

# PROCEEDINGS OF SPIE

[SPIEDigitalLibrary.org/conference-proceedings-of-spie](https://SPIEDigitalLibrary.org/conference-proceedings-of-spie)

## An ensemble-based approach for breast mass classification in mammography images

Patricia B. Ribeiro, João P. Papa, Roseli A. F. Romero

Patricia B. Ribeiro, João P. Papa, Roseli A. F. Romero, "An ensemble-based approach for breast mass classification in mammography images," Proc. SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis, 101342N (21 March 2017); doi: 10.1117/12.2250083

**SPIE.**

Event: SPIE Medical Imaging, 2017, Orlando, Florida, United States

# An Ensemble-based Approach for Breast Mass Classification in Mammography Images

Patricia B. Ribeiro, João P. Papa\*

São Paulo State University, Department of Computing  
Bauru, São Paulo - Brazil  
Email: patriciabellin@yahoo.com.br, papa@fc.unesp.br

Roseli A. F. Romero

University of São Paulo, Department of Computing Science  
São Carlos, São Paulo - Brazil  
Email: rfrance@icmc.usp.br

## ABSTRACT

Mammography analysis is an important tool that helps detecting breast cancer at the very early stages of the disease, thus increasing the quality of life of hundreds of thousands of patients worldwide. In Computer-Aided Detection systems, the identification of mammograms with and without masses (without clinical findings) is highly needed to reduce the false positive rates regarding the automatic selection of regions of interest that may contain some suspicious content. In this work, we introduce a variant of the Optimum-Path Forest (OPF) classifier for breast mass identification, as well as we employed an ensemble-based approach that can enhance the effectiveness of individual classifiers aiming at dealing with the aforementioned purpose. The experimental results also comprise the naïve OPF and a traditional neural network, being the most accurate results obtained through the ensemble of classifiers, with an accuracy nearly to 86%.

## 1. INTRODUCTION

The interpretation of lesions in a mammogram is a complex task that very much depends on the experience of the person in charge of the diagnosis.<sup>1</sup> In order to be more robust and less prone to errors, Computer-Aided Diagnosis-based (CAD) tools have been the subject of extensive research.<sup>2-4</sup> Such systems make use of computer vision and machine learning techniques that can enhance desired properties of an image that may not be able to be observed in naked eyes. Although the literature is rich and a number of works can be referred, there is still room for improvements.

Recently, Ribeiro et al.<sup>5</sup> introduced the Optimum-Path Forest (OPF)<sup>6,7</sup> classifier in the context of breast mass identification. OPF models the problem of pattern classification as a graph partition task, where *prototype* nodes compete among themselves in order to conquer the remaining samples. Each prototype becomes the root of an optimum-path tree, which is a partition of the graph that gather the most strongly connected nodes according to some *path-cost function*. That version of the OPF is parameterless and user-friendly, which means there is no need for fine-tuning parameters, which turns out to be considerably faster than Support Vector Machines for training step.

In this work, we cope with the problem of automatic identification of breast masses in mammography images by means of supervised pattern recognition techniques. Although one can refer to a number of related works in the literature, we present two main contributions to the field: (i) to introduce the Optimum-Path Forest classifier with  $k$ -nearest neighbors graph, hereinafter called  $OPF_{knn}$ ,<sup>8</sup> for breast mass identification, as well as (ii) to present an ensemble-driven approach that boosts the performance of individual classifiers. The ensemble

---

\*The authors would like to thank Capes, FAPESP grant #2014/16250-9, and CNPq grant #306166/2014-3.

is composed of the naïve OPF classifier (i.e. the one that makes use of a complete graph) and a well-known neural network with multi-layer perceptron (MLP).

The remainder of this paper is organized as follows. Sections 2 and 3 present the theoretical background and the methodology adopted in this work, respectively. Section 4 discusses the experiments, and conclusions and final remarks are stated in Section 5.

## 2. OPF FORMULATION

Nowadays, one can find two distinct versions of OPF for supervised learning: (i) the former approach that makes use of a full connectedness graph,<sup>6,7</sup> and (ii) a variant that employs a  $k$ -neighborhood graph.<sup>8-10</sup> In fact, they also differ with respect to the other two main principles, i.e. the heuristic to find out prototypes and path-cost function. The next sections present both techniques in deeper details.

### 2.1 OPF with Complete Graph

Let  $\mathcal{D} = \mathcal{D}^{tr} \cup \mathcal{D}^{ts}$  be a  $\lambda$ -labeled dataset such that  $\mathcal{D}^{tr}$  and  $\mathcal{D}^{ts}$  stand for the training and testing sets, respectively. Additionally, let  $\mathbf{s} \in \mathcal{D}$  be an  $n$ -dimensional sample that encodes features extracted from a certain data, and  $d(\mathbf{s}, \mathbf{v})$  be a function that computes the distance between two samples  $\mathbf{s} \in \mathbf{v}$ ,  $\mathbf{v} \in \mathcal{D}$ .

Let  $\mathcal{G}^{tr} = (\mathcal{D}^{tr}, \mathcal{A})$  be a graph derived from the training set, such that each node  $\mathbf{v} \in \mathcal{D}^{tr}$  is connected to every other node in  $\mathcal{D}^{tr} \setminus \{\mathbf{v}\}$ , i.e.  $\mathcal{A}$  defines an adjacency relation known as *complete graph* (Figure 1a illustrates such training graph), in which the arcs are weighted by function  $d(\cdot, \cdot)$ . We can also define a path  $\pi_s$  as a sequence of adjacent and distinct nodes in  $\mathcal{G}^{tr}$  with terminus at node  $\mathbf{s} \in \mathcal{D}^{tr}$ . Notice a *trivial path* is denoted by  $\langle \mathbf{s} \rangle$ , i.e. a single-node path.

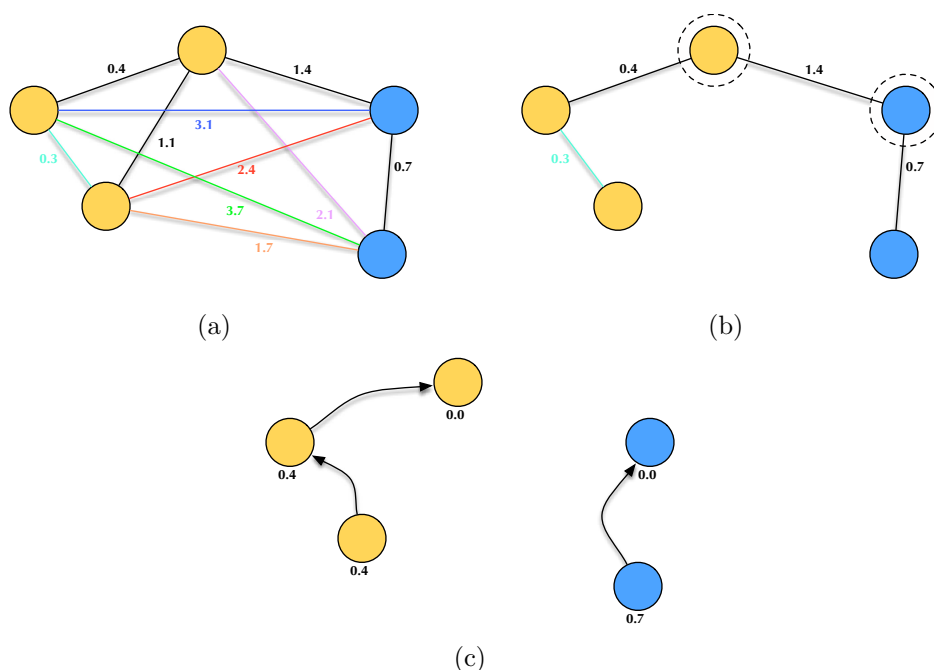


Figure 1. Illustration of the OPF working mechanism: (a) a two-class (orange and blue labels) training graph with weighted arcs, (b) a MST with prototypes highlighted, and (c) optimum-path forest generated during the training phase with costs over the nodes (notice the prototypes have zero cost).

Let  $f(\pi_s)$  be a path-cost function that essentially assigns a real and positive value to a given path  $\pi_s$ , and  $\mathcal{S}$  be a set of prototype nodes. Roughly speaking, OPF aims at solving the following optimization problem:

$$\min f(\pi_s), \forall \mathbf{s} \in \mathcal{D}^{tr}. \quad (1)$$

The good point is that one does not need to deal with mathematical constraints, and the only rule to solve Equation 1 concerns that all paths must be rooted at  $\mathcal{S}$ . Therefore, we must choose two principles now: how to compute  $\mathcal{S}$  (prototype estimation heuristic) and  $f(\pi)$  (path-cost function).

Since prototypes play a major role, Papa et al.<sup>6</sup> proposed to position them at the regions with the highest probabilities of misclassification, i.e. at the boundaries among samples from different classes. In fact, we are looking for the nearest samples from different classes, which can be computed by means of a Minimum Spanning Tree (MST) over  $\mathcal{G}^{tr}$ . The MST has interesting properties, which ensure OPF can be errorless during training when all arc-weights are different to each other.<sup>7</sup> Figure 1b depicts a MST with prototypes highlighted.

Finally, with respect to the path-cost function, OPF requires  $f$  to be a smooth one.<sup>7</sup> Previous experience in image segmentation led the authors to use a chain code-invariant path-cost function, that basically computes the maximum arc-weight along a path, being denoted as  $f_{max}$  and given by:

$$\begin{aligned} f_{max}(\langle s \rangle) &= \begin{cases} 0 & \text{if } \mathbf{s} \in \mathcal{S} \\ +\infty & \text{otherwise,} \end{cases} \\ f_{max}(\pi_s \cdot (\mathbf{s}, \mathbf{t})) &= \max\{f_{max}(\pi_s), d(\mathbf{s}, \mathbf{t})\}, \end{aligned} \quad (2)$$

where  $\pi_s \cdot (\mathbf{s}, \mathbf{t})$  stands for the concatenation between path  $\pi_s$  and arc  $(\mathbf{s}, \mathbf{t}) \in \mathcal{A}$ . In short, by computing Equation 2 for every sample  $\mathbf{s} \in \mathcal{D}^{tr}$ , we obtain a collection of optimum-path trees (OPTs) rooted at  $\mathcal{S}$ , which then originate an optimum-path forest. A sample that belongs to a given OPT means it is more strongly connected to it than to any other in  $\mathcal{G}^{tr}$ . Roughly speaking, the OPF training step aims at solving Equation 2 in order to build the optimum-path forest, as displayed in Figure 1c.

## 2.2 OPF with $k$ -nearest neighbors Graph

Papa and Falcão<sup>8-10</sup> came up with the idea of using a  $k$ -nearest neighbors ( $k$ -nn) graph instead of a full connected one concerning supervised OPF technique. However, in order to deal with such adjacency relation, one should consider changing the two remaining principles, i.e. the path-cost function and heuristic to estimate prototypes. Since we can not guarantee a connected graph with the  $k$ -nn adjacency, we can not longer use a MST to estimate prototypes. The approach firstly presented by Papa and Falcão<sup>8</sup> estimates prototypes in the highest density regions, which is somehow related to elect prototypes nearly to the centroids of the clusters. Therefore, OPF with  $k$ -nn graph can be understood as a dual version of OPF with complete graph (Equation 1), since it aims at maximizing the cost of each sample, as follows:

$$\max f(\pi_s), \forall \mathbf{s} \in \mathcal{D}^{tr}. \quad (3)$$

One main difference regarding OPF with  $k$ -nn graph concerns weighting the nodes as well, since the OPF presented in the previous section does weight the arcs among samples only. Now, we should redefine the training and testing graphs as follows:  $\mathcal{G}^{tr} = (\mathcal{D}^{tr}, \mathcal{A}_k)$ , and  $\mathcal{G}^{ts} = (\mathcal{D}^{ts}, \mathcal{A}_k)$ , where  $\mathcal{A}_k$  stands for a  $k$ -nearest neighbors-based adjacency relation. Additionally, let  $\rho(\mathbf{s})$  be the probability density function (pdf) of a given sample  $\mathbf{s} \in \mathcal{D}^{tr}$  computed as follows:

$$\rho(\mathbf{s}) = \frac{1}{\sqrt{2\pi\sigma^2k}} \sum_{\forall \mathbf{t} \in \mathcal{A}_k(\mathbf{s})} \exp\left(\frac{-d(\mathbf{s}, \mathbf{t})}{2\sigma^2}\right), \quad (4)$$

where  $\mathcal{A}_k(\mathbf{s})$  stands for the  $k$ -nearest neighbors of sample  $\mathbf{s}$ , and  $\sigma = d_{max}/3$ . In this case,  $d_{max}$  denotes the maximum arc-weight in  $\mathcal{G}^{tr}$ . Notice  $\rho(\mathbf{s})$  considers all adjacent nodes for the probability computation purposes, since a Gaussian function covers 99.7% of the samples within  $d(\mathbf{s}, \mathbf{t}) \in [0, 3\sigma]$ . Also, since  $\mathcal{A}_k$  may not be symmetric, we force it to avoid overclustering the feature space prior to its classification.

After computing the pdf for each training node, the competition process among prototypes takes place through the path-cost function  $f_{min}$ , given as follows:

$$\begin{aligned}
f_{min}(\langle t \rangle) &= \begin{cases} \rho(t) & \text{if } t \in \mathcal{S} \\ \rho(t) - 1 & \text{otherwise} \end{cases} \\
f_{min}(\pi_s \cdot (s, t)) &= \min\{f_{min}(\pi_s), \rho(t)\}.
\end{aligned} \tag{5}$$

The rationale behind the above equation can be explained in two steps: (i) the first one is related to the upper formulation, which computes the initial cost of each sample in  $\mathcal{D}^{tr}$  using the pdf, and (ii) the lower equation that propagates the optimum-paths to the samples. Concerning step (i), at the very beginning of the training phase, OPF with  $k$ -nn graph computes the pdf for every sample in  $\mathcal{D}^{tr}$  and stores such values in the priority queue  $Q$ . In order to avoid overclustering (a prototype may be located at a plateau of densities), when a new sample pops out of  $Q$ , it is verified whether this sample has a predecessor or not. In the former case, i.e. we have no predecessor, it means such sample is a prototype (notice in this new OPF we remove samples with the highest costs from  $Q$ ), and thus we keep its original density value; otherwise we subtract one unit from its value.

In regard to step (ii), suppose sample  $s$  is trying to conquer another sample  $t$ . The conquering process concerns  $s$  offering the minimum value between its cost  $f_{min}(\pi_s)$  and the pdf value of  $t$ , i.e.  $\rho(t)$ . Since the prototype holds the higher cost of its optimum-path tree, the idea is to conquer samples with lower costs. Finally, the sample that maximizes  $f_{min}$  for  $t$  will be its conqueror.

### 3. METHODS

In this section, we describe the methodology used to evaluate both OPF<sub>knn</sub> and the ensemble-based approach. Additionally, we present deeper details about the dataset and the statistical evaluation.

#### 3.1 Dataset

A dataset composed of 120 images containing regions of interest of several sizes was used in agreement with medical reports supplied for each mammography image,<sup>11</sup> being 60 images from patients containing suspect masses and 60 without masses. In this work, we are interested in classifying each image in two classes: with (positive) and without breast mass (negative), being the data of the former class composed of images with benign and malign samples. The original digital mammograms were obtained from films digitized by a Lumiscan (Lumisys, Inc.) scanner, with 12 bits of contrast resolution and spatial resolution of 0.15mm and 0.075 mm per pixel.

#### 3.2 Feature Extraction

In the context of breast masses identification, the texture plays an important role,<sup>12–14</sup> thus providing good measures that describe an image based on the variation of intensity or subtle changes between the object and the image's background.

We used the well-known Haralick texture features approach proposed by Haralick,<sup>15</sup> which aim at representing the co-occurrence matrix of grey-levels for each image's pixel, wherein the objective is to compute the probability of the combined occurrence among grey levels in different angles. For each image (specifically the region of interest, i.e. the masses), we extracted the Haralick texture features, being the best ones selected using a Gaussian distribution-based approach.<sup>14,16</sup> In this method, the smaller the overlap of the curves the better the feature represents each image. For each feature, two Gaussian distributions are fitted, being one for the positive images of the dataset and the other one for the negative samples. Thus, these two Gaussian distributions are compared to the amount of overlap between them. Further, the selected features are then used as the input to the supervised methods employed in this work.

### 3.3 Statistical Evaluation

In this work, we performed a multiple hold out procedure over 20 runnings in order to provide data for further statistical analysis, being the dataset normalized within the range  $[0, 1]$ . We used 75% of the dataset for training, and the remaining 25% for classification purposes. Also, the accuracy measure employed in this work was the one proposed by Papa et al.,<sup>6</sup> which considers unbalanced datasets.

Let  $N_i$  be the number of samples from class  $i$  in a given dataset  $\mathcal{D}$ ,  $i = 1, 2, \dots, C$ , where  $C$  stands for the number of classes. Also, we can define the partial errors  $e_{i,1}$  and  $e_{i,2}$  as follows:

$$e_{i,1} = \frac{FP_i}{|\mathcal{D}| - N_i} \quad \text{and} \quad e_{i,2} = \frac{FN_i}{N_i}, \quad (6)$$

where  $FP_i$  and  $FN_i$  stand for the false positives and false negatives of class  $i$ , respectively. Roughly speaking, partial  $e_{i,1}$  computes the percentage of samples from class  $j$  that were misclassified as belonging to class  $i$ ,  $i \neq j$ , and partial  $e_{i,2}$  computes the percentage of misclassified samples from class  $i$ . Finally, we can merge the partials as follows:

$$E_i = e_{i,1} + e_{i,2}. \quad (7)$$

The final accuracy  $Acc \in [0, 1]$  is given by equation below:

$$Acc = 1 - \frac{\sum_{i=1}^C E_i}{2C}. \quad (8)$$

Additionally, the results have been analyzed by means of the Wilcoxon signed-rank test<sup>17</sup> with significance level of 0.05. Notice the percentages used to partition the dataset were empirically chosen.

## 4. RESULTS

In this section, we present the experiments conducted to assess the effectiveness of both OPF<sub>knn</sub> and the ensemble for breast masses identification. As aforementioned, we have selected the best set of features using the method described in Section 3.2. The following set of features have been chosen as the best representative: Contrast (estimation of intensity variation between pairs of image points), Sum Average (average background tones), Sum Variance (variation in image's background tones), Sum Entropy (it indicates the amount of information among pairs of pixels), Difference entropy (is an indicator of the amount of clutter among the background pixels), and Information of Measure Correlation (it is a correlation indicator based on the measure of entropy for independent elements and pairs of elements in the image).

In regard to the pattern recognition techniques, we used a MLP neural network, OPF and OPF<sub>knn</sub>. All these techniques were used to compose the ensemble either. With respect to the MLP network, we used an architecture composed of one hidden layer with 8 neurons, and two neurons in the output layer, as well as a learning rate of 0.7. Since OPF<sub>knn</sub> has the size of the  $k$ -neighborhood as a parameter, we used the learning procedure proposed by Papa and Falcão<sup>9</sup> to find out reasonable values for  $k \in [1, 20]$ . Roughly speaking, this algorithm aims at finding the value of  $k$  that maximizes the classification accuracy over a validating set. With respect to the MLP implementation, we used the Fast Artificial Neural Network (FANN) library, and concerning OPF and OPF<sub>knn</sub> we used LibOPF.<sup>18</sup>

Finally, with respect to the ensemble of classifiers, we used a weighted majority voting approach, as follows: since MLP obtained the best individual results, we decided to weight its decision as 0.5; followed by OPF<sub>knn</sub> with weight of 0.3, and OPF with weight of 0.2. Although these values were empirically chosen, the idea is to give large weights to the best classifiers.

Table 4 presents the mean accuracy results, being the most accurate results according to the Wilcoxon test in bold. Clearly, the best results were obtained by means of the ensemble of classifiers, being around 6.9% more accurate than the second best approaches (MLP and OPF<sub>knn</sub> were considered statistically similar to each other). Another interesting observation concerns the OPF<sub>knn</sub> classifier, which obtained better results than naïve OPF. However, the latter approach has been the fastest one for training purposes, closely followed by OPF<sub>knn</sub>, as presented in Table 4. OPF has been around 3,114.97 times faster than MLP for training purposes, and around 48.29 times faster than OPF<sub>knn</sub> for the very same learning process.

Hold out #	MLP	OPF	OPF <sub>knn</sub>	Ensemble
1	86.67	76.67	86.67	93.33
2	83.33	80.00	83.33	90.00
3	86.67	73.33	56.67	86.67
4	56.67	66.67	63.33	90.00
5	90.00	96.67	76.67	96.67
6	76.67	63.33	63.33	83.33
7	76.67	53.33	86.67	93.33
8	93.33	80.00	80.00	93.33
9	93.33	86.67	80.00	96.67
10	86.67	93.33	86.67	90.00
11	73.33	63.33	83.33	73.33
12	90.00	80.00	83.33	90.00
13	80.00	80.00	86.67	90.00
14	86.67	80.00	86.67	93.33
15	76.67	83.33	70.00	90.00
16	63.33	80.00	86.67	83.33
17	86.67	70.00	83.33	26.00
18	86.67	80.00	80.00	90.00
19	73.33	83.33	80.00	90.00
20	80.00	83.33	83.33	86.67
Average	81.33±9.57	77.67±10.32	79.33±8.95	<b>86.30±15.12</b>

Table 1. Mean accuracy concerning the breast mass classification.

Hold out #	MLP	OPF	OPF <sub>knn</sub>
1	0.523387	0.000458	0.008893
2	0.655113	0.000415	0.02033
3	0.059134	0.000435	0.02212
4	0.32443	0.000425	0.021417
5	5.081678	0.000442	0.023092
6	4.955903	0.000394	0.022833
7	0.786171	0.000479	0.021188
8	0.160995	0.000373	0.02197
9	0.122259	0.000463	0.022711
10	0.261161	0.000418	0.022885
11	0.836253	0.000346	0.019774
12	4.90849	0.000389	0.020872
13	0.677381	0.000437	0.021165
14	4.909609	0.000327	0.023917
15	0.099265	0.000573	0.01942
16	1.504622	0.000437	0.01983
17	0.130734	0.000427	0.021054
18	0.10642	0.000421	0.019617
19	0.472649	0.000429	0.019544
20	0.06359	0.000464	0.020416
Average	1.3319622	<b>0.0004276</b>	0.0206524

Table 2. Mean computational load (seconds) concerning the training time.

## 5. CONCLUSIONS

This work evaluated the  $OPF_{knn}$  in the context of breast masses identification in mammography images, as well as we presented an ensemble-based approach that can enhance the effectiveness of individual classifiers. Although naïve OPF is extremely faster for training, it has obtained the worst results, being  $OPF_{knn}$  the one with the best trade-off between effectiveness and efficiency. In addition, the proposed ensemble of classifiers obtained the best results so far, thus showing considerably accurate recognition rates.

## REFERENCES

- [1] Kopans, D., [*Breast Imaging*], Lippincott-Raven (1998).
- [2] Ribeiro, P. B., Romero, R. A., Oliveira, P. R., Schiabel, H., and Verçosa, L. B., “Automatic segmentation of breast masses using enhanced ICA mixture model,” *Neurocomputing* (2013).
- [3] Soares, F., Janela, F., Pereira, M., Seabra, J., and Freire, M., “Classification of breast masses on contrast-enhanced magnetic resonance images through log detrended fluctuation cumulant-based multifractal analysis,” (2013).
- [4] Choi, J. Y., Kim, D. H., Plataniotis, K. N., and Ro, Y. M., “Computer-aided detection (cad) of breast masses in mammography: combined detection and ensemble classification,” *Physics in Medicine and Biology* **59**(14), 3697 (2014).
- [5] Ribeiro, P. B., Costa, K. A. P., Papa, J. P., and Romero, R. A. F., “Optimum-path forest applied for breast masses classification,” in [*IEEE 27th International Symposium on Computer-Based Medical Systems*], 52–55 (2014).
- [6] Papa, J., Falcão, A. X., and SUZUKI, C. T. M., “Supervised pattern classification based on optimum-path forest,” *International Journal of Imaging Systems and Technology*, 120–131 (2009).
- [7] Papa, J. P., Falcão, A. X., Albuquerque, V. H. C., and Tavares, J. M. R. S., “Efficient supervised optimum-path forest classification for large datasets,” *Pattern Recognition* **45**(1), 512–520 (2012).
- [8] Papa, J. P. and Falcão, A. X., “A new variant of the optimum-path forest classifier,” in [*Advances in Visual Computing*], Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Porikli, F., Peters, J., Klosowski, J., Arns, L., Chun, Y., Rhyne, T.-M., and Monroe, L., eds., *Lecture Notes in Computer Science* **5358**, 935–944, Springer Berlin Heidelberg (2008).
- [9] Papa, J. P. and Falcão, A. X., “A learning algorithm for the optimum-path forest classifier,” in [*Graph-Based Representations in Pattern Recognition*], Torsello, A., Escolano, F., and Brun, L., eds., *Lecture Notes in Computer Science* **5534**, 195–204, Springer Berlin Heidelberg (2009).
- [10] Papa, J. P., Fernandes, S. E. N., and Falcão, A. X., “Optimum-path forest based on k-connectivity: Theory and applications,” *Pattern Recognition Letters* (2016).
- [11] “Online mammographic images database for development and comparison of cad schemes,” *Journal of Digital Imaging* **24**(3) (2011).
- [12] Mudigonda, N. R., Rangayyan, R. M., and Desautels, J. E. L., “Gradient and texture analysis for the classification of mammographic masses,” *IEEE Transactions on Medical Imaging* **19**(10), 1032–1043 (2000).
- [13] S.E. Grigorescu, N. P. and Kruizinga, P., “Comparison of texture features based on gabor filters,” *IEEE Transactions on Image Processing* **11**(10), 1160–1167 (2002).
- [14] Ribeiro, P. B., Costa, K. A. P., Papa, J. P., and Romero, R. A. F., “Optimum-path forest applied for breast masses classification,” *IEEE 27th International Symposium on Computer Based Medical Systems (CBMS)* **27**, 52–55 (2014).
- [15] Haralick, R., Shanmugam, K., and Dinstein, I., “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3**, 610–621 (1973).
- [16] Patrocínio, A., Schiabel, H., Benatti, R., Goes, C., and Nunes, F., “Investigation of clustered microclassification features for an automated classifier as part of a mammography cad scheme,” in [*Proceedings of the 22nd Annual International Conference of the IEEE in Engineering in Medicine and Biology Society*], (2000).
- [17] Harris, T. and Hardin, J. W., “Exact wilcoxon signed-rank and wilcoxon mann-whitney ranksum tests,” *Stata Journal* **13**(2), 337–343 (2013).



- [18] Papa, J. P., S., C. T. N., and Falcão, A. X., *LibOPF: A library for the design of optimum-path forest classifiers* (2009). Software version 2.0 available at <http://www.ic.unicamp.br/~afalcao/LibOPF>.