Optimum-Path Forest: A Novel and Powerful Framework for Supervised Graph-based Pattern Recognition Techniques

João Paulo Papa and Alexandre Xavier Falcão

Institute of Computing – University of Campinas
P.O. Box 6176 – 13083-970 – Campinas – SP – Brazil

papa.joaopaulo@gmail.com, afalcao@ic.unicamp.br

Abstract. *We present here a novel framework for graph-based pattern recognition techniques called Optimum-Path Forest (OPF), which has been demonstrated to be superior than traditional supervised pattern recognition techniques, such as Artificial Neural Networks using Multilayer Perceptrons and Support Vector Machines, in terms of both accuracy and execution times. The OPF-based classifiers model the problem of the pattern recognition as a computation of an optimum-path forest in a graph induced by the dataset samples, achieving very good results in complex situations, i.e., in which we have a large amount of overlapped regions. Results in several real and synthetic datasets show the robustness of the OPF-based classifiers against the above ones.*

1. Introduction

Patterns are usually represented by feature vectors obtained from samples of a dataset [Duda et al. 2000]. Two fundamental problems in pattern recognition are: (i) unsupervised classification, in which we have to identify natural groups (clustering) composed by samples with similar patterns and (ii) supervised one, in which we have to classify each sample in one of c possible classes (labels). Our focus is on the supervised learning approaches, in which we have, basically, three typical cases in 2D feature spaces using two classes: (a) linearly separable, (b) piecewise linearly separable, and (c) non-separable classes with arbitrary shapes. Any reasonable approach should handle (a) and (b), being (c) the most interesting challenge. An artificial neural network with multi-layer perceptrons (ANN-MLP), for example, can address (a) and (b), but not (c) [Haykin 1994]. As an unstable classifier, collections of ANN-MLP can improve its performance up to some unknown limit of classifiers. Support vector machines (SVMs) have been proposed to overcome the problem, by assuming linearly separable classes in a higher-dimensional feature space [Boser et al. 1992]. Its computational cost rapidly increases with the training set size and the number of support vectors. As a binary classifier, multiple SVMs are required to solve a multi-class problem. Tang and Mazzoni [Tang and Mazzoni 2006] proposed a method to reduce the number of support vectors in the multi-class problem. Their approach suffers from slow convergence and high computational cost, because they first minimize the number of support vectors in several binary SVMs, and then share these vectors among the machines. Another point is that, in all SVM approaches, the assumption of separability may also not be valid in any space of finite dimension.

We have proposed a framework for supervised classifiers based on *optimum-path forest* (OPF) [Papa et al. 2007, Papa et al. 2008a, Papa and Falcão 2008, Papa et al. 2009d, Papa and Falcão 2009a], which allows the design of fast, simple

and native multi-class classifiers, which can handle some degree of overlapping between classes. The training set is thought of as a graph, whose nodes are the samples and arcs are defined by some adjacency relation. The arcs are weighted by the distances between the feature vectors of their corresponding nodes and the nodes can also be weighted by some probability density function. Any sequence of distinct samples forms a path connecting the terminal nodes and a *connectivity function* assigns a cost to that path (e.g., the maximum arc-weight along it). The idea is to identify prototypes in each class such that every sample is assigned to the class of its most strongly connected prototype. That is, the one which offers to it an optimum-cost path, considering all possible paths from the prototypes. Our framework can handle all three aforementioned cases (linearly separable, piecewise linearly separable, and non-separable classes) with a predefined arc-weight function and prototypes estimation methodology.

Other contribution of this work concerns learning algorithms, which can teach a classifier from its errors on a third evaluation set without increasing the size of the training set. As the samples in the test set can not be seen during the project, the evaluation set is necessary for this purpose. The basic idea is to randomly interchange samples of the training set with misclassified samples of the evaluation set, retrain the classifier and evaluate it again, repeating this procedure during a few iterations. The effectiveness is measured by comparing the results on the unseen test set before and after the learning algorithm. It is expected an improvement in performance for any stable classifier. This paper describes the supervised OPF-based classifiers and their learning algorithms in Section 2, shows results that compare the OPF-based classifiers with SVM and ANN-MLP in Section 3, and states conclusions in Section 4.

2. Optimum-Path Forest Classifiers

Let Z_1, Z_2, and Z_3 be training, evaluation, and test sets with $|Z_1|$, $|Z_2|$, and $|Z_3|$ samples of a given dataset. We use samples as points, images, voxels, and contours in this paper. This division of the dataset is necessary to validate the classifier and evaluate its learning capacity from the errors. Z_1 is used to project the classifier and Z_3 is used to measure its accuracy, being the labels of Z_3 kept unseen during the project. A pseudo-test on Z_2 is used to teach the classifier by randomly interchanging samples of Z_1 with misclassified samples of Z_2. After learning, it is expected an improvement in accuracy on Z_3.

Our problem consists of using S, (v, d) and Z_1 to project an optimal classifier which can predict the correct label $\lambda(s)$ of any sample $s \in Z_3$. The OPF-based classifiers create a discrete optimal partition of the feature space such that any sample $s \in Z_3$ can be classified according to this partition. This partition is an optimum path forest (OPF) computed in $\Re^n$ by the image foresting transform (IFT) algorithm [Falcão et al. 2004]. Let (Z_1, A) be a graph whose nodes are the training samples and any pair of samples defines an arc in $A = Z_1 \times Z_1$. A path is a sequence of distinct samples $\pi_t = \langle s_1, s_2, \ldots, t\rangle$ with terminus at a sample t, and is said *trivial* if $\pi_t = \langle t\rangle$. We assign to each path π_t a cost $f(\pi_t)$ given by a connectivity function f. A path π_t is said optimum if $f(\pi_t) \leq f(\tau_t)$ for any other path τ_t. We also denote by $\pi_s \cdot \langle s, t\rangle$ the concatenation of a path π_s and an arc (s, t). We have presented two OPF-based classifiers: OPF using complete graph (OPF_{cpl}) and OPF using k-nn graph (OPF_{knn}), in which they differ in the adjacency relation, path-cost function and prototypes estimation methodology. The OPF framework allows us to design different classifiers changing one/all the above aspects.

2.1. Optimum-Path Forest classifier using complete graph

The OPF$_{cpl}$ uses a complete graph as adjacency relation, i. e., each pair of node is connected by an arc (full connectedness graph). The arcs do not need to be stored and so the graph does not need to be explicitly represented. The path-cost function used is f_{max}, given by

$$\begin{aligned} f_{\max}(\langle s \rangle) &= \begin{cases} 0 & \text{if } s \in S, \\ +\infty & \text{otherwise} \end{cases} \\ f_{\max}(\pi_s \cdot \langle s,t \rangle) &= \max\{f_{\max}(\pi_s), d(s,t)\} \end{aligned} \quad (1)$$

such that $f_{\max}(\pi_s \cdot \langle s,t \rangle)$ computes the maximum distance between adjacent samples along the path $\pi_s \cdot \langle s,t \rangle$. The minimization of $f_{\max}$ assigns to every sample $t \in Z_1$ an optimum path $P^*(t)$ from the set $S \subset Z_1$ of prototypes, whose minimum cost $C(t)$ is

$$C(t) = \min_{\forall \pi_t \in (Z_1, A)} \{f_{\max}(\pi_t)\}. \quad (2)$$

The minimization of $f_{\max}$ is computed by the OPF$_{cpl}$ algorithm, which is an extension of the general image foresting transform (IFT) algorithm [Falcão et al. 2004] from the image domain to the feature space, here specialized for $f_{\max}$. This process assigns one optimum path from S to each training sample t in a non-decreasing order of minimum cost, such that the graph is partitioned into an optimum-path forest P (a function with no cycles which assigns to each $t \in Z_1 \backslash S$ its predecessor $P(t)$ in $P^*(t)$ or a marker nil when $t \in S$. The root $R(t) \in S$ of $P^*(t)$ can be obtained from $P(t)$ by following the predecessors backwards along the path, but its label is propagated during the algorithm by setting $L(t) \leftarrow \lambda(R(t))$. We say that S^* is an optimum set of prototypes when the OPF$_{cpl}$ algorithm minimizes the classification errors in Z_1. S^* can be found by exploiting the theoretical relation between minimum-spanning tree (MST) and optimum-path tree for $f_{\max}$ [Papa 2008]. By computing a MST in the complete graph (Z_1, A), we obtain a connected acyclic graph whose nodes are all samples of Z_1 and the arcs are undirected and weighted by the distances d between adjacent samples). The spanning tree is optimum in the sense that the sum of its arc weights is minimum as compared to any other spanning tree in the graph. In the MST, every pair of samples is connected by a single path which is optimum according to $f_{\max}$. That is, the minimum-spanning tree contains one optimum-path tree for any selected root node. The optimum prototypes are the closest elements of the MST with different labels in Z_1. By removing the arcs between different classes, their adjacent samples become prototypes in S^* and OPF$_{cpl}$ algorithm can compute an optimum-path forest in Z_1.

For any sample $t \in Z_3$ (or $s \in Z_2$), we consider all arcs connecting t with samples $s \in Z_1$, as though t were part of the training graph. Considering all possible paths from S^* to t, we find the optimum path $P^*(t)$ from S^* and label t with the class $\lambda(R(t))$ of its most strongly connected prototype $R(t) \in S^*$. This path can be identified incrementally, by evaluating the optimum cost $C(t)$ as

$$C(t) = \min\{\max\{C(s), d(s,t)\}\}, \ \forall s \in Z_1. \quad (3)$$

Let the node $s^* \in Z_1$ be the one that satisfies Equation 3 (i.e., the predecessor $P(t)$ in the optimum path $P^*(t)$). Given that $L(s^*) = \lambda(R(t))$, the classification simply assigns $L(s^*)$ as the class of t. An error occurs when $L(s^*) \neq \lambda(t)$.

2.2. Optimum-Path Forest classifier using k-nn graph

Let $k \geq 1$ be a fixed number for the time being. An k-nn relation A_k is defined as follows. A sample $t \in Z_1$ is said adjacent to a sample $s \in Z_1$, if t is k-nearest neighbor of s according to $d(s,t)$. The pair (Z_1, A_k) then defines a k-nn graph for training. The arcs (s,t) are weighted by $d(s,t)$ and the nodes $s \in Z_1$ are weighted by a density value $\rho(s)$, given by

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

where $\sigma = \frac{d_f}{3}$ and d_f is the maximum arc weight in (Z_1, A_k). This parameter choice considers all nodes for density computation, since a Gaussian function covers most samples within $d(s,t) \in [0, 3\sigma]$. However the density value $\rho(s)$ be calculated with a Gaussian kernel, the use of the k-nn graph allows the proposed OPF to be robust to possible variations in the shape of the classes.

We define $f(\pi_t)$ such that its maximization for all nodes $t \in Z_1$ results into an optimum-path forest with roots at the maxima of the pdf, forming a root set $\mathcal{R}$. We expect that each class be represented by one or more roots (maxima) of the pdf. Each optimum-path tree in this forest represents the influence zone of one root $r \in \mathcal{R}$, which is composed by samples more strongly connected to r than to any other root. We expect that the training samples of a same class be assigned (classified) to an optimum-path tree rooted at a maximum of that class. The path-value function is defined as follows.

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

where $\delta = \min_{\forall (s,t) \in A_k | \rho(t) \neq \rho(s)} |\rho(t) - \rho(s)|$. The root set $\mathcal{R}$ is obtained on-the-fly. The method also uses the Image Foresting Transform (IFT) [Falcão et al. 2004] to maximize $f_1(\pi_t)$ and obtain an optimum-path forest P — a predecessor map with no cycles that assigns to each sample $t \notin \mathcal{R}$ its predecessor $P(t)$ in the optimum path $P^*(t)$ from $\mathcal{R}$ or a marker nil when $t \in \mathcal{R}$.

For any sample $t \in Z_3$ (or $s \in Z_2$), we consider the k-nearest neighbors connecting t with samples $s \in Z_1$, as though t were part of the graph. Considering all possible paths from $\mathcal{R}$ to t, we find the optimum path $P^*(t)$ with root $R(t)$ and label t with the class $\lambda(R(t))$. This path can be identified incrementally, by evaluating the optimum cost $V(t)$ as

$$V(t) = \max\{\min\{V(s), \rho(t)\}\}, \ \forall s \in Z_1. \quad (6)$$

Let the node $s^* \in Z_1$ be the one that satisfies the above equation. Given that $L(s^*) = \lambda(R(t))$, the classification simply assigns $L(s^*)$ to t.

2.3. Learning algorithm

There are many situations that limit the size of Z_1: large datasets, limited computational resources, and high computational time as required by some approaches. Mainly in applications with large datasets, it would be interesting to select for Z_1 the most informative

samples, such that the accuracy of the classifier is little affected by this size limitation. It is also important to show that a classifier can improve its performance along time of use, when we are able to teach it from its errors. This section presents a general learning algorithm which uses a third evaluation set Z_2 to improve the composition of samples in Z_1 without increasing its size.

From an initial choice of Z_1 and Z_2, the algorithm projects an instance I of a given classifier from Z_1 and evaluates it on Z_2. The misclassified samples of Z_2 are randomly selected and replaced by samples of Z_1 (under certain constraints). This procedure assumes that the most informative samples can be obtained from the errors. The new sets Z_1 and Z_2 are then used to repeat the process during a few iterations T. The instance of classifier with highest accuracy is selected along the iterations. The accuracy values $\mathcal{L}(I)$ obtained for each instance I form a *learning curve*, whose non-decreasing monotonic behavior indicates a positive learning rate for the classifier. Afterwards, by comparing the accuracies of the classifier on Z_3, before and after the learning process, we can evaluate its learning capacity from the errors. Notice that this learning algorithm can be used for any classifier, and here we applied it to OPF$_{cpl}$, OPF$_{knn}$, SVM and ANN-MLP.

3. Experimental results

The OPF-based algorithms were published in several Qualis A and B conferences and journals, and applied to a large range of applications, since that several medical applications [Papa et al. 2008b, Spadotto et al. 2008, Falcão et al. 2006, Falcão et al. 2007, Falcão et al. 2008, Cappabianco et al. 2008, Papa et al. 2009c, Iliev et al. 2009], biometrics recognition problems [Montoya-Zegarra et al. 2008b, Chiachia et al. 2009, Papa et al. 2009a], including face recognition and fingerprint identification, texture recognition [Montoya-Zegarra et al. 2008a, Montoya-Zegarra et al. 2007], rainfall estimation [Freitas et al. 2009, Papa et al. 2009b, Freitas et al. 2007] and land use classification [Pisani et al. 2009]. This last work have received a prize for one of the best accepted papers. Notice that due to the lack of space, some publications were omitted.

Table 1 shows the accuracies and execution time in seconds of the OPF$_{cpl}$, OPF$_{knn}$, SVM and ANN-MLP classification in some real and synthetic problems. The databases were splitted in 30% for training, 20% for evaluating (learning) and 50% for testing the algorithms. The accuracy is obtained as an average over 10 runnings for each experiment. In each running, training, evaluating and test sets were randomly generated. The results are displayed in the following format: $x \pm y(z)$, in which x, y and z are, respectively, the mean accuracy and its standard deviation and the mean execution time (normalized with respect to the dataset size). Several other results can be found in the aforementioned publications and in the João Paulo Papa's PhD. thesis [Papa 2008]. Note that OPF-based algorithms are extremely fast and accurate. On average, the results indicate that the OPF-based classifiers were about 72 times faster than SVM and 443 times faster than ANN-MLP [Papa 2008]. One of the main advantages of the OPF-based classifiers is that it does not interprets the classification task as a separating hyperplanes problem, such that SVM and ANN-MLP. The SVM characteristics turn its quadratic optimization problem into a suffering task in situations where large datasets is required. Still, the problem increases when the parameters of the nonlinear mapping (kernel functions) have to be found. Generally, this procedure is carried out by exhaustive search, which can make the training phase time prohibitive.

Dataset	OPF$_{cpl}$	OPF$_{knn}$	SVM	ANN-MLP
MPEG-7 Shape	87.24±0.01(0.0019)	**87.65±0.01(0.0038)**	87.18±0.01(0.2859)	79.23±0.04(0.5043)
CONE-TORUS Synthetic	**88.38±0.01(0.0010)**	86.28±0.02(0.0040)	87.95±0.03(0.1281)	85.58±0.04(1.8540)
SATURN Synthetic	88.70±0.02(0.0011)	**89.30±0.02(0.0011)**	**89.30±0.03(0.1445)**	88.10±0.04(0.3828)
WISCONSIN BREAST CANCER	94.17±0.12(0.0019)	94.75±0.11(0.0019)	**96.07±0.13(0.0091)**	93.26±1.10(0.01505)

Table 1. Mean accuracy, standard deviation and execution time in seconds for OPF$_{cpl}$, OPF$_{knn}$, SVM and ANN-MLP classifiers.

4. Conclusions

We introduced a novel and powerful framework for graph-based supervised classification, in which one can design OPF-based classifiers by just changing the path-cost function, adjacency relation and/or prototypes estimation methodology. We presented here two variants of OPF-based algorithms: OPF$_{cpl}$ and OPF$_{knn}$. The first one uses f_{max} as the path-cost function, complete graph as adjacency relation and MST approach for prototypes estimation, and the last one uses f_{min}, a k-nn adjacency relation and the prototypes are located in the maxima regions of the data feature space. We also developed the LibOPF, a library for the design of OPF-based classifiers, available in `http://www.ic.unicamp.br/~afalcao/libopf`.

We compared OPF-based classifiers with SVM and ANN-MLP using several datasets. The advantage of OPF-based techniques over the others in computational time is notorious and impressive, which is crucial in the case of large datasets. It can be more or less accurate than SVM, depending on the case, but its accuracy is usually superior to those of ANN-MLP. OPF-based classifiers also present some interesting properties. It is fast, simple, multi-class and does not make any assumption about the shape of the classes, and can handle some degree of overlapping between classes. An actual work concerns with reducing the training set without compromising the accuracy over the test set [Papa and Falcão 2009b, Papa et al. 2009b]. The main idea is to simulate a learning algorithm that can mark the training samples that participated from the classification process in the evaluation set. The remaining samples can be discarded from the training project or even so moved to the evaluating set to improve learning.

The OPF framework, as well its variants, learning algorithms and applications were evaluated and published in 14 international and 2 national conferences, 4 journals, 3 technical reports, 2 national and 1 international patent involving the automatic identification of human intestinal parasites in the last 2 years [Papa 2008], and have received a prize for the best accepted papers for remote sensing image classification [Pisani et al. 2009]. Nowadays, we have also submitted 1 conference and 3 journal articles.

References

Boser, B., Guyon, I., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Workshop on Computational Learning Theory*, pages 144–152, New York, NY, USA. ACM Press.

Cappabianco, F., Falcão, A., and Rocha, L. (2008). Clustering by optimum path forest and its application to automatic gm/wm classification in mr-t1 images of the brain. In *ISBI*, pages 428–431.

Chiachia, G., Marana, A., Papa, J., and Falcão, A. (2009). Infrared face recogniton by optimum-path forrest. In *Proc. of the 16th Intl. Conf. on Systems, Signals, and Image Processing*. Santorini, Greece (accepted).

Duda, R., Hart, P., and Stork, D. (2000). *Pattern Classification*. Wiley-Interscience.

Falcão, A., Stolfi, J., and Lotufo, R. (2004). The image foresting transform: Theory, algorithms, and applications. *IEEE Trans. on PAMI*, 26(1):19–29.

Falcão, A., Suzuki, C., Gomes, J., Candido, L., Shimizu, S., and Papa, J. (2008). Sistema para diagnóstico de parasitos intestinais por análise computadorizada de imagens e uso de referido sistema. Protocol number PI018080046387.

Falcão, A., Suzuki, C., Gomes, J., Papa, J., Candido, L., and Shimizu, S. (2006). Sistema para diagnóstico de parasitos intestinais por análise computadorizada de imagens. INPI Protocol- PI0605465-0.

Falcão, A., Suzuki, C., Gomes, J., Papa, J., Candido, L., and Shimizu, S. (2007). A system for diagnosing intestinal parasites by computerized image analysis. PCT Protocol - PCT/BR2007/000272.

Freitas, G., Ávila, A., Pinto, H., Papa, J., and Falcão, A. (2009). Optimum-path forest-based models for rainfall estimation. In *Proceedings of the 16th International Conference on Digital Signal Processing*. Santorini, Greece (accepted).

Freitas, G. M., Ávila, A. M. H., Pinto, H. S., and Papa, J. P. (2007). Classifiers evaluation for agricultural areas monitoring. In *Proceedings of the XV Brazilian Congress on Agrometeorology*, volume 1, pages 25–29. Aracaju, Brazil.

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall.

Iliev, A., Scordilis, M., Papa, J., and Falcão, A. (2009). Spoken emotion recognition through optimum-path forest classification using glottal features. *Computer, Speech, and Language*. accepted for publication.

Montoya-Zegarra, J., Papa, J., Leite, N., Torres, R., and Falcão, A. (2007). Rotation-invariant texture recognition. In *3rd International Symposium on Visual Computing*, volume Part II, LNCS 4842, pages 193–204, Lake Tahoe, Nevada, CA, USA. Springer.

Montoya-Zegarra, J., Papa, J., Leite, N., Torres, R., and Falcão, A. (2008a). Learning how to extract rotation-invariant and scale-invariant features from texture images. *EURASIP Journal on Advances in Signal Processing*, 2008:1–16.

Montoya-Zegarra, J., Papa, J., Leite, N., Torres, R., and Falcão, A. (2008b). Novel approaches for exclusive and continuous fingerprint classification. In *Proc. of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology*, pages 386–397, Berlin, Heidelberg. Springer-Verlag.

Papa, J. (2008). Supervised pattern classification using optimum-path forest. PhD. Thesis - Institute of Computing, University of Campinas, available at http://www.ic.unicamp.br/˜jpaulo/files/tese_papa.pdf.

Papa, J., Falcão, A., Levada, A., Corrêa, D., Salvadeo, D., and Mascarenhas, N. (2009a). Fast and accurate holystic face recognition through optimum-path forest. In *16th International Conference on Digital Signal Processing*. Santorini, Greece (accepted).

Papa, J., Falcão, A., Suzuki, C., and Mascarenhas, N. (2008a). A discrete approach for supervised pattern recognition. In *12th International Workshop on Combinatorial Image Analysis*, volume 4958, pages 136–147. LNCS Springer Berlin/Heidelberg.

Papa, J., Freitas, G., Falcão, A., and Ávila, A. (2009b). Robust training patterns pruning for optimum-path forest and its application for satellite-based rainfall estimation. *IEEE Geoscience and Remote Sensing Letters*. (accepted for publication).

Papa, J., Spadotto, A., ao, A. F., and Pereira, J. (2008b). Optimum path forest classifier applied to laryngeal pathology detection. In *Proc. of the 15th Intl. Conf. on Systems, Signals, and Image Processing*, volume 1, pages 249–252, Bratislava, Slovakia. IEEE.

Papa, J., Spadotto, A., Falcão, A., Guido, R., and Pereira, J. (2009c). Robust biomedical signal recognition though optimum-path forest. *Artificial Intelligence in Medicine*. submitted.

Papa, J. P. and Falcão, A. X. (2008). A new variant of the optimum-path forest classifier. In *ISVC '08: Proceedings of the 4th International Symposium on Advances in Visual Computing*, pages 935–944, Berlin, Heidelberg. Springer-Verlag.

Papa, J. P. and Falcão, A. X. (2009a). A learning algorithm for the optimum-path forest classifier. In *7th IAPR Workshop on Graph-based Representations in Pattern Recognition*, volume 5534, pages 195–204, Berlin, Heidelberg. Springer-Verlag.

Papa, J. P. and Falcão, A. X. (2009b). On the training patterns pruning for optimum-path forest. In *15th Intl. Conf. on Image Analysis and Processing*. Salerno, Italy (accepted).

Papa, J. P., Falcão, A. X., Miranda, P. A. V., Suzuki, C. T. N., and Mascarenhas, N. D. A. (2007). Design of robust pattern classifiers based on optimum-path forests. In *Mathematical Morphology and its Applications to Image and Signal Processing (ISMM'07)*, pages 337–348. MCT/INPE.

Papa, J. P., Falcão, A. X., and Suzuki, C. T. N. (2009d). Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology*, 19:120–131.

Pisani, R., Papa, J., Zimback, C., Falcão, A., and Barbosa, A. (2009). Land use classification using optimum-path forest. In *Proceedings of the 14th Brazilian Symposium on Remote Sensing*, pages 7063–7070. MCT-INPE.

Spadotto, A., Pereira, J., Guido, R., Papa, J., Falcão, A., Gatto, A., Cola, P., and Schelp, A. (2008). Oropharyngeal dysphagia identification using wavelets and optimum path forest. In *Proc. of the 3rd IEEE Intl. Symp. on Communications, Control and Signal Processing*, pages 735–740.

Tang, B. and Mazzoni, D. (2006). Multiclass reduced-set support vector machines. In *Proceedings of the 23th International Conference on Machine learning*, pages 921–928. ACM Press.