Drosophila melanogaster Gender Classification Based on Fractal Dimension

Francisco Gerardo Medeiros Neto,
Ítalo Rodrigues Braga
Federal University of Ceará,
Sobral, Ceará, Brazil
E-mails: fcogmneto@gmail.com,
italorodriguesb@gmail.com

Matthew Henry Harber
GeoPoll, Denver, CO, USA
E-mail: matthew.h.harber@gmail.com

Iális Cavalcante de Paula Júnior
Federal University of Ceará,
Sobral, Ceará, Brazil
E-mail: ialis@ufc.br

***Abstract*—Biometrics, previously used only in human identification, can help experts in the analysis of biological images. Flies of the genus *Drosophila* have become model organisms by almost global presence and short life cycle. Facial recognition techniques and geometric morphometry can be used in image processing for classification. The latter requires human interaction. This work details a methodology based on stationary wavelet transform, Canny filter and fractal dimension aimed to infer the gender of *Drosophila melanogaster* based on images of their wings. The combination of variation in the training and test samples and classification methods showed the proposed algorithm's accuracy rate, 90%, outperformed other methods. The proposed methodology proved efficient by using a reduced number of attributes and did not require human interaction for feature extraction (landmarks).**

Keywords - stationary wavelet transform; Canny filter; fractal dimension; classification;

I. Introduction

Biometrics is the identification of human beings from physical and/or behavioral traits [1], [2]. The use of biometric tools can extract phenotypic information from biological images [3]. Digitally extracted image features can discriminate data into groups, such as gender or genotype [3]. In addition to humans, this technique can be used to identify individuals from other animal species, such as flies of the species *Drosophila melanogaster* (fruit fly) [1]–[3]. These flies have a short life cycle and a global presence, which made this species, for over 100 years, a model genetic organism [1]. Studies on flies can help biologists and entomologists analyze the development of different insects and how they may impact nature [2]. The similarity between male and female wings makes it difficult for experts to classify [1]. Proteins and genes similar to the wings of *Drosophila* are found in humans [2] and the application of facial recognition methods in the *Drosophila* wing obtained satisfactory results [1]–[3]. Another technique is geometric morphometric analysis, which requires a prior knowledge of the biological system and landmarks in the image [3].

This paper proposes an automatic gender classification method based on the fractal dimension computed from the segmentation of *Wavelet* components. In Section II, the literature is reviewed. Section III details the proposed methodology: image database, features extracted and classifiers. Section IV presents the results, their discussion and compared against other methods. Section V shows the conclusions and future works.

II. Related works

This section presents works from the literature for image processing of fly wings, wavelet transform, Canny filter and fractal dimension. Studies on computer-aided classification of fly wings images are recent. Payne et al [1] applied the Genetic and Evolutionary Feature Extraction – Machine Learning ($GEFE_{ML}$) [4], created for facial recognition and based on a Local Binary Pattern (LBP), and obtained a 73.16% accuracy. Ahmad et al. [2] compared LBP and Modified Local Binary Pattern (MLBP) reaching 90% and 89.5% accuracy, respectively.

Sonnenschein et al. [3] tested the Bioimage Classification and Annotation Tool (BioCAT) [5] and WINGMACHINE [6] softwares. BioCAT achieved a success rate above 80% for gender but low rates (maximum 52%) for genotype, while WINGMACHINE achieved over 90% for gender and over 80% for genotype, due to landmarks and semi-landmarks (which requires human interaction). Other software for fly wing analysis are FijiWings [7] and MorphoJ [8], both developed in Java and open source.

Wavelets are used for signal analysis and can be used in many areas. Figueiredo et al. [9] and Khalid et al. [10] identified lesions on retinal fundus and skin images, respectively. Bankhead et al. [11] segmented retinal vessels and Demirhan and Güller [12] segmented brain magnetic resonance imaging. Nguyen, Kam and Cheng [13] identified cracks in concrete. Examples of wavelet families are the cubic B-spline [9], [11], [13], Daubechie [12] and Morlet [14].

Another feature used in pattern recognition is fractal dimension. Bruno et al. [15] have recognized Brazilian vegetation with box-counting and Minkowski multiscale methods. Polychronaki et al. [16] analyzed the onset of epileptic seizures on electroencephalogram signals with three signal fractal dimension estimation techniques.

These two concepts can also be combined. Acharya et al. [14] compared the fractal dimension and scales of wavelets of electrocardiograms signals to analyze heart diseases. Bayraktar, Poor and Sircar [17] calculated the fractal dimension from

the data wavelet of the financial market. Murty, Reddy and Babu [18] combined Canny filter and Hough transform to select the area of the iris before the wavelet decomposition and reached 100% accuracy. For classification, common methods are Random Forest (RF), Support-Vector Machine (SVM) [2], K-Nearest Neighbors (KNN) [18], Linear Discriminant Analysis (LDA) [15] and neural networks [3].

III. Materials and methods

This section describes the proposed approach. First, the image database is presented. Then, a brief overview of the concepts of wavelet transform, Canny filter segmentation and fractal dimension is shown. The next topics are the classification methods and the proposed methodology.

A. Image database

The image database of [3], available in [19], was used in this paper's approach. The images are the right and left wings of male and female *Drosophila melanogaster* from the Samarkand region (SAM) [3]. In addition to the wild specimens, there are samples with genotypic variations: Epidermal growth factor receptor (*Egfr*), mastermind (*mam*), *Star* and thickveins (*tkv*). *Egfr*, *mam* and *Star* are lethal homozygotes while *tkv* has a qualitative defect [3]. Samples of these genotypes are shown in Figure 1. The images are divided into 4 groups, taken with Olympus and Leica microscopes and both with 20X and 40X zoom. For the proposed methodology, we selected the images of the group Olympus 40X with a total of 2269 images of 1360 by 1024 pixels. At the bottom right of the images there is a $1mm$ size marking (see Figure 1) which is removed before the feature extraction (see Subsection III-F). The images used in [2] and [1] are similar to those in the selected group. The physical similarity of genotype variation is shown in the Figure 2.

B. Wavelets

The traditional Wavelet transform uses subsampling, decreasing the signal to $1/2^n$ of the original size, where n is the wavelet level [12]. The stationary wavelet transform (SWT) uses filter upsampling at each iteration before signal convolution [12], [20]. The filter upsampling by a factor of 2, denoted by $[]_{\uparrow 2}$, is calculated using:

$$\begin{aligned} h_{i+1}(k) &= [h]_{\uparrow 2^i} * h_i(k) \\ g_{i+1}(k) &= [g]_{\uparrow 2^i} * g_i(k), \end{aligned} \tag{1}$$

and the convolution:

$$\begin{aligned} s_{i+1}(k) &= h_{i+1}(k) * s_i(k) \\ d_{i+1}(k) &= g_{i+1}(k) * s_i(k), \end{aligned} \tag{2}$$

where $i = 0, 1, ..., I$ [12], [20].This feature maintains the original signal size and makes the transform invariant to translation [12], [20]. The components of the transform contain middle frequency information, which is important for segmentation [12]. It is also known as undecimated wavelet transform (UWT) [21] and a widely used variant is the isotropic undecimated wavelet transform (IUWT) [9], [11], [13]. In this paper's proposed methodology, we selected the SWT of the Haar type because it is conceptually simple and fast [18]. The number of decomposition levels varies with filter and image size used. Two levels were used with approximation, horizontal, vertical and diagonal components, totaling 8 images. An example of this transform is shown in Figure 3.

C. Canny filter

Canny filter [22] uses a Gaussian operator with standard deviation of σ for edge detection and it is applied to wavelet components, as shown in Figure 4. The segmentation algorithm follows the steps [22], [23]:

1) Application of Gaussian filter with σ width;
2) Calculation of image gradients, e. g. with Sobel operators [24], [25];
3) Reduction of non edge candidate pixel segments to 0;
4) Hysteresis thresholding: Points above the upper threshold are considered contour and points inside the threshold and connected to contour points are also labeled.

The level of detail of the segmentation increases as the wavelet transform level increases. As the noise level of the images varies, the σ for each component varies, but it was fixed for both levels of the wavelet transform (see Figures 4 and 5).

D. Fractal Dimension

The fractal dimension measures the level of complexity in a shape or texture [15]. Unlike shapes in traditional geometry with integer dimensions, the fractal dimension is fractional, occupying spaces between two Euclidean dimensions [14], [15]. A common way of calculating this value is with the Haussdorf dimension, defined as [15]:

$$d_f = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)}, \tag{3}$$

where $N(\epsilon)$ is the number of cubes of sides ϵ. An algorithm to find this dimension is box-counting, which is based on the number of boxes $N(\epsilon)$ of size ϵ needed to fill the image and is given by [15], [26]:

$$d \sim -\frac{\log N(\epsilon)}{\log \epsilon}. \tag{4}$$

E. Classifiers

The following classifiers were used on this paper's approach [27]:

- RF: a set of unpruned classification trees, uses boostrapping of training data and random selection of features [2];
- SVM: the separation of the data by hyperplanes in a multidimensional space using a kernel function [2];
- MultiLayer Perceptron (MLP): a type of neural network capable of solving non-linear problems [28], [29];
- LDA: projection of the data in a subspace, reducing the dimensionality and maximizing the distance between classes [28], [30];

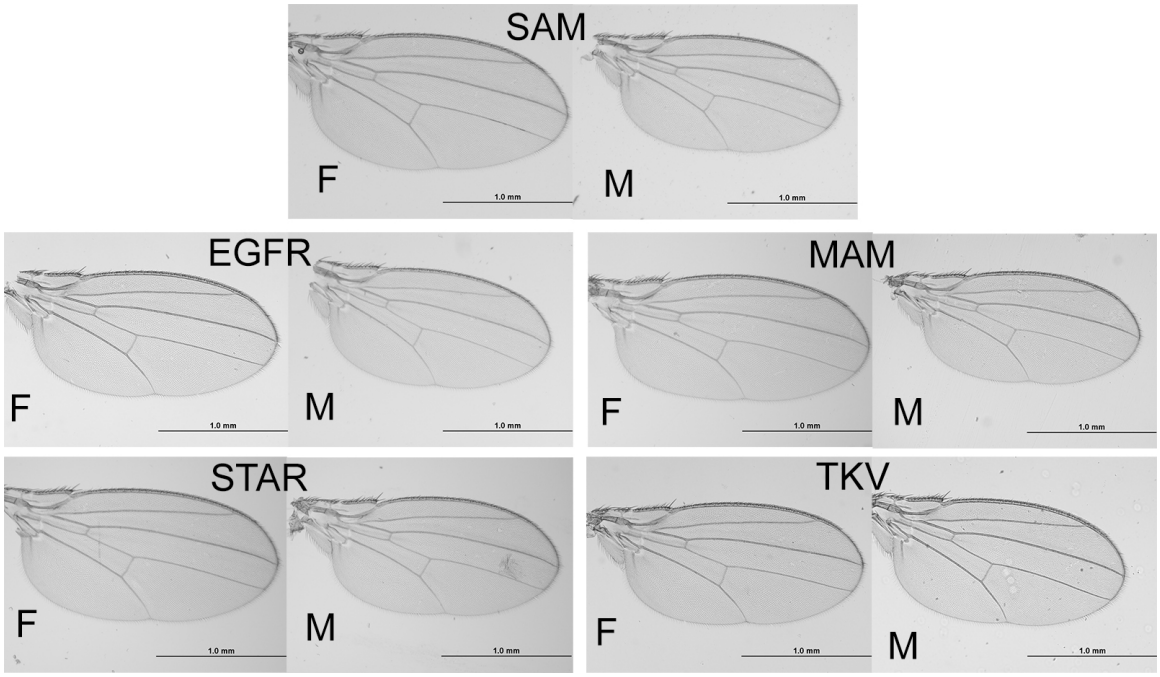

Fig. 1. Samples of *Drosophila melanogaster* wings of both genders and mutant genotypes found in Samarkand, where F and M represent female and male [19].

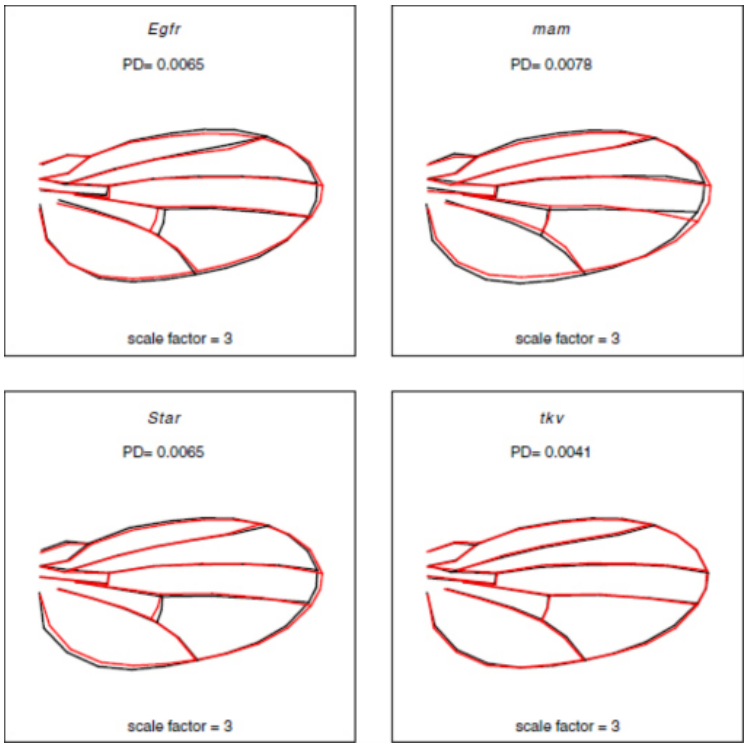

Fig. 2. Procustres distances (PD) magnitudes and mean vector of wild SAM (black) and genotype variations (red), magnified three-fold for better visualization [3]. Source: Sonnenschein et al. [3].

- Quadratic Discriminant Analysis (QDA): similar to LDA but divides data with quadratic surface and does not use the covariance matrix [28], [30];
- KNN: assignment of a class to unknown data from the most frequent class of samples with lower k distances [28].

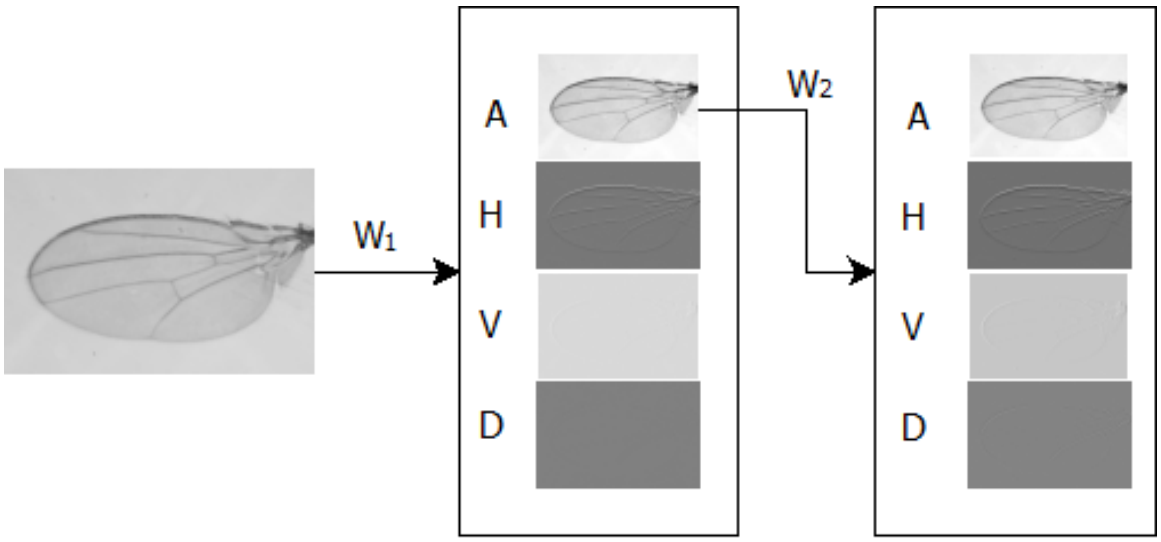

Fig. 3. Application of stationary wavelet transform in a preprocessed image of the database [19] where A represents the approximation component, H, horizontal, V, Vertical and D, diagonal and W_i is the level i of the transform. The components maintains the size of the original image.

F. Proposed methodology

The following steps are used for this paper's proposed methodology:

1) Opening the image file in gray level;
2) Applying of median filter with a structuring element of a disk with a radius 7 for noise removal. Then the image is cut 900 pixels high, with the center of the axes in the upper left corner to remove the $1mm$ marking;
3) Decompositing the filtered image with a stationary Wavelet transform of Haar of 2 levels, generating 8 outputs (horizontal, vertical and diagonal components of each level);
4) Segmentating each image from the previous step from the Canny filter with $\sigma = 21$ to approximation and horizontal components, $\sigma = 4$ for vertical and $\sigma = 3$ for diagonal. These values were found by exhaustive manual search;
5) Calculating the fractal dimension from the edges detected in the previous step;
6) Splitting the data by gender or genotype, shuffling and selecting the samples according to the Table I;
7) Analyzing of the accuracy by cross-validation. Each classifier (see III-E) is run 10 times, varying the training and test data.

Figure 5 shows an example of application of the proposed methodology.

IV. Results and discussion

The proposed methodology reached high success rates for classification by gender, above 90%, even with a reduced number of attributes, as shown in Table I. Eight features are used compared to 50 in [3] or over 700 in [2]. In Table III, we compare the results obtained with works that classified fly wings. This paper's approach requires reduced computational cost and reached high accuracy compared to others studies. By increasing the number of samples per class, the success rate remained consistent. Sonnenschein *et al*. [3] achieved high hit rates when using landmark, which requires user interaction, in opposition to the objective of the proposed approach.

The LDA method was better than other classifiers with SAM samples while QDA was better with all genotypes. The SVM technique obtained better results than RF in both approaches. The KNN method achieved the lowest success rates, followed by MLP. During the tests, the linear kernel classification in SVM was less efficient than the polynomial kernel. The parameters of the classifiers and related to randomness of the data shuffle were kept constant during the tests due to reproducibility.

The used images vary in wing size, orientation and location relative to the background. The approximation and horizontal components of the second level obtained better segmentation, whereas the vertical and diagonal components had a higher noise presence (see Figures 4 and 5). In the first level, the extraction of the contour of the components, more visibly in the horizontal and diagonal, was less efficient. Canny filter σ decrease can improve that result. The Haar wavelet used in the proposed methodology is considered simple [18]. The use of more complex families such as Daubechie or cubic B-Spline can improve the result of the segmentation and classification. The fractal dimension extraction algorithm was written in Cython [31] and the extraction parallelized to optimize performance.

V. Conclusions and future works

The proposed methodology for gender classification was efficient in comparison with studies from the literature. Success rates above 90% with reduced number of attributes and automatic classification are possible. A simple contour detection filter, such as Canny, combined with decomposition

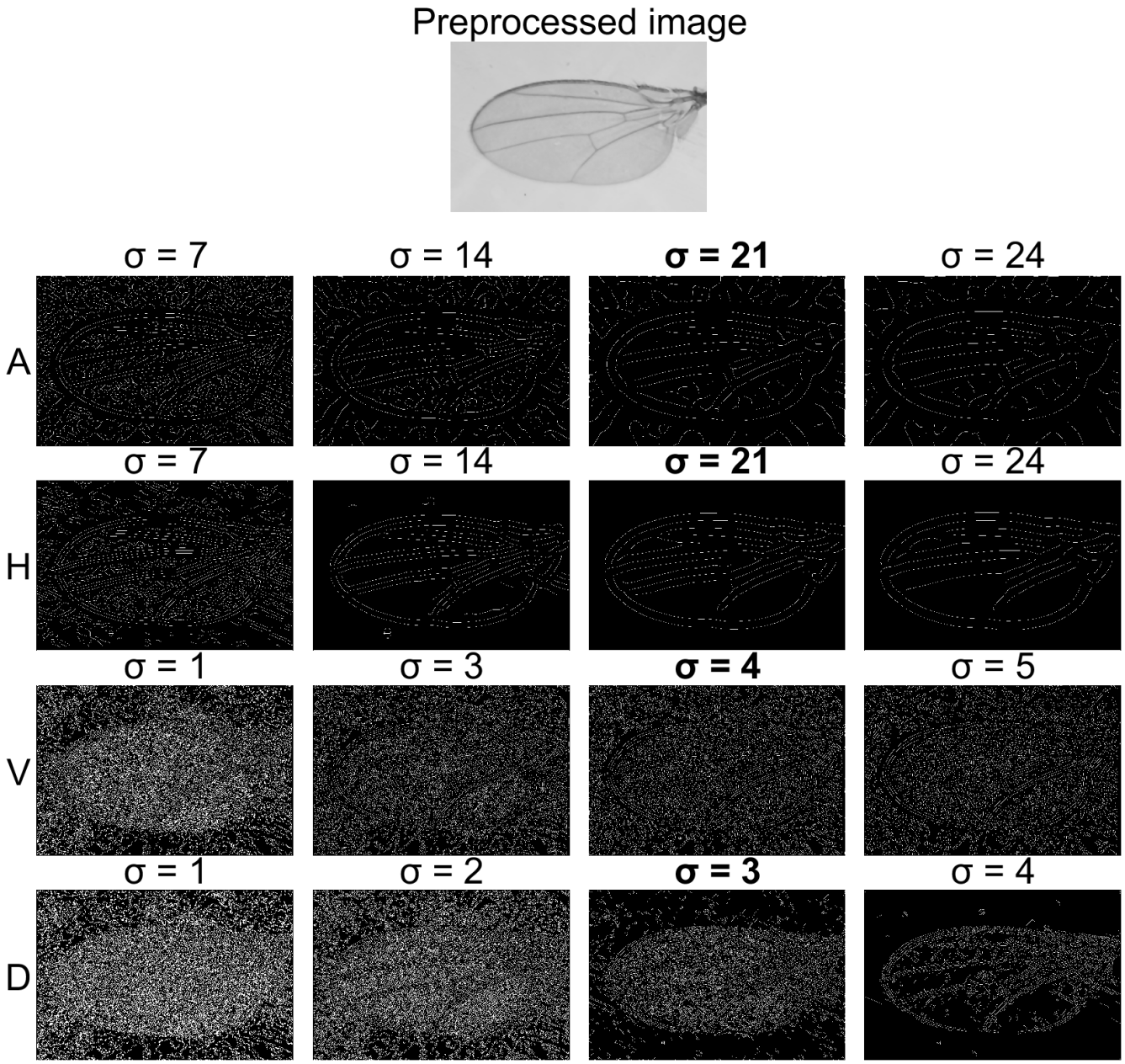

Fig. 4. Segmented level 2 wavelet components (approximation - A, horizontal - H, vertical - V and diagonal - D) of an image from [19] with Canny filter. Highlighted values are used in the proposed methodology.

of a simple wavelet family, such as Haar, achieved satisfactory accuracy. Linear classifiers obtained better results in gender identification.

For future work, we will review other Wavelet families and parameters of the segmentation and classification techniques for a performance improvement. We will also perform the analysis with other images databases for the validation of the proposed methodology.

ACKNOWLEDGMENT

The first author thanks CAPES-Brazil for a master degree scholarship and for the financial support for scientific research. The first author thanks Alceu Costa for the box-counting fractal dimension algorithm available online.

REFERENCES

[1] M. Payne, J. Turner, J. Shelton, J. Adams, J. Carter, H. Williams, C. Hansen, I. Dworkin, and G. Dozier, "Fly wing biometrics," in *2013 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, April 2013, pp. 42–46.

[2] F. Ahmad, K. Roy, B. O'Connor, J. Shelton, G. Dozier, and I. Dworkin, "Fly wing biometrics using modified local binary pattern, svms and random forest," *International Journal of Machine Learning and Computing*, vol. 4, no. 3, pp. 279–285, 2014.

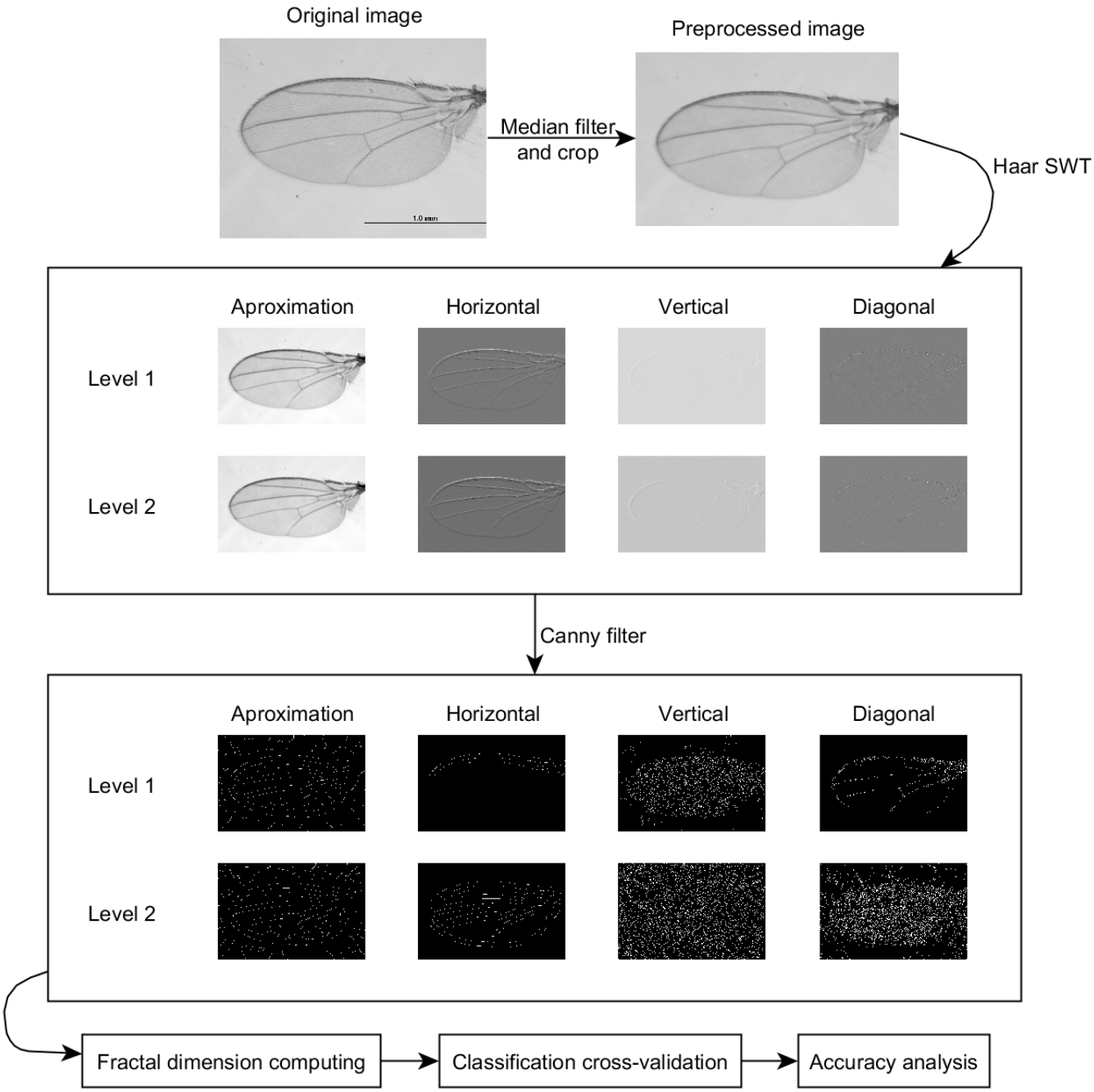

Fig. 5. Proposed methodology on wing from [19].

[3] A. Sonnenschein, D. VanderZee, W. R. Pitchers, S. Chari, and I. Dworkin, "An image database of drosophila melanogaster wings for phenomic and biometric analysis," *GigaScience*, vol. 4, no. 1, p. 25, 2015. [Online]. Available: http://dx.doi.org/10.1186/s13742-015-0065-6

[4] J. Shelton, K. Bryant, S. Abrams, L. Small, J. Adams, D. Leflore, A. Alford, K. Ricanek, and G. Dozier, "Genetic & evolutionary biometric security: Disposable feature extractors for mitigating biometric replay attacks," *Procedia Computer Science*, vol. 8, pp. 351 – 360, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050912000737

[5] J. Zhou, S. Lamichhane, G. Sterne, B. Ye, and H. Peng, "Biocat: a pattern recognition platform for customizable biological image classification and annotation," *BMC Bioinformatics*, vol. 14, 2013. [Online]. Available: http://dx.doi.org/10.1186/1471-2105-14-291

[6] D. Houle, J. Mezey, P. Galpern, and A. Carter, "Automated measurement of drosophila wings," *BMC Evolutionary Biology*, vol. 3, no. 1, p. 25, 2003. [Online]. Available: http://dx.doi.org/10.1186/1471-2148-3-25

[7] A. C. Dobens and L. L. Dobens, "Fijiwings: An open source toolkit for semiautomated morphometric analysis of insect wings," *G3: Genes|Genomes|Genetics*, vol. 3, no. 8, pp. 1443–1449, 2013. [Online]. Available: http://www.g3journal.org/content/3/8/1443.abstract

[8] C. P. Klingenberg, "Morphoj: an integrated software package for geometric morphometrics," *Molecular Ecology Resources*, vol. 11, no. 2, pp. 353–357, 2011. [Online]. Available: http://dx.doi.org/10.1111/j.1755-0998.2010.02924.x

[9] I. Figueiredo, S. Kumar, C. Oliveira, J. Ramos, and B. Engquist,

TABLE I
AVERAGE ACCURACY OBTAINED WITH THE PROPOSED APPROACH AND ALL GENOTYPES. HIGHLIGHTED NUMBERS WERE SELECTED FOR COMPARISON IN TABLE III.

Classifier	Samples per genotype	Test rate (%)	Mean acc. (%)	Maximum acc. (%)	Classifier	Samples per genotype	Test rate (%)	Mean acc. (%)	Maximum acc. (%)
RF	180 (90 per gender)	20	87.28	89.44	*RF*	200 (100 per gender)	50	86.76	89.00
SVM			88.28	90.00	*SVM*			88.14	89.60
MLP			87.11	89.44	*MLP*			87.62	89.00
LDA			87.94	90.56	*LDA*			88.00	89.20
QDA			88.00	91.67	*QDA*			87.38	89.20
KNN			87.06	90.00	*KNN*			86.68	89.00
RF	300 (150 per gender)	50	86.91	89.20	*RF*	400 (200 per gender)	50	**87.86**	88.80
SVM			88.52	89.20	*SVM*			**88.15**	89.20
MLP			87.71	89.33	*MLP*			88.15	89.20
LDA			88.03	89.33	*LDA*			87.76	88.60
QDA			88.60	90.40	*QDA*			88.67	90.30
KNN			86.75	87.87	*KNN*			87.49	88.50

TABLE II
AVERAGE ACCURACY OBTAINED WITH THE PROPOSED APPROACH (SAM SAMPLES ONLY). HIGHLIGHTED NUMBERS WERE SELECTED FOR COMPARISON IN TABLE III.

Classifier	Samples per gender	Test rate (%)	Mean acc. (%)	Maximum acc. (%)	Classifier	Samples per gender	Test rate (%)	Mean acc. (%)	Maximum acc. (%)
RF	90	20	91.39	97.22	*RF*	100	50	91.10	96.00
SVM			94.44	97.22	*SVM*			95.50	99.00
MLP			91.94	97.22	*MLP*			94.00	98.00
LDA			95.28	100.00	*LDA*			94.30	98.00
QDA			93.33	97.22	*QDA*			91.10	94.00
KNN			90.83	97.22	*KNN*			91.20	95.00
RF	150	50	92.33	95.33	*RF*	200	50	**93.45**	96.00
SVM			93.73	94.67	*SVM*			**95.55**	97.00
MLP			92.67	95.33	*MLP*			94.45	96.50
LDA			94.13	96.00	*LDA*			95.80	98.00
QDA			92.07	94.67	*QDA*			94.90	96.00
KNN			91.80	94.67	*KNN*			93.85	97.00

"Automated lesion detectors in retinal fundus images," *Computers in Biology and Medicine*, vol. 66, pp. 47 – 65, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482515002851

[10] S. Khalid, U. Jamil, K. Saleem, M. U. Akram, W. Manzoor, W. Ahmed, and A. Sohail, "Segmentation of skin lesion using cohen–daubechies–feauveau biorthogonal wavelet," *SpringerPlus*, vol. 5, no. 1, p. 1603, 2016. [Online]. Available: http://dx.doi.org/10.1186/s40064-016-3211-4

[11] P. Bankhead, C. N. Scholfield, J. G. McGeown, and T. M. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement," *PLoS ONE*, vol. 7, no. 3, pp. 1–12, 03 2012. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0032435

[12] A. Demirhan and İnan Güler, "Combining stationary wavelet transform and self-organizing maps for brain {MR} image segmentation," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 2, pp. 358 – 367, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0952197610001740

[13] H. N. Nguyen, T. Y. Kam, and P. Y. Cheng, "A novel automatic concrete surface crack identification using isotropic undecimated wavelet transform," in *Intelligent Signal Processing and Communications Systems (ISPACS), 2012 International Symposium on*, Nov 2012, pp. 766–771.

[14] R. A. U., P. S. Bhat, N. Kannathal, A. Rao, and C. M. Lim, "Analysis of cardiac health using fractal dimension and wavelet transformation," *ITBM-RBM*, vol. 26, no. 2, pp. 133 – 139, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1297956205000227

[15] O. M. Bruno, R. de Oliveira Plotze, M. Falvo, and M. de Castro, "Fractal dimension applied to plant identification," *Information Sciences*, vol. 178, no. 12, pp. 2722 – 2733, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025508000364

[16] G. E. Polychronaki, P. Y. Ktonas, S. Gatzonis, A. Siatouni, P. A. Asvestas, H. Tsekou, D. Sakas, and K. S. Nikita, "Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection," *Journal of Neural Engineering*, vol. 7, no. 4, p. 046007, 2010. [Online]. Available: http://stacks.iop.org/1741-2552/7/i=4/a=046007

[17] E. Bayraktar, H. V. Poor, and K. R. Sircar, "Estimating the fractal dimension of the s&p 500 index using wavelet analysis," *International Journal of Theoretical and Applied Finance*, vol. 07, no. 05, pp. 615–643, 2004. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S021902490400258X

[18] P. S. C. Murty, E. S. Reddy, and I. R. Babu, "Iris recognition system using fractal dimensions of haar patterns," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 2, no. 3, pp. 75–84, 2009.

TABLE III
COMPARISON OF THE RESULTS OBTAINED WITH WORKS IN LITERATURE WITH ALL GENOTYPES (†) AND ONLY SAM SAMPLES (*).

Methodology	Features	Classifier	Accuracy (%)
[1]	$GEFE_{ML}$	Manhattan distance	73.16
[2]	MLBP	RF	90.00
	LBP	SVM (linear)	89.50
[3]	Hessians	RF (10 trees)	85.00
		SVM (linear)	81.70
Proposed	Fractal dimension	RF (15 trees)	**87.86**†
		SVM (polynomial)	**88.15**†
		RF (15 trees)	**93.45***
		SVM (polynomial)	**95.55***

[19] A. Sonnenschein, D. VanderZee, W. R. Pitchers, S. Chari, and I. Dworkin, "Supporting material and data for "an image database of drosophila melanogaster wings for phenomic and biometric analysis"," 2015. [Online]. Available: https://doi.org/10.5524/100141

[20] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Transactions on Image Processing*, vol. 4, no. 11, pp. 1549–1560, Nov 1995.

[21] L. Ebadi and H. Z. M. Shafri, "A stable and accurate wavelet-based method for noise reduction from hyperspectral vegetation spectrum," *Earth Science Informatics*, vol. 8, no. 2, pp. 411–425, 2015. [Online]. Available: http://dx.doi.org/10.1007/s12145-014-0168-0

[22] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov 1986.

[23] Şaban Öztürk and B. Akdemir, "Comparison of edge detection algorithms for texture analysis on glass production," *Procedia - Social and Behavioral Sciences*, vol. 195, pp. 2675 – 2682, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042815039567

[24] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, "scikit-image: image processing in Python," *PeerJ*, vol. 2, p. e453, 6 2014. [Online]. Available: http://dx.doi.org/10.7717/peerj.453

[25] "skimage 0.12.2 docs," Documentação, 2017. [Online]. Available: http://scikit-image.org/docs/stable/

[26] R. F. Voss, "Fractals in nature: From characterization to simulation," in *The Science of Fractal Images*, H.-O. Peitgen and D. Saupe, Eds. New York, NY: Springer New York, 1988, pp. 21–70. [Online]. Available: https://doi.org/10.1007/978-1-4612-3784-6_1

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28] "Documentation of scikit-learn 0.18," Documentação, 2017. [Online]. Available: http://scikit-learn.org/stable/documentation.html

[29] S. F. Crone, S. Lessmann, and R. Stahlbock, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing," *European Journal of Operational Research*, vol. 173, no. 3, pp. 781 – 800, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221705006739

[30] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, "Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions," *Current Applied Physics*, vol. 11, no. 3, pp. 740 – 745, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1567173910004153

[31] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn, and K. Smith, "Cython: The best of both worlds," *Computing in Science Engineering*, vol. 13, no. 2, pp. 31–39, March 2011.