XL CILAMCE
IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING
NOVEMBER 11-14, 2019
Praiamar Natal Hotel & Convention
Natal, RN-BRAZIL

A COMPARISON OF DIFFERENT CLASSIFICATION STRATEGIES IN MEDICAL IMAGES OF SPECULAR MICROSCOPY TO DETECT GUTTAE IN EARLY STAGES OF FUCHS' DYSTROPHY

Marlon Woelffel Candoti
marlonwoelffel@gmail.com
IFES – Instituto Federal do Espírito Santo
Rodovia ES-010 – Km 6.5 – Manguinhos – CEP 29173-087 – Serra – ES – Brazil
Diego Luchi
Diego.lucchi@gmail.com
UFES – Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514 – Goiabeiras – CEP: 29075-910 – Vitória – ES – Brazil
Flávio Garcia Pereira
flaviogarciap@gmail.com
IFES – Instituto Federal do Espírito Santo
Rodovia ES-010 – Km 6.5 – Manguinhos – CEP 29173-087 – Serra – ES – Brazil
Daniel Cruz Cavalieri
dcruzcavalieri@gmail.com
IFES – Instituto Federal do Espírito Santo
Rodovia ES-010 – Km 6.5 – Manguinhos – CEP 29173-087 – Serra – ES – Brazil

Abstract. Fuchs' endothelial dystrophy, or Fuchs‘ dystrophy, is a slowly progressive corneal disease that usually affects both eyes. Although in many cases early signs of the disease can be seen in people aged 20-30, the disease rarely affects vision until the person reaches the age 50-60. The tests to diagnose the disease are: Biomicroscopy, and Specular Microscopy. In both cases, it is possible to find the morphological changes characteristics of the disease. In this context, this paper compares the use of three distinct machine learning techniques to perform the Fuchs‘ dystrophy diagnosis on Specular Microscopy images in order to reduce the time spent in a manual analysis of the specialist. The approaches used in this work were: Convolutional Neural Networks (CNN); Support Vector Machines (SVM) with features extracted from Histogram of Oriented Gradients (HOG) and by Speeded Up Robust Features (SURF). The dataset consists of 123,200 images of both eyes of different people, obtained over 9 years at Hospital Evangélico de Vila Velha, Espírito Santo, Brazil. Due to the absence of labels in the original dataset, only 2400 images were analyzed and labeled with the help of a specialist. In this subset, only in 1165 exams the Fuchs' dystrophy is present. A cross-validation approach using 10-folds was performed and the results were evaluated through the accuracy, area under the Receiver Operating Characteristic (ROC) curve, precision, recall and F1 score metrics with the CNNs outperforming the other methods.

Keywords: Fuchs‘ dystrophy, Deep Learning, Medical Image, Image Processing

1 Introduction

Fuch's Endothelial Dystrophy, also known as Fuch's Dystrophy, is a disease that was first reported in 1910 by Ernst Fuchs[1]. He was analyzing thirteen elderly patients with cases of bilateral central corneal clouding.

Out of all dystrophies this is one of the most common and it is the most common indication for endothelium corneal transplantation. It affects the lower level of the cornea (Fig. 1), the endothelium, and is usually ranked in stages to help the specialists to know the severity of the disease. In the early stages, where this work is focused, guttae appears in the middle of the cornea and spread towards the periphery followed by a painless decrease in vision. In the final stages, the person starts to feel episodes of pain and visual acuity is severely compromised. But only the appearance of guttae does not necessarily causes the dystrophy. It is known that, because of the natural processing of the cell aging, it is common for elderly patients to have guttae but not the dystrophy[2].

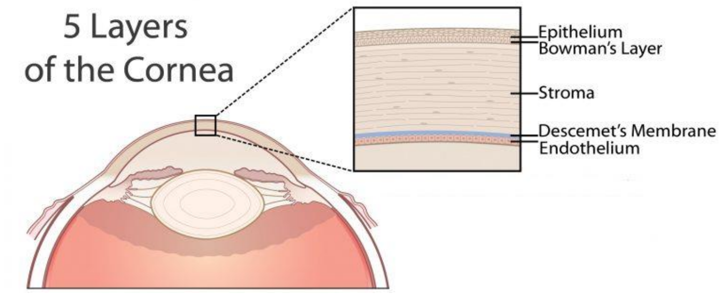
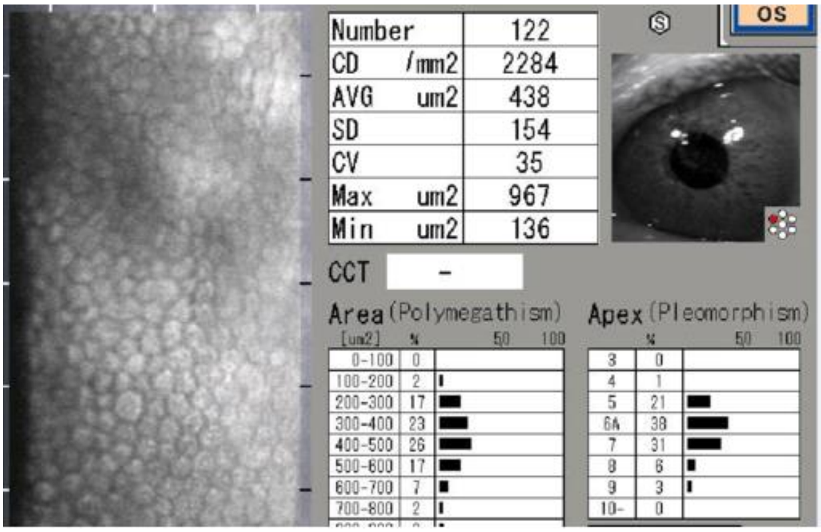


Figure 1. Corneal layers

The most common test to diagnose is the Specular Microscopy, which is a non-invasive instrument that projects a light onto the cornea and capture the endothelium reflection of this light. Then an image is constructed where it is possible to analyze the population, the shape and the size of the cells[3]. Usually 6 images of each eye are taken, 1 from the center and 5 from the side margins of the eye. Figure 2 shows the output of the instrument. This is a type of test that always is done before and after a surgery. Because the endothelium always suffer some damage from the surgery, the specialist needs to know the health of the cornea before and after the surgery.

In this scenario the present work proposes an automatic way of diagnose one of the diseases the specialist looks when it is diagnosing an exam. This way, it could help auxiliating when there is some doubt about the exam, reducing the number of false positives and false negatives, and reducing the time spent on the diagnose. The proposed method in this work is utilizing a Convolutional Neural Network (CNN) to search for images with guttae. Because we lack cases for comparison, as discussed in section 2, the other two methods will serve as a benchmark so we can compare our CNN with something. These other methods involves extracting features from the image, using Histogram of Oriented Gradients (HOG) and Speeded Up Robust Features (SURF), and then classifying using a Support Vector Machine (SVM).



Number	122
CD /mm2	2284
AVG um2	438
SD	154
CV	35
Max um2	967
Min um2	136

CCT -

Area (Polymegathism)

[um2]	%
0-100	0
100-200	2
200-300	17
300-400	23
400-500	26
500-600	17
600-700	7
700-800	2

Apex (Pleomorphism)

	%
3	0
4	1
5	21
6A	38
7	31
8	6
9	3
10-	0

Figure 2. Specular Microscopy view

2 Literature Review

2.1 Previous work

Much is being done of artificial intelligence applied in ophthalmology. Daniel [4] brings a review of different cases where AI is applied to diseases like glaucoma, diabetic retinopathy and others. Even though lots of study has been done, until now the works that are closely related to ours are rather limited. The others involve image segmentation of specular microscopy image and automatic recognition of corneal layers.

A closely related problem was addressed by Sharif [5]. The authors presented a system capable of preprocessing the images and classify which corneal layers the image belongs or if this image presents some abnormality. In this stage, the method works like a binary classification problem and does not point out which abnormality. Out of 7 abnormalities it was studied, one of them is Fuch's Dystrophy. In the binary classification task, the proposed method, based on an ensemble of artificial neural networks and adaptive neuro fuzzy inference system, achieved 100% accuracy. He also described the abnormal and normal images with some statistical metrics such as mean, standard deviation, smoothness, skewness, energy and entropy and showed that with this metrics some differences between them can be viewed. Elbita [6] proposed a neural network to classify the corneal layers using four different statistical texture-based feature extraction techniques applied to original and pre-processed images. The accuracy achieved was 99.4%. Alfredo [7] used neural networks to classify the corneal layers as well. He used Hu variables on binarized images and Zernike moments without the binarization. Siti [8] used deep learning to categorize if the epithelium was injured or not using hyperspectral image.

Anna [9] was the first to use a U-net-based convolutional neural network to segment specular microscopy images of the endothelium cells. She used patches of the image of size 32x32 as input for training. As a result, she obtained DICE of 0.85 and AUC of 0.92. After that, Juan [10] did a comparison between two methods based on convolutional neural networks to segment images of the endothelium cells. It was compared a sliding-windows CNN and a U-net-based CNN. The images were preprocessed with a contrast limited adaptive histogram equalization (CLAHE) to enhance the contrast of the images. He discovered that U-net had the smallest error rate with AUC of 0.9938 compared to 0.9921 from sliding-window approach.

2.2 Convolutional Neural Networks

The CNN was first proposed by Lecun [11] in 1998. In 2012, CNNs became the state of the art on image classification tasks when Krizhevsky [12] drew attention to the famous “AlexNet” model by getting a top-5 error rate of 15.3%, outperforming the second-best entry by 26.2% [13]. Since then, several works have been published in many applications in different fields. In 2014, K. Simonyan & A. Zisserman [14] released the “VGG16”. In 2015, Szegedy [15] developed the “Inception V2”. And many others have been proposed [16].

The basic CNN architecture is usually divided in some parts. In the convolutional layer, a kernel is convolved over the image to extract the most important features. The result is then passed to an activation function, which is normally a non-linear transformation over the data. After the convolutional layer, a pooling layer is responsible for extracting the most dominant features, which are rotational and positional invariant. After that, another convolutional layer is put so more features are extracted. The more are put; the more features are extracted. In the last part there is a dense connected neural network that has the job of classifying those features based on previously given labels. The neurons have complete connection to all the activations from the previous layers. Their activations can hence be computed with a matrix multiplication followed by a bias offset.

2.3 Histogram of Oriented Gradient

The Histogram of Oriented Gradient was proposed in 2005 when Dalal [17] presented the algorithm for a human detection task. Essentially, HOG tries to categorize local object appearance and shape by the distribution of local intensity gradients or edge directions. The image is divided into small connected regions called cells. Each cell is a group of pixels, usually a square group. Within each cell, the gradients for each pixel are calculated. These gradients points to the direction of change in intensity and its magnitude shows how big this change is. Then, with these gradients, a representation (histogram) of this cell is made. Each cell has a fixed number of gradient orientation bins. Each pixel in the cell votes for a gradient orientation bin with a vote proportional to the gradient magnitude at that pixel. The last step is the normalization, which takes local groups of cells (blocks) and normalize them, so that results in better invariance to changes in brightness.

2.4 Speeded Up Robust Features (SURF)

The Speeded Up Robust Features was designed and proposed by H. Bay [18]. It is a method for local, similarity invariant representation and comparison of images. The main interesting fact is that SURF relies on box filters as an approximation of Gaussian filters. This makes the computations really fast enabling real-time operations to be performed. The SURF is divided in two parts: feature detection and feature description.

2.4.1 Feature detection

In the detection, the aim is to find points of interest. To find these points SURF uses Hessian matrix because it has a good performance in computation and accuracy. The Hessian matrix is defined in Eq. (1).

$$H(x,\sigma) = \; [L_{xx}(x,\sigma)\; L_{xx}(x,\sigma)\; L_{xx}(x,\sigma)\; L_{xx}(x,\sigma)\;], \tag{1}$$

where $L_{xx}(x,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial 2}{\partial x2}\; g(\sigma)$ with the image I in point x and is similar for the others. With the determinant of this matrix, the location of the points of interest and the scale are determined. But, to speed up the computation, the Gaussian second order derivative is approximated with box filters. The lowest scale used is a 9 x 9 box that represents a Gaussian with σ = 1.2.

Scale-spaces are usually implemented as pyramids. The images are repeatedly smoothed with Gaussian and subsampled afterwards in order to achieve higher levels of the pyramid. But, using box filters, the image remains the same and only the size of the box filters is altered to perform the computations. Therefore, the scale space is analyzed by upscaling the box filter instead of iteratively reduce the image size. The points of interest are then located at different scales using a non-maximum suppression in a 3 x 3 x3 neighborhood.

2.4.2 Feature Descriptor

For the orientation, SURF uses a Haar-wavelet in x and y directions in a circle of radius 6 times the scale the point of interest was found. Then it is calculated the sum of vertical and horizontal wavelet responses in a scanning area of sizes $\frac{\pi}{3}$. The area with the largest value will be the orientation of the point.

For the descriptor components, a square region of size 20 times the scale is constructed around the keypoint with the same orientation as the circle. The region is then divided into 4×4 square sub-regions. For each sub-region, a horizontal and vertical Haar wavelet is performed at a 5×5 regularly spaced sample points. Then the wavelet responses are summed up over each region. Also, the polarity changes are represented as the sum of the absolute responses. This results in a descriptor vector for all 4×4 subregions of length 64.

2.5 Bag of Visual Words

The idea of Bag of Visual Words (BVW) model [19][20] is that an image can be divided in groups and each group represents something (visual words). It is analogous as the bag of words largely used in natural language processing. After all the descriptors was extracted from the images, it is used an unsupervised approach, so they are grouped together in k mutually exclusive clusters. In this work it is used the K-means algorithm with a fixed size of 100 clusters. Then each cluster center represents a feature, i.e. a visual word. This way, all the centers are our bag of visual words. This step is necessary because we want to group together those descriptors that are the same and almost the same. With these clusters a histogram, where each bin represents the number of times a visual word appeared on the image, is created. Then this histogram is used as input for a supervised approach because it encodes the image as a simple feature vector. This work uses a SVM with a linear kernel.

3 Methodology

3.1 Development Process

This present work proposes a comparison between 3 methods for classifying whether there is or not a guttae presence in a Microscopy Specular image of a patient. By that, we should note how each algorithm was used and about the images.

3.1.1 The images

Originally, the images are RGB of size 266 x 480. Due to hardware limitations, the images were resized by half its size (133 x 240) for the CNN. In the other 2 methods the image was resized of 1, ½ and ⅓ of its original size. Another thing, we transformed the image from RGB to grayscale, to prevent the color to influence in the feature extraction. This transformation also tries to take away the influence of the red lines on the image. These red lines are just the machine trying to segment the cells. But this segmentation is not perfect because there are areas not segmented the way it should and there are areas segmented the way it shouldn't. Besides, these red lines could introduce a lot of unnecessary noise. Figure 3 shows an example of this problem. On the left side, there is a fault on the segmentation. On the right side, it was tried to segment where it should be automatically segmented.

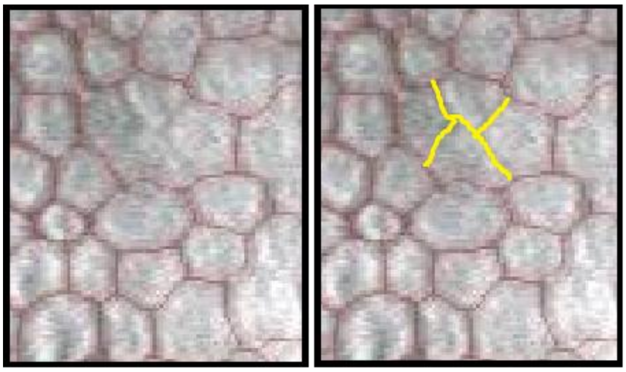


Figure 3. Segmentation problem

3.1.2 The CNN architecture

The CNN makes use of 3 convolutional layers with 2 x 2 kernels, a rectifier linear unit activation (ReLu) function and a max pooling operation after each convolution. After that there is 2 dense layers with 256 and 2 neurons with ReLu and Softmax activation functions respectively. For the 256 layer we use a dropout rate of 0.3, to prevent overfitting. The optimizer used was Adam [21] with learning rate of 0.001 and a decay of 0. To create and train the CNN, the keras library for python was used.

3.1.3 SURF and HOG

The extraction of the SURF features was performed with the OpenCV library, while the HOG was using the implementation available in the sk-image library.

For both, the images were resized so that it was from 1 to 3 times smaller than the original size. After the features are retrieved, BVW are used to cluster these features and then a SVM is used to train and classify the instances. In our tests, SURF was best combined with a SVM with a linear kernel. And HOG was best combined with SVM with linear kernel without BVW.

For HOG, the orientations and the number of cells per block was kept constant at 8 and 4x4 respectively. The pixels per cell used was 16x16 and 8x8. The normalization used was the L2-norm.

3.1.4 Training and evaluation process

A 10-fold cross validation were used for training and evaluation purposes. In this process we divided the dataset in 10 equally sized folds of samples, so we could use 9 for training and 1 for test. The same folds for train and test were used for all three approaches, so that we could not bias our evaluation. The results were evaluated through the accuracy, area under the ROC curve, precision, recall and F1 score metrics.

4 Case Study

4.1 Dataset

The dataset is private and is property of Hospital Evangélico de Vila Velha, Espírito Santo, Brazil. It is consisted of 123000 images of both eyes of different people.

First, it was necessary to crop the images from the specular microscopy test. As it was shown in Fig. 2 the image has a lot of unnecessary information. Because of this, the images were cropped so only the image on the left was used on the training process.

Then, due to the lack of label, we had to label the images. With the help of a specialist, 2665 images were labeled following his instructions. The goal was to separate the dataset in two classes: the images with guttae and the images without. On the images with guttae it was picked up images with huge guttae, because the minor ones could be easily confused with other types of problems. On the images without, it was selected those without any signal of guttae. Figure 4 shows examples of these two classes. On the left there is an endothelium without any signals of guttae. On the right, there is an endothelium with guttae.

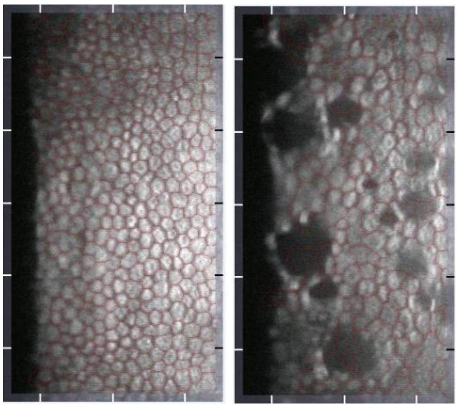


Figure 4. Example of endothelium with and without guttae

4.2 Results

Table 1 shows the mean results of the metrics accuracy, precision, recall, f1 score and the Area Under the ROC Curve (AUC). The accuracy is shown as percentage and the other are in a range from 0 to 1 (better). Below the results, the standard deviation is shown in parentheses to represent the variations around the mean of the folds in the cross validation. After that, the confusion matrix of each are shown below on Table 2, Table 3 and Table 4. The 0 (zero) represents the image without guttae and 1 (one) represents image with. The left side represents the actual values, and the upper side represents the predicted values.

Table 1. Metrics of each algorithm

	Accuracy	Precision	Recall	F1 Score	AUC
CNN (scale ½)	87.68 (2.20)	0.884 (0.034)	0.826 (0.053)	0.854 (0.028)	0.875 (0.023)
HOG (scale 1)	63.53 (3.04)	0.584 (0.035)	0.573 (0.054)	0.570 (0.042)	0.685 (0.032)
HOG (scale ½)	64.99 (1.20)	0.611 (0.019)	0.553 (0.052)	0.578 (0.026)	0.697 (0.017)
HOG (scale ⅓)	65.17 (2.58)	0.634 (0.043)	0.486 (0.036)	0.549 (0.032)	0.689 (0.031)
SURF + BOVW (scale 1)	79.55 (2.15)	0.816 (0.028)	0.686 (0.051)	0.746 (0.032)	0.869 (0.018)
SURF + BOVW (scale ½)	79.21 (2.30)	0.815 (0.039)	0.681 (0.046)	0.741 (0.030)	0.857 (0.020)
SURF + BOVW (scale ⅓)	78.80 (2.37)	0.814 (0.046)	0.670 (0.027)	0.734 (0.027)	0.854 (0.019)

Table 2. HOG + SVM confusion matrix

HOG + SVM (1)	0	1	HOG + SVM (½)	0	1	HOG + SVM (⅓)	0	1
0	1026	474	0	1088	412	0	1170	330
1	497	668	1	521	644	1	598	567

Table 3. SURF + BOW confusion matrix

SURF + BOVW (1)	0	1	SURF + BOVW (½)	0	1	SURF + BOVW (⅓)	0	1
0	1320	180	0	1317	183	0	1319	181
1	365	800	1	371	794	1	384	781

Table 4. CNN confusion matrix

CNN (½)	0	1
0	1374	126
1	202	963

4.3 Result Analysis

The results show that CNN outperformed the other methods in a large range. To corroborate with this, Pawara [22] had shown similar results when comparing local descriptors with famous architectures of CNN.

The local descriptors (SURF) had shown good average performance. But it failed to detect the condition when the condition is positive. And even though we resize the images, the metrics were getting down. This is a little straightforward because the more we resize the image, the more noise appears making it difficult for the SURF to find points of interest that really makes sense.

The HOG features were the worst between these 3 methods. It is shown the results when pixel per cell is 16x16. Its major accuracy was 65.17% with the worst recall of 0.486. During our experiments, we tried the parameter pixel per cell 8x8 as well, but it did not show better results from the above already mentioned.

5 Conclusions

With this work it was possible to observe that local descriptors and neural networks using images with little preprocessing are a good starting point to classify images with or without guttae in the early stages of Fuchys Dystrophy on specular microscopy images.

Therefore, this study concludes the comparison between different approaches that was thought to be useful. Due to the lack of previous work in classifying this type of image, a lot of improvements can be done in future works. We can try to extract other types of features from the images as well as use other techniques to segment the guttae. This work is just the beginning of a series of works that should be done in order to try to categorize the cornea's healthy through images of specular microscopy.

References

[1] ADAMIS, Anthony P. et al. Fuchs' endothelial dystrophy of the cornea. Survey of ophthalmology, v. 38, n. 2, p. 149-168, 1993.

[2] ELHALIS, Hussain; AZIZI, Behrooz; JURKUNAS, Ula V. Fuchs endothelial corneal dystrophy. The ocular surface, v. 8, n. 4, p. 173-184, 2010.

[3] MCCAREY, Bernard E.; EDELHAUSER, Henry F.; LYNN, Michael J. Review of corneal endothelial specular microscopy for FDA clinical trials of refractive procedures, surgical devices and new intraocular drugs and solutions. Cornea, v. 27, n. 1, p. 1, 2008.

[4] HOGARTY, Daniel T.; MACKEY, David A.; HEWITT, Alex W. Current state and future prospects of artificial intelligence in ophthalmology: a review. Clinical & experimental ophthalmology, v. 47, n. 1, p. 128-139, 2019. Clinical and Experimental Ophthalmology 2019; 47: 128–139

[5] SHARIF, Mhd Saeed et al. Medical image classification based on artificial intelligence approaches: a practical study on normal and abnormal confocal corneal images. Applied Soft Computing, v. 36, p. 269-282, 2015.

[6] ELBITA, A. et al. Automatic classification of cell layers in corneal confocal microscopy images. In: Ophthalmic Image Analysis Workshop, University of Liverpool, UK. 2011.

[7] RUGGERI, Alfredo; PAJARO, Simone. Automatic recognition of cell layers in corneal confocal microscopy images. Computer methods and programs in biomedicine, v. 68, n. 1, p. 25-35, 2002.

[8] MD NOOR, Siti et al. Hyperspectral image enhancement and mixture deep-learning classification of corneal epithelium injuries. Sensors, v. 17, n. 11, p. 2644, 2017.

[9] FABIJAŃSKA, Anna. Segmentation of corneal endothelium images using a U-Net-based convolutional neural network. Artificial intelligence in medicine, v. 88, p. 1-13, 2018.

[10] VIGUERAS-GUILLÉN, Juan P. et al. Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation. BMC Biomedical Engineering, v. 1, n. 1, p. 4, 2019.

[11] LECUN, Yann et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, v. 86, n. 11, p. 2278-2324, 1998.
[12] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. 2012. p. 1097-1105.
[13] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. ILSVRC-2012, 2012. URL http://www.image-net.org/challenges/LSVRC/2012/.
[14] SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[15] SZEGEDY, Christian et al. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 2818-2826.
[16] ALOM, Md Zahangir et al. The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164, 2018.
[17] DALAL, Navneet; TRIGGS, Bill. Histograms of oriented gradients for human detection. In: international Conference on computer vision & Pattern Recognition (CVPR'05). IEEE Computer Society, 2005. p. 886--893.
[18] BAY, Herbert; TUYTELAARS, Tinne; VAN GOOL, Luc. Surf: Speeded up robust features. In: European conference on computer vision. Springer, Berlin, Heidelberg, 2006. p. 404-417.
[19] TSAI, Chih-Fong. Bag-of-words representation in image annotation: A review. ISRN Artificial Intelligence, v. 2012, 2012.
[20] CSURKA, Gabriella et al. Visual categorization with bags of keypoints. In: Workshop on statistical learning in computer vision, ECCV. 2004. p. 1-2.
[21] KINGMA, Diederik P.; BA, Jimmy. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
[22] PAWARA, Pornntiwa et al. Comparing Local Descriptors and Bags of Visual Words to Deep Convolutional Neural Networks for Plant Recognition. In: ICPRAM. 2017. p. 479-486.