to find key points by smoothing the image with a Gaussian filter. Our proposed algorithm uses wavelet transform subbands at different levels as a tool to create scale-space of image to be used to extract the local features, and it also sets the contrast threshold parameter value to 0.002. The image is decomposed at level four using the 2D-DWT (Symlets) as a tool for multiresolution decomposition, which provides a robust analysis of the image based on the important aspects of scaling and shifting. Moreover, the DWT is able to overcome the perspective shifts caused by changing the camera angle, because it is the most perfect method used for localized time and frequency to analyze the image. Thus, it facilitates calculating the keypoints, which are mostly represented as the anomalies of a particular pixel intensity value against its neighbors' intensity values. Scale-space of an image is generated as follows, after decomposing the image at level one using DWT, the first scale is obtained by reconstructing the image without the level one details (horizontal, vertical, and diagonal) whereas the next scale is obtained by decomposing the image at level two and then reconstructing it without the details of levels one and two, and so on. Similar considerations hold for the other octaves. One-level image decomposition is illustrated in Figure 14.
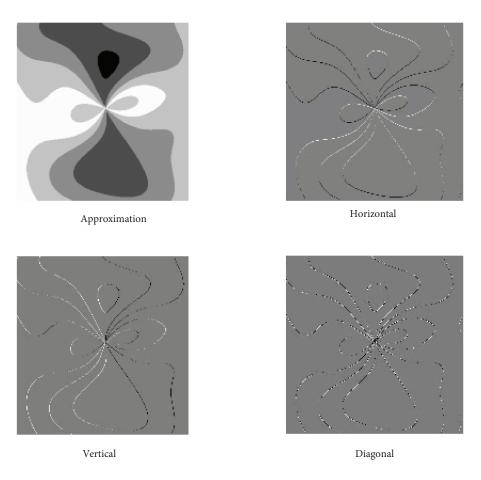
Approximation

Horizontal

Vertical

Diagonal

Figure 14. Decomposition of an image using first-level wavelet transform.

4. Evaluation

The performance of the algorithm has been evaluated using various metrics. Firstly, since it is sought to enlarge the value of true positives and to decrease the number of false positives, adopting the following computation, the proportion of cases that are genuinely positive and algorithm results that are also positive (recall), the proportion of the overall relevant results that are correctly classified by the system (precision), and the proportion of true results (positives and negatives) among all cases are examined (accuracy) [52], as displayed in 15–17, respectively.

Secondly, paying attention to examination of true negative rates, by measuring the following, the proportion of the genuinely negative cases and the algorithm test results that are also negative (specificity), and the receiver operating characteristic (ROC) curve, which is commonly used in the literature that can define false detection rates (1–precision) [53, 54] are calculated, as denoted in 18 and 19, respectively.

$$Sensitivity(recall) = \frac{True\ Positive(TP)}{True\ positive(TP) + False\ Negitive(FN)} \quad (15)$$

$$percision = \frac{True\ Positive(TP)}{True\ positive(TP) + False\ positive(FP)} \quad (16)$$

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{17}$$

$$Specificity = \frac{True\ Negitive(TN)}{True\ Negitive(TN) + False\ positive(FP)} \tag{18}$$

$$1 - Precision = \frac{False\ positive(FP)}{False\ positive(FP) + True\ positive(TP)} \tag{19}$$

5. Dataset

The dataset is images obtained by scanning the cornea called corneal topography which is used by ophthalmologists for the examination of the eye in clinics. Each patient's eye data, which contain four corneal maps and a set of parameters, are stored separately. This information has an important role in the diagnostic process. The dataset is collected using a device called Pentacam, which is based on imaging Scheimpflug. The camera scans the eye from several angles in a circular path to provide information about the anterior and posterior segments of the cornea in the shape of maps and provides a fast screening report. The Oculus Pentacam can be updated and upgraded to suit the needs of the user [55]. The dataset comprises 1848 images of different maps such as sagittal, corneal thickness, elevation front, and elevation back maps, collected and labeled scientifically with the help of ophthalmologists. Each image has a resolution of $351 \times 336 \times 3$.

6. Practical experiments

In this paper, corneal topography is used as data of the human cornea, consisting of four different maps, as shown in Figure 15.

The first map, called the sagittal map, provides information about the astigmatism disease of the cornea. The second map, called the corneal thickness map, reads the extent of the corneal thinness in various places of the cornea and gives the coordinates of the pupil center. The third and fourth maps are the elevation front and elevation back maps, which provide information about the front and the back cornea surfaces condition, as shown in Figure 16 and therefore could be used in diagnosing different corneal diseases. Each map is analyzed separately to diagnose the condition as normal or abnormal using the SVM algorithm with "linear" kernel function, C=100, and gamma. Accordingly, it can diagnose many morphological and clinical findings such as cornea within rule astigmatism and oblique astigmatism diseases from the first map, thin corneas disease from the second map, and lastly, isolated land and like tang shape diseases from the third and fourth maps. Different maps provide an opportunity to examine the algorithm in a big range of data. Therefore, we perform four types of examinations for a more accurate and reliable diagnosis, to discover the performance of the proposed method, and to achieve a perfect comparison with the standard SIFT algorithm.

7. Results and discussion

7.1. Results

This section compares the SWFT achievements to those of the standard SIFT in controlled experiments. The evaluation of the diagnosis of the different corneal diseases is calculated as well. Support vector machine (SVM) is exploited as a classification tool to diagnose cornea diseases. In the real world, there are a variable number of patient cases available; therefore, to build a system that simulates reality and to assess the effect of classification
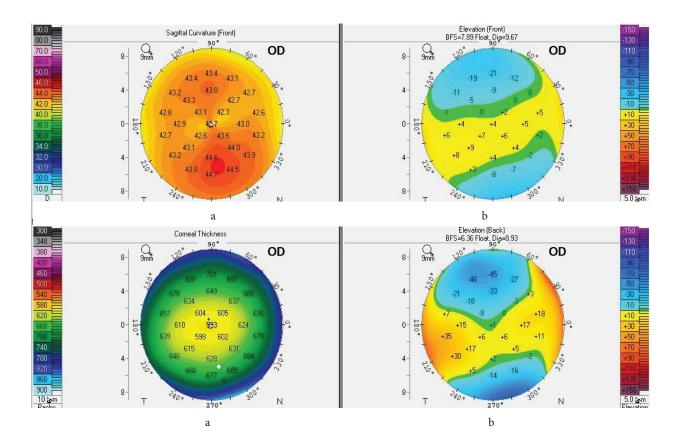
Figure 15. Human corneal topography from Pentacam device. a) Sagittal map, b) corneal thickness map, c) elevation front map, d) elevation back map.

accuracy when the training size changes, the practical experiments are conducted in four different scenarios as follows: 10% of the dataset for training and 90% for testing, 50% of the dataset for training and 50% for testing, 70% of the dataset for training and 30% for testing, or 90% of the dataset for training and 10% for testing. The results are obtained using cross-validation for each strategy. Tables 1–4 show the comparison results between the proposed and the standard algorithms using the four different strategies.

Table 1. Achievements of the algorithm on the sagittal map to diagnose the corneal-astigmatism using all strategies (%).

Measurements	Strategies of dataset							
	10%90%		50%50%		70%30%		90%10%	
	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT
Recall	67.47	89.47	80.39	84.15	85.48	91.37	83.86	88.29
Precision	61.90	80.95	76.63	78.70	81.53	81.53	78.86	85.56
Accuracy	69.23	84.61	77.17	80.30	82.20	85.59	80.11	85.99
1-precision	38.09	19.04	23.36	21.29	18.64	18.64	21.13	14.43
Specificity	63.60	80.00	73.68	76.28	78.57	80	76.43	83.43

Figure 16. Human corneal disease examples. a) Astigmatism, b) thin cornea, c) like tongue, d) isolated land shape.

The proposed algorithm has also been tested using two other classifiers, the naive Bayes and the K-nearest neighbors (KNN) classifiers, as it applied using the four maps of corneal topography, taking into account the balancing of data in case a big gap occurs between the numbers of the instances in classes. The accuracy of the diagnosis is calculated and the cross-validation error is estimated for both classifiers using 10-fold cross-validation. The results are shown in Table 5.

Our proposal has also been tested on three common datasets, namely Skin Cancer MNIST: HAM10000 [56], Chest X-Ray Images (Pneumonia) [57], and MIAS MiniMammographic Databases [58]. It should be

Table 2. Achievements of the algorithm on the corneal thickness map to diagnosing the corneal-thinning disease using all strategies (%).

Measurements	Strategies of dataset							
	10%90%		50%50%		70%30%		90%10%	
	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT
Recall	71.73	77.48	76.19	80.76	79.36	87.09	75.0	85.95
Precision	65.02	69.48	70.79	74.33	74.62	76.05	68.18	77.27
Accuracy	66.39	69.05	74.22	78.22	75.20	80	70.0	81.81
1-precision	34.97	30.51	29.20	25.66	25.37	23.94	31.80	22.72
Specificity	60.0	58.86	72.5	76.03	70.68	73.01	65.0	79.16

Table 3. Achievements of the algorithm on the corneal elevation front map to diagnosing the cornea-isolated-land disease using all strategies (%).

Measurements	Strategies of dataset							
	10%90%		50%50%		70%30%		90%10%	
	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT
Recall	94.66	96.47	96.59	100	96.15	100	96.96	100
Precision	82.60	85.09	95.50	97.19	93.45	97.19	94.11	97.14
Accuracy	82.33	85.57	93.69	97.74	91.72	97.74	92.85	97.67
1-precision	17.30	14.90	4.49	2.80	6.54	2.80	5.88	2.85
Specificity	53.71	59.32	82.60	89.79	75.86	89.66	77.77	88.88

Table 4. Achievements of the algorithm on the corneal elevation back map to diagnosing the like tange diseases using all strategies (%).

Measurements	Strategies of dataset							
	10%90%		50%50%		70%30%		90%10%	
	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT
Recall	91.59	92.21	93.75	96.12	90.58	95.40	96.42	100
Precision	79.56	82.11	78.94	81.57	84.61	91.20	90.00	93.33
Accuracy	75.00	77.63	83.40	86.30	84.61	91.66	91.66	95.83
1-precision	20.43	17.88	21.05	18	15.38	8.79	10	6.66
Specificity	15.15	18.33	71.68	75	75.86	85.96	85.0	90.00

noted that the Skin Cancer MNIST dataset contains several categories. Therefore, it is classified on the basis of category. For example, benign keratosis (bkl), dermatofibroma (df), melanocytic nevi (nv), and basal cell carcinoma (bcc). For comparison and testing, each of the two categories is selected separately whereas the Chest X-Ray and MIAS MiniMammographic datasets are labeled as normal and abnormal. The proposed algorithm and the standard SIFT are used to be compared. Finally, the accuracy of the diagnosis is calculated using three classifiers, the Naive Bayes, KNN, and SVM classifiers, and the results are shown in Table 6.

Table 5. Accuracy of cornea disease diagnosis by applying the proposed algorithm on all corneal topography maps (%), and loss cross-validation error.

Maps \Methods	Naive Bayes classifier				KNN classifier			
	Accuracy (%)		Loss cross-validation		Accuracy (%)		Loss cross-validation	
	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT
Sagittal	81.0	98.7	0.29	0.25	82.8	92.1	0.34	0.13
Corneal thickness	47.0	92.8	0.51	0.34	73.1	78.5	0.34	0.26
Elevation front	98.7	99.7	0.16	0.02	98.1	98.7	0.32	0.01
Elevation back	82.6	99.8	0.24	0.01	91.4	92.4	0.17	0.14

Table 6. Comparison of accuracy of diagnosis between the proposed algorithm and standard SIFT tested on different datasets and employing different classifiers (%).

	Chest X-ray		MIAS MiniMammographic		Skin cancer MNIST					
Classifiers	Normal vs abnormal		Normal vs abnormal		Bkl vs df		Bcc vs bkl		Bcc vs nv	
	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT	SIFT	SWFT
SVM	93.4	95.8	73.0	72.5	73.5	75.8	70.7	75.4	89.1	94.9
Naïve Bayes	91.0	94.8	81.4	83.3	90.7	94.2	89.0	96.5	95.5	98.5
KNN	94.0	93.1	73.1	73.9	80.7	85.7	81.5	84.5	90.0	94.5

7.2. Discussion

The previous section showed that the SWFT algorithm, in general, has significantly high accuracy and reliability compared to the standard SIFT local feature descriptor. From the tables above, the probability of correct diagnosis (recall) using the proposed algorithm with various strategies gives more realistic results than the standard algorithm, where the recall average values for diagnosing cornea diseases for all strategies and maps are 91.54 and 86.01 using SWFT and SIFT, respectively. The statistics tell that the proposed algorithm detects better the relevant diseases among the retrieved instances via testing using all the strategies, where the precision average values for diagnosing cornea diseases using all maps are 84.29 and 79.76 using SWFT and SIFT, respectively. The accuracy metric refers to the correct predictions of the proposed algorithm. The reports indicate that the SWFT algorithm achieved the highest performance among all cases examined to detect both true results (positive and negative) for all the strategies, where the average accuracy values are 84.85 and 80.61 using SWFT and SIFT algorithms, respectively.

The SWFT algorithm achieved a lower percentage of the false detection cases relative to the overall cases compared with the standard SIFT algorithm, where the averages of the false detection values (1-precision) are 15.68 and 20.22 using SWFT and SIFT algorithms, respectively. The results also showed a system using the SWFT algorithm achieves a higher percentage in detecting healthy corneas that are correctly diagnosed as not having any disease than using the SIFT algorithm via using all strategies, the averages of specificity are 75.23 and 68.63 using SWFT and SIFT algorithms, respectively. Figure 6 indicates that the proposed algorithm outperforms the standard algorithm even when tested with different datasets, except in very rare cases. The reason might be attributed to the nature of the data trained. Despite the minor errors, it is still an algorithm that has an explicit positive impact on improving the efficiency of the system performance, especially with the corneal topography datasets.

Figures 17a–17d plot recall and precision to illustrate the comparison results of practical experiments between the SIFT and SWFT algorithms using sagittal, corneal thickness, elevation front, and elevation back maps, which clearly shows how the proposed algorithm outperforms its counterpart.
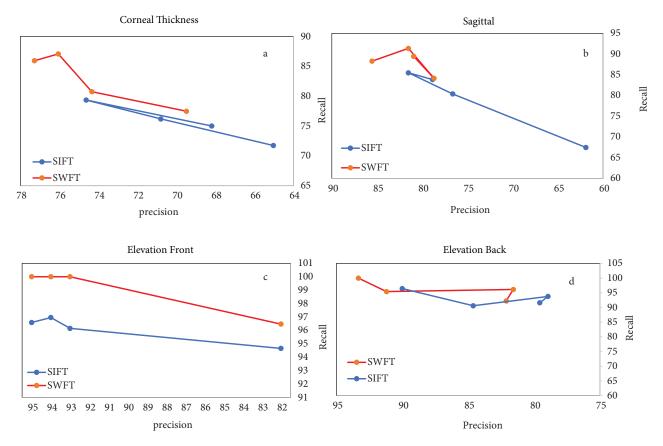

Figure 17. Ploting the comparison of the SIFT and SWFT algorithms performance with different strategies of training data (10%, 50%, 70%, 90%) using a) sagittal map, b) corneal thickness map, c) elevation front map, d) elevation back map.

8. Conclusion and future work

The use of computer-aided tools is essential for increasing accuracy and rapidity of diagnosing corneal diseases. The most significant part of the diagnostic systems is extracting the features from medical images, which is one of the most powerful techniques to increase the reliability of such systems. Accordingly, this paper focused on extracting local features from medical images using the proposed SWFT algorithm, which is an alternative representation for the local features descriptor for the SIFT algorithm. Precise extraction of the features considerably helps eliminate the diagnosis mistakes due to erroneous data that may occur for any reason and thus maintain the safety of patients' eyes, preventing damage, and aid specialists make the right decisions. Compared to the standard representation, the expensive experiments indicate that the SWFT algorithm is more distinctive, a significant achievement for real-world applications, and it is efficient in extracting features from images thereby increasing the efficiency of the system. Employing wavelet transform as a multiresolution analysis to produce images with different scales has a significant and clear impact on the extracting more distinct features that boost the system's ability to outperform the standard SIFT descriptor and achieve a

high performance in distinguishing between categories. Ultimately, the use of machine-learning algorithms to diagnose corneal diseases using corneal topography has yielded impressive results, so we recommend using them by ophthalmologists to help giving more accurate and immediate decisions.

As a future plan, satisfactory results led us to think carefully about the development of our proposal. A different set of SVM parameters will be manipulated and its effectiveness will be observed in developing the proposed method. The future work will also include developing the algorithm for generating image scales by using the GAN model instead of using the difference of Gaussians, where the parameters of the GAN networks can be manipulated to generate a realistic image suitable for use as different scales of image.

References

[1] Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized Medical Imaging and Graphics 2007; 31 (4-5): 198-211.

[2] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in Medicine 2001; 23 (1): 89-109.

[3] Yang AY, Chow J, Liu J. Focus: sensory biology and pain: corneal innervation and sensation: the eye and beyond. The Yale Journal of Biology and Medicine 2018; 91 (1): 13.

[4] Nemeth G, Vajas A, Kolozsvari B, Berta A, Modis Jr L. Anterior chamber depth measurements in phakic and pseudophakic eyes: Pentacam versus ultrasound device. Journal of Cataract and Refractive Surgery 2006; 32 (8): 1331-1335.

[5] Fontes BM, Ambrósio Jr R, Jardim D, Velarde GC, Nosé W. Corneal biomechanical metrics and anterior segment parameters in mild keratoconus. Ophthalmology 2010; 117 (4): 673-679.

[6] Cortes C, Vapnik V. Support-vector networks. Machine Learning 1995; 20, 273-297. doi: 10.1007/BF00994018

[7] Sinjab MM. Corneal Tomography in Clinical Practice (Pentacam System): Basics and Clinical Interpretation. Jaypee Brothers, Medical Publishers Pvt. Limited; 2018.

[8] Verma M, Raman B. Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval. Multimedia Tools and Applications 2018; 77 (10): 11843-11866.

[9] Arya KV, Rajput SS, Upadhyay S. Noise-robust low-resolution face recognition using sift features. In: Computational Intelligence: Theories, Applications and Future Directions-Volume II 2019; 1: 645-655.

[10] Kamdi VD, Raut AD, Pardakhe V, Jagtap VA. A study of SURF features for" face recognition. International Journal of Electronics, Communication and Soft Computing Science and Engineering 2018: 90-94.

[11] Dai-Hong J, Lei D, Dan L, San-You Z. Moving-object tracking algorithm based on PCA-SIFT and optimization for underground coal mines. IEEE Access 2019; 7: 35556-35563.

[12] Cheng Q, Shao K, Li C, Li S, Li J et al. Distributed system architecture for high-resolution remote sensing image retrieval by combining deep and traditional features. Image and Signal Processing for Remote Sensing 2018; 1: 1078918-1078919. doi: 10.1117/12.2323310

[13] Xie B, Qin J, Xiang X, Li H, Pan L. An image retrieval algorithm based on gist and SIFT Features. IJ Network Security 2018; 20 (4): 609-616.

[14] Lowe DG. Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision 1999; Kerkyra, Greece; 1999. pp. 1150-1157.

[15] Morel JM, Yu G. On the consistency of the SIFT method. 2011.

[16] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors. InProceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2004; 2: II-II.

[17] Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J et al. A comparison of affine region detectors. International Journal of Computer Vision 2005; 65 (1-2): 43-72.

[18] Bay H, Tuytelaars T, Van Gool L. Surf: speeded up robust features. In: European Conference on Computer Vision 2006; New York, USA; 2006. pp. 404-417.

[19] Morel JM, Yu G. ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences 2009; 2 (2): 438-69.

[20] Tang H, Tang F. AH-SIFT: Augmented histogram based SIFT descriptor. In: 2012 19th IEEE International Conference on Image Processing 2012; New York, USA; 2012. pp. 2357-2360.

[21] Bauer J, Sünderhauf N, Protzel P. Comparing several implementations of two recently published feature detectors. IFAC Proceedings Volumes 2007; 40 (15): 143-148.

[22] Potthast C, Breitenmoser A, Sha F, Sukhatme GS. Active multi-view object recognition and online feature selection. InRobotics Research 2018; 1: 471-488.

[23] Dong J, Liu W, Zhu Y. Application of SIFT feature vector matching algorithm in binocular stereo vision system. In: Fifth Symposium on Novel Optoelectronic Detection Technology and Application; Xi'an, China; 2019. pp. 110231U-110241U.

[24] Alhwarin F, Ristić Durrant D, Gräser A. VF-SIFT: very fast SIFT feature matching. In: Joint Pattern Recognition Symposium; Heidelberg, Germany; 2010. pp. 222-231.

[25] Abdel-Hakim AE, Farag AA. CSIFT: A SIFT descriptor with color invariant characteristics. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) 2006; 2: 1978-1983.

[26] Brown M, Lowe DG. Invariant features from interest point groups. In: BMVC 2002; 4.

[27] Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification. InEuropean Conference on Computer Vision 2006, pp. 490-503.

[28] Daubechies I. The wavelet transform, time-frequency localization and signal analysis. IEEE Transactions on Information Theory 1990; 36 (5): 961-1005.

[29] Subasi A, Yaman E. EMG Signal classification using discrete wavelet transform and rotation forest. In: International Conference on Medical and Biological Engineering 2019.pp. 29-35.

[30] Rhif M, Ben Abbes A, Farah IR, Martínez B, Sang Y. Wavelet transform application for/in non-stationary time-series analysis: A review. Applied Sciences 2019; 9 (7): 1345-1346.

[31] Meng Y, Ye C, Yu S, Qin J, Zhang J et al. Sugarcane node recognition technology based on wavelet analysis. Computers and Electronics in Agriculture 2019; 158: 68-78.

[32] Maity R, Suman M. Predictability of hydrological systems using the wavelet transformation: Application to drought prediction. In: Hydrology in a Changing World 2019, pp. 109-137.

[33] Bharath I, Devendiran S, Mathew AT. Bearing condition monitoring using tunable q-factor wavelet transform, spectral features and classification algorithm. Materials Today: Proceedings 2018; 5 (5): 11476-11490. doi: 10.1016/j.matpr.2018.02.115

[34] Koh JE, Acharya UR, Hagiwara Y, Raghavendra U, Tan JH et al. Diagnosis of retinal health in digital fundus images using continuous wavelet transform (CWT) and entropies. Computers in Biology and Medicine 2017; 84: 89-97. doi: 10.1016/j.compbiomed.2017.03.008

[35] Amiri M, Lina JM, Pizzo F, Gotman J. High frequency oscillations and spikes: separating real HFOs from false oscillations. Clinical Neurophysiology 2016; 127 (1): 187-196. doi: 10.1016/j.clinph.2015.04.290

[36] Starck JL, Murtagh F, Fadili J. Sparse Image and Signal Pprocessing: Wavelets and Related Geometric Multiscale Analysis. USA: Cambridge University Press, 2015.

[37] Bhosale NV, Kanade SS. A comparative analysis on hardware system of 2D discrete wavelet transformation for the image denoising application. In: International Conference on Intelligent Data Communication Technologies and Internet of Things; New York, USA; 2018. pp. 1076-1087

[38] Acharya UR, Raghavendra U, Fujita H, Hagiwara Y, Koh JE et al. Automated characterization of fatty liver disease and cirrhosis using curvelet transform and entropy features extracted from ultrasound images. Computers in Biology and Medicine 2016; 79: 250-258. doi: 10.1016/j.compbiomed.2016.10.022

[39] Kashyap N, Sinha GR. Image watermarking using 3-level discrete wavelet transform (DWT). International Journal of Modern Education and Computer Science 2012; 4 (3): 50-51. doi: 10.5815/ijmecs.2012.03.07

[40] Cortes C, Vapnik V. Support-vector networks. Machine Learning 1995; 20(3): 273-297. doi: 10.1007/BF00994018

[41] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters 1999; 9(3): 293-300. doi: 10.1023/A:1018628609742

[42] Dadi HS, Pillutla GM. Improved face recognition rate using HOG features and SVM classifier. IOSR Journal of Electronics and Communication Engineering 2016; 11(4): 34-44. doi: 10.9790/2834-1104013444

[43] Mun S, Park S, Han DK, Ko H. Generative adversarial network based acoustic scene training set augmentation and selection using SVM hyper-plane. Proceeding of DCASE 2017;1: 93-97.

[44] Fung GM, Mangasarian OL. Multicategory proximal support vector machine classifiers. Machine Learning 2005; 59(1-2): 77-97. doi: 10.1007/s10994-005-0463-6

[45] Bergman S. The Kernel Function and Conformal Mapping. USA: American Mathematical Society Publication, 1970.

[46] Lane F, Azad RM, Ryan C. Principled evolutionary algorithm search operator design and the kernel trick. In: 2016 IEEE Symposium Series on Computational Intelligence (SSCI); Athens, Greece; 2016. pp. 1-9.

[47] Achirul Nanda M, Boro Seminar K, Nandika D, Maddu A. A comparison study of kernel functions in the support vector machine and its application for termite detection. Information 2018; 9(1): 5.

[48] Khalaf M, Hussain AJ, Alafandi O, Al-Jumeily D, Alloghani M et al. An application of using support vector machine based on classification technique for predicting medical data sets. International Conference on Intelligent Computing 2019; 3: 580-591.

[49] Schölkopf B, Smola AJ, Bach F. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. USA: MIT Press, 2002.

[50] Liu Y, Lian J, Bartolacci MR, Zeng QA. Density-based penalty parameter optimization on C-SVM. The Scientific World Journal 2014; 1: 1-20. doi: 10.1155/2014/851814

[51] Kaestner CA. A. Support vector machines and kernel functions for text processing. Revista de Informática Teórica e Aplicada 2013; 20(3): 130-154. doi: 10.22456/2175-2745.39702

[52] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS One 2015; 10 (3): e0118432-.e0118432. doi: 10.1371/journal.pone.0118432

[53] Agarwal S, Roth D. Learning a sparse representation for object detection. In: European Conference on Computer Vision 2002; USA; 2012. pp. 113-127.

[54] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006; 27 (8): 861-874. doi: 10.1016/j.patrec.2005.10.010

[55] Hashemi H, Mehravaran S. Day to day clinically relevant corneal elevation, thickness, and curvature parameter using the orbscan II scanning slit topographer and the pentacam scheimpflug imaging device. Middle East African Journal of Ophthalmology 2010; 1: 44-45. doi: 10.4103/0974-9233.61216

[56] Tschandl, P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data 2018; 5: 180161. doi: 10.1038/sdata.2018.161

[57] Kermany D, Zhang K, Goldbaum M. Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification. Mendeley Data 2018, 2.

[58] Suckling J, Astley S, Betal D, Cerneaz N, Dance DR et al. Mammographic Image Analysis Society MiniMammographic Database, 2005, v1.21 [Dataset].