Chapter 1

Recent advances on optimum-path forest for data classification: supervised, semi-supervised and unsupervised learning

João Paulo Papa, Willian Paraguassu Amorim, Alexandre Xavier Falcão, and João Manuel R. S. Tavares

papa@fc.unesp.br, paraguassuec@gmail.com
afalcao@ic.unicamp.br, tavares@fe.up.pt

Although one can find several pattern recognition techniques out there, there is still room for improvements and new approaches. In this book chapter, we revisited the Optimum-Path Forest (OPF) classifier, which has been evaluated over the last years in a number of applications that consider supervised, semi-supervised and unsupervised learning problems. We also presented a brief compilation of a number of previous works that employed OPF in different research fields, that range from remote sensing image classification to medical data analysis.

1. Introduction

Pattern recognition techniques have been paramount in the last years, mainly due to the increasing number of applications that often require some decision-making mechanism based on artificial intelligence. Depending on the amount of knowledge one has about the training set, we can divide pattern recognition techniques in three main segments: (i) supervised ones, which are allowed to use the information of all training set, i.e., the samples' label; (ii) semi-supervised techniques, which have a partially labeled training set only; and (iii) unsupervised or the so-called clustering techniques, in which one has no labeled samples.[1]

The number of techniques out there has increased yearly, but the most widely used ones still remain on a reduced set of them. Support Vector Machines (SVMs),[2] Artificial Neural Networks[3] and Bayesian classifiers[1] seem to be the most frequent employed techniques. While the latter two might be the most traditional ones, SVMs have gained popularity in the last decade due to their good power of generalization over unseen data.

Despite their popularity, some very famous techniques often require a high computational burden for learning the model that best fits to the training data. In addition, a considerable number of techniques have several parameter to be fine-tuned also, turning their usage a laborious and time-consuming task. Some years ago, Papa et al.[4,5] and Rocha et al.[6] proposed the supervised and unsupervised versions of the Optimum-Path Forest (OPF) classifier, respectively. The main idea

behind OPF-based classification is to rule a competition process among some key samples (*prototypes*), which aim at conquering the remaining samples offering to them optimum-path costs according to some path-cost function. The final result of such competition is an optimum-path tree (OPT) rooted at each prototype node, being the collection of all OPTs an optimum-path forest that names the classifier.

Roughly speaking, OPF can be understood as a framework to the design of classifiers based on optimum-path connectivity, not as a sole classifier. Therefore, depending on the adjacency relation, path-cost function and methodology used to choose prototypes, one can design a new OPF classifier. Supervised versions with full-connected and k-nearest neighbors (k-nn) graphs have been addressed by Papa et al.[4] and Papa and Falcão,[7] respectively. The unsupervised version was firstly introduced by Rocha et al.,[6] and further used in a number of different applications, such intrusion detection in computer networks,[8] theft identification in power distribution systems, land-use classification in satellite imagery,[9] medical image analysis,[10] and anomaly detection,[11] just to name a few. In regard to the supervised approach, one can refer to several works, such as supervised intrusion detection in computer networks,[12] hyperspectral band selection,[13] non-technical losses detection in power distribution systems,[14] automatic aquatic weed recognition,[15] and as wrapper approaches for feature selection.[16,17]

Some large-scale-oriented approaches have been proposed also. Papa et al.,[18] for instance, proposed a learning algorithm that allows OPF to design more compact and representative training sets, and Papa et al.[5] introduced a faster OPF classification for the test phase. Later on, Osaku et al.[19] presented a contextual-based OPF to classify remote sensing images, and Pereira et al.[20] introduced a sequential learning approach for the same context using supervised OPF. A multi-label version of the OPF classifier was applied for video classification by Pereira et al.,[21] as well as an OPF-based video summarization approach was proposed by Martins et al.[22] Amorim[23] presented a semi-supervised learning algorithm for the Optimum-Path Forest classifier. Roughly speaking, the idea is to train OPF over the training set, and then label the samples of the unlabeled set. Soon after, the OPF is trained again over the whole training set, i.e., the original labeled and the predicted samples. Their approach outperformed standard OPF in some semi-supervised tasks.

In this work, we present a brief compilation of some recent works on Optimum-Path Forest classification considering supervised, semi-supervised and unsupervised learning problems. The remainder of this chapter is divided as follows. Section 2 presents the OPF background theory, and Section 3 discusses some experimental results. Finally, Section 4 states conclusions and future works.

2. Optimum-Path Forest Classification

In this section, we present the main theoretical background regarding supervised, semi-supervised and unsupervised learning through Optimum-Path Forest.

2.1. *Supervised Learning*

The OPF framework is a recent highlight to the development of pattern recognition techniques based on graph partitions. The nodes are the data samples, which are represented by their corresponding feature vectors, and are connected according to some predefined adjacency relation. Given some key nodes (prototypes), they will compete among themselves aiming at conquering the remaining nodes. Thus, the algorithm outputs an optimum path forest, which is a collection of optimum-path trees (OPTs) rooted at each prototype. This work employs the OPF classifier proposed by Papa et al.,[4,5] which is explained in more details as follows.

Let $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ be a labeled dataset, such that $\mathcal{D}_1$ and $\mathcal{D}_2$ stands for the training and test sets, respectively. Let $\mathcal{S} \subset \mathcal{D}_1$ be a set of prototypes of all classes (i.e., key samples that best represent the classes). Let $(\mathcal{D}_1, A)$ be a complete graph whose nodes are the samples in $\mathcal{D}_1$, and any pair of samples defines an arc in $A = \mathcal{D}_1 \times \mathcal{D}_1$ (Figure 1a)*. Additionally, let π_s be a path in $(\mathcal{D}_1, A)$ with terminus at sample $s \in D_1$.

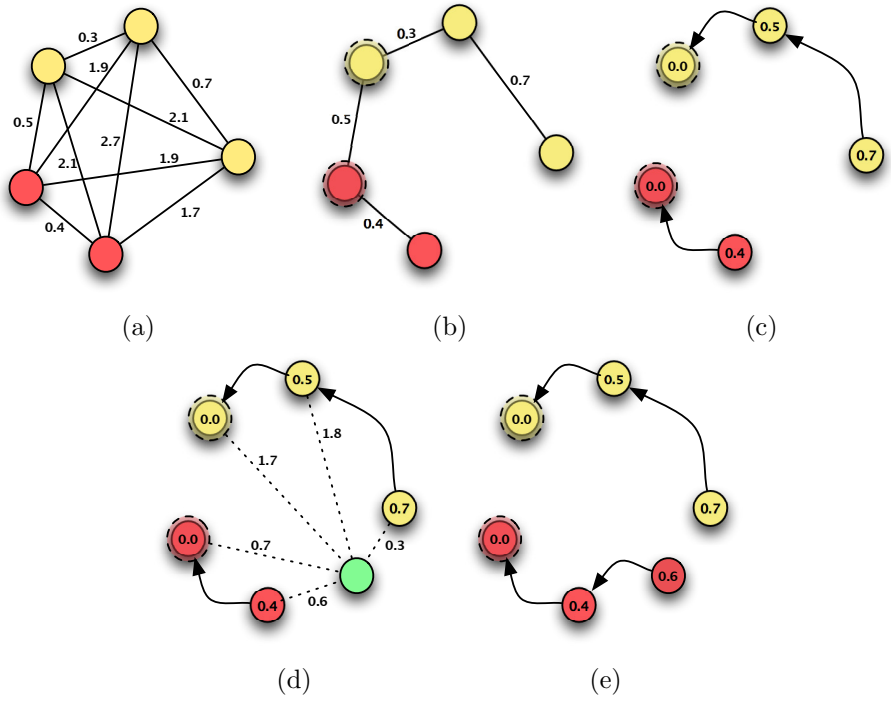

Fig. 1. (a) Training set modeled as a complete graph, (b) a minimum spanning tree computation over the training set (prototypes are highlighted), (c) optimum-path forest over the training set, (d) classification process of a "green" sample, and (e) test sample is finally classified.

The OPF algorithm proposed by Papa et al.[4,5] employs the path-cost function

*The arcs are weighted by the distance between their corresponding nodes.

f_{max} due to its theoretical properties for estimating prototypes (Section 2.1.1 gives further details about this procedure):

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

where $d(s,t)$ stands for a distance between nodes s and t, such that $s,t \in \mathcal{D}_1$. Therefore, $f_{max}(\pi_s)$ computes the maximum distance between adjacent samples in π_s, when π_s is not a trivial path. In short, the OPF algorithm tries to minimize $f_{max}(\pi_t)$, $\forall t \in \mathcal{D}_1$.

2.1.1. *Training*

We say that $\mathcal{S}^*$ is an optimum set of prototypes when the OPF algorithm minimizes the classification errors for every $s \in \mathcal{D}_1$. We have that S^* can be found by exploiting the theoretical relation between the minimum-spanning tree and the optimum-path tree for f_{max}.[24] The training essentially consists of finding S^* and an OPF classifier rooted at S^*. By computing a Minimum Spanning Tree (MST) in the complete graph $(\mathcal{D}_1, A)$ (Figure 1b), one obtain a connected acyclic graph whose nodes are all samples of $\mathcal{D}_1$ and the arcs are undirected and weighted by the distances d between adjacent samples. In the MST, every pair of samples is connected by a single path, which is optimum according to f_{max}. Hence, 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 $\mathcal{D}_1$ (i.e., elements that fall in the frontier of the classes, as highlighted in Figure 1b). By removing the arcs between different classes, their adjacent samples become prototypes in S^*, and OPF algorithm can define an optimum-path forest with minimum classification errors in $\mathcal{D}_1$ (Figure 1c).

2.1.2. *Classification*

For any sample $t \in \mathcal{D}_2$, we consider all arcs connecting t with samples $s \in \mathcal{D}_1$, as though t were part of the training graph (Figure 1d). 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 $R(t) \in S^*$ (Fig. 1e). This path can be identified incrementally, by evaluating the optimum cost $C(t)$ as follows:

$$C(t) = \min\{\max\{C(s), d(s,t)\}\}, \ \forall s \in D_1. \quad (2)$$

Let the node $s^* \in \mathcal{D}_1$ be the one that satisfies Equation 2 (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. *Unsupervised Learning*

Let $\mathcal{D}$ be an unlabeled dataset now. A graph $(\mathcal{D}, \mathcal{A}_k)$ can be defined such that the arcs $(s,t) \in \mathcal{A}$ connect k-nearest neighbors in the feature space. Now, the nodes $s \in \mathcal{D}$ are weighted by a probability density value $\rho(s)$:

$$\rho(s) = \frac{1}{\sqrt{2\pi\sigma^2}|\mathcal{A}_k(s)|} \sum_{\forall t \in \mathcal{A}_k(s)} \exp\left(\frac{-d^2(s,t)}{2\sigma^2}\right), \tag{3}$$

where $|\mathcal{A}_k(s)| = k$, $\sigma = \frac{d_f}{3}$, and d_f is the maximum arc weight in $(\mathcal{D}, \mathcal{A}_k)$. This parameter choice considers all adjacent nodes for density computation, since a Gaussian function covers most samples within $d(s,t) \in [0, 3\sigma]$. Moreover, since $\mathcal{A}_k$ is asymmetric, symmetric arcs must be added to it on the plateaus of the probability density function (pdf) in order to guarantee a single root per maximum.

The traditional method to estimate a pdf is by Parzen-window. Equation (3) can provide the Parzen-window estimation based on an isotropic Gaussian kernel when we define the arcs by $(s,t) \in \mathcal{A}_k$ if $d(s,t) \leq d_f$. However, this choice presents problems with the differences in scale and sample concentration. Solutions for this problem lead to adaptive choices of d_f depending on the region of the feature space.[25] By taking into account the k-nearest neighbors, the method handles different concentrations and reduces the scale problem to the one of finding the best value of k, say k^* within $[k_{\min}, k_{\max}]$, for $1 \leq k_{\min} < k_{\max} \leq |\mathcal{D}|$.

The solution proposed by Rocha *et al.*[6] to find k^* considers the minimum graph cut among all clustering results for $k \in [1, k_{\max}]$ ($k_{\min} = 1$), according to the normalized measure $GC(\mathcal{A}_k, L, d)$ suggested by Shi and Malik:[26]

$$GC(\mathcal{A}_k, L, d) = \sum_{i=1}^{c} \frac{W_i'}{W_i + W_i'}, \tag{4}$$

$$W_i = \sum_{\forall (s,t) \in \mathcal{A}_k | L(s) = L(t) = i} \frac{1}{d(s,t)}, \tag{5}$$

$$W_i' = \sum_{\forall (s,t) \in \mathcal{A}_k | L(s) = i, L(t) \neq i} \frac{1}{d(s,t)}, \tag{6}$$

where $L(t)$ is the label of sample t, W_i' uses all arc weights between cluster i and other clusters, and W_i uses all arc weights within cluster $i = 1, 2, \ldots, c$.

The method defines a path π_t as a sequence of adjacent samples starting from a root $R(t)$ and ending at a sample t, being $\pi_t = \langle t \rangle$ a trivial path and $\pi_s \cdot \langle s,t \rangle$ the concatenation of π_s and arc (s,t). It assigns to each path π_t a value $f(\pi_t)$ given by a connectivity function f. A path π_t is considered optimum if $f(\pi_t) \geq f(\tau_t)$ for any other path τ_t.

Among all possible paths π_t from the maxima of the pdf, the method assigns to t a path whose minimum density value along it is maximum. That is, the method

finds $V(t) = \max_{\forall \pi_t \in (\mathcal{D}, \mathcal{A}_k)}\{f(\pi_t)\}$ for $f(\pi_t)$ defined by:

$$f(\langle t \rangle) = \begin{cases} \rho(t) & \text{if } t \in \mathcal{R} \\ \rho(t) - \delta \text{ otherwise,} \end{cases}$$
$$f(\langle \pi_s \cdot \langle s, t \rangle \rangle) = \min\{f(\pi_s), \rho(t)\}, \tag{7}$$

for $\delta = \min_{\forall (s,t) \in \mathcal{A}_k | \rho(t) \neq \rho(s)} |\rho(t) - \rho(s)|$ and $\mathcal{R}$ being a root set, discovered on-the-fly, with one element per each maximum of the pdf. It should be noted that higher values of δ reduce the number of maxima. We are setting $\delta = 1.0$ and scaling real numbers $\rho(t) \in [1, 1000]$ in this work. The OPF algorithm maximizes the connectivity map $V(t)$ by computing an optimum-path forest — a predecessor map P with no cycles that assigns to each sample $t \notin \mathcal{R}$ its predecessor $P(t)$ in the optimum path from $\mathcal{R}$ or a marker *nil* when $t \in \mathcal{R}$.

2.3. *Semi-supervised Learning*

This section describes the approach proposed by Amorim et al.,[27] which is a variant of the first semi-supervised OPF introduced by Amorim et al.[23]

Let our dataset $\mathcal{D}$ be divided again into subsets $\mathcal{D}_1$ for the design of the classifier (training) and $\mathcal{D}_2$ for testing its generalization ability. Set $\mathcal{D}_1 = \mathcal{D}_1^l \cup \mathcal{D}_1^u$ will also consist of labeled $\mathcal{D}_1^l$ and unlabeled $\mathcal{D}_1^u$ subsets of samples. For training, the method must connect samples from $\mathcal{D}_1$ into a graph and propagate labels to $\mathcal{D}_1^u$ such that the classifier will be an optimum-path forest rooted at $\mathcal{D}_1^l$. The classification of new samples from $\mathcal{D}_2$ is done by evaluating extended optimum paths. The algorithms for each step are described next.

2.3.1. *Training*

In regard to the training set, we first consider an adjacency relation $A = \mathcal{D}_1 \times \mathcal{D}_1$, which defines a complete and weighted graph $(\mathcal{D}_1, A, d)$, and computes a Minimum Spanning-Tree $(\mathcal{Z}_1, B, d)$ as the input graph for optimum-path forest computation. In $(\mathcal{D}_1, B, d)$, the number of arcs $|B| \ll |A|$ makes the optimum-path forest computation considerably fast. The arcs in B already connect the closest labeled and unlabeled samples, but a node $t \in \mathcal{D}_1^u$ can still be reached by paths from nodes of $\mathcal{D}_1^l$ with distinct labels. Therefore, labeled nodes will compete with each other and the label $L(t) \leftarrow \lambda(s)$ assigned to t will come from its most closely connected node $s \in \mathcal{D}_1^l$. A label propagation error occurs when $\lambda(t) \neq \lambda(s)$.

In short, we first consider a graph $(\mathcal{D}_1, A, d)$ with connectivity function f_{mst}, which generates an MST $(\mathcal{Z}_1, B, d)$, and then we consider the MST with connectivity

function $f_{\max}$, as follows:

$$f_{mst}(\langle s \rangle) = \begin{cases} 0 & \text{for one arbitrary } s \in \mathcal{Z}_1, \\ +\infty & \text{otherwise,} \end{cases}$$
$$f_{mst}(\pi_s \cdot \langle s,t \rangle) = d(s,t), \tag{8}$$
$$f_{\max}(\langle s \rangle) = \begin{cases} 0 & \text{if } s \in \mathcal{Z}_1^l, \\ +\infty & \text{otherwise,} \end{cases}$$
$$f_{\max}(\pi_s \cdot \langle s,t \rangle) = \max\{f_{\max}(\pi_s), d(s,t)\}. \tag{9}$$

2.3.2. *Classification*

For classification, the optimum paths π_s must be extended to new samples $t \in \mathcal{Z}_2$ by considering

$$C(t) = \min_{\forall s \in \mathcal{D}_1} \{\max\{C(s), d(s,t)\}\}, \tag{10}$$

and assigning to t the label $L(t) = L(s^*)$ of the sample $s^* \in \mathcal{D}_1$ for which $\pi_{s^*} \cdot \langle s^*, t \rangle$ is optimum. Note that classification considers that t is connected to all nodes in $\mathcal{D}_1$, rather than t being an additional node to the MST. Therefore, classification is based on the same rule used for the supervised OPF classifier, but now using a much larger set $\mathcal{D}_1 > \mathcal{D}_1^l$. Moreover, by following the order of nodes in $\mathcal{D}_1$, the evaluation of $C(t)$ can halt whenever $C(s) \geq \max\{C(s^*), d(s^*,t)\}$ for some previous $s^* \in \mathcal{D}_1$.

In Figure 2, we can better understand the practical behavior of the semi-supervised OPF. Given the selected labeled and unlabeled datasets (Figure 2(a)), we applied the semi-supervised OPF, and the propagation of the unlabeled samples can be seen in Figure 2(b). Figure 2(c) shows the addition of test samples within each class, and Figure 2(d) shows the classification results. Figure 2(e) depicts the addition of test samples within each class together with test samples outside the contour of the unlabeled samples, and Figure 2(f) shows the classification results.

3. Experiments

In this section, we compile some experiments conducted in recent works with respect to supervised, semi-supervised and unsupervised learning.

3.1. *Supervised Classification of Radar Images*

These results were obtained by Pereira et al.,[20] which evaluated the OPF classifier in the context of stacked sequential learning. Figure 3a displays the radar image used in that work, as well as its ground-truth version in 3b. The data is a dual-polarized (HH and HV) image obtained from ALOS-PALSAR radar covering the area of Duque de Caxias, RJ-Brazil.

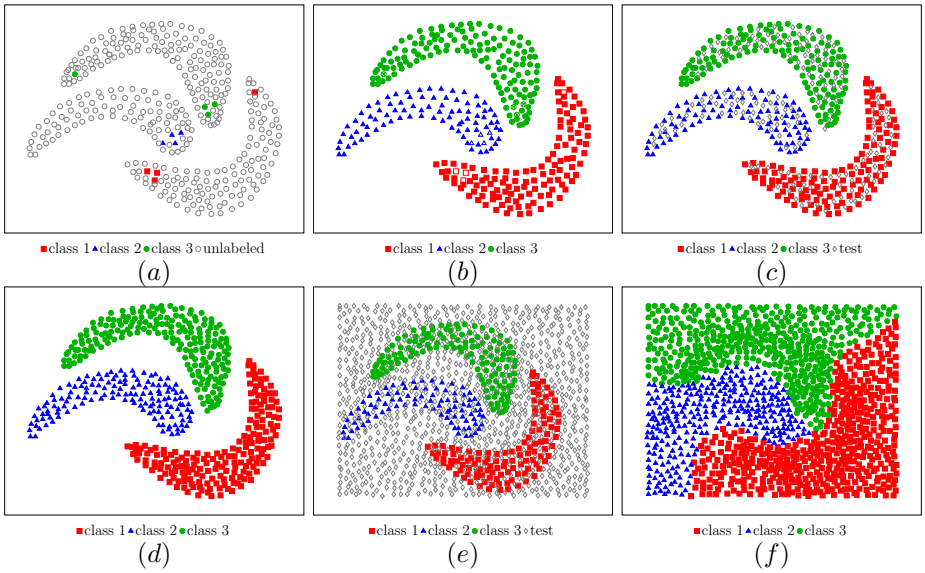

Fig. 2. A practical behavior of semi-supervised OPF technique. (a) Labeled and unlabeled data sets, (b) label propagation (c) test samples within each class, (d) result of semi-supervised OPF classification, (e) test samples within each class and outside the border of the unlabeled samples, (f) result of semi-supervised OPF classification.

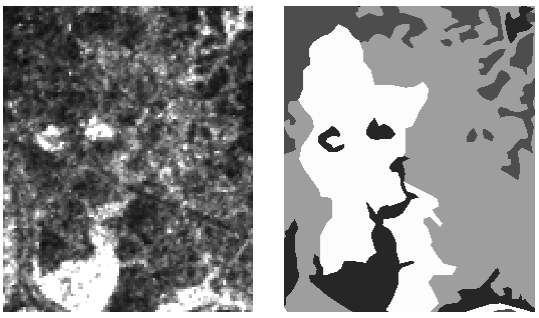

Fig. 3. ALOS-PALSAR images: (a) original (HV) and its (b) ground truth.

The authors compared standard OPF† classifier against with four different sequential learning approaches: OPF with Stacked Sequential Learning[28] (OPF-SSL), OPF with Sliding Window[29] (OPF-SW), OPF with Multi-Scale Sequential Learning and Multi-resolution-based decomposition (OPF-MSSL-MR),[30] and OPF with Multi-Scale Sequential Learning and Pyramid-based decomposition (OPF-MSSL-PY).[30] They also evaluated the influence of different training set sizes with 1%,

†It was employed the LibOPF, which is an open-source library to the design of optimum-path forest-based classifiers.

3% and 5% of the entire image, being the remaining pixels used to compose the test set. In order to allow a robust statistical evaluation, the authors performed a cross-validation procedure over 20 runnings for further computation of the Wilcoxon signed-rank test.[31] Additionally, each pixel has been described by its RGB values to compose the dataset samples for pattern classification purposes[‡].

Table 1 presents the mean accuracy results using 1%, 3% and 5% of the dataset for training, being the remaining samples used for classification purposes[§]. The most accurate techniques considering the Wilcoxon signed-rank test are in bold. Roughly speaking, better results concerning sequential learning methodologies can be observed, as well as OPF-SSL improved its recognition rates when the authors used more training samples, since it is based on the labelling of neighboring samples, but its recognition rate using 5% of training samples dropped when compared to the experiment using 3% of the training set, meaning it may be affected by misclassified samples generated by standard OPF in the initial phase.

Technique	Accuracy (1%)		Accuracy (3%)		Accuracy (5%)	
	CBERS	ALOS-PALSAR	CBERS	ALOS-PALSAR	CBERS	ALOS-PALSAR
OPF	68.22±0.0117	75.13±0.0555	67.33±0.0159	75.22±0.1746	66.42±0.0160	76.05±0.0627
OPF-SSL	60.81±0.0112	75.38±0.0669	63.04±0.0124	76.26±0.00507	61.57±0.0080	77.32±0.0335
OPF-MSSL-MR	73.70±0.0067	**76.34±0.0432**	75.84±0.0027	76.69±0.0419	76.94±0.0041	**78.23±0.0370**
OPF-MSSL-PY	**74.82±0.0063**	75.46±0.0350	**76.90±0.0040**	76.41±0.0653	**78.09±0.0025**	76.69±0.0938
OPF-SW	60.81±0.0112	75.09±0.0804	**77.13±0.0041**	69.80±0.0039	**78.17±0.0040**	77.22±0.0203

3.2. *Unsupervised Recognition of Thefts in Power Distribution Systems*

In this work, Passos et al.[11] evaluated the unsupervised OPF in the context of thefts (i.e., non-technical losses) in power distribution systems. In that work, the authors used two labeled datasets obtained from a Brazilian electric power company, named B_i and B_c. The former is a dataset composed of 3,178 industrial profiles, and the latter contains 4,948 commercial profiles. Each industrial and commercial profile is represented by eight features, as follows:

(1) Demand Billed (DB): demand value of the active power considered for billing purposes, in kilowatts (kW);
(2) Demand Contracted (DC): the value of the demand for continuous availability requested from the energy company, which must be paid whether the electric power is used by the consumer or not, in kilowatts (kW);
(3) Demand Measured or Maximum Demand (D_{max}): the maximum actual demand for active power, verified by measurement at fifteen-minute intervals during the

[‡]In this work, the authors used an 8-neighborhood system for OPF-SSL and OPF-SW techniques, and an 11- and a 3-neighborhood systems for OPF-MSSL-MR and OPF-MSSL-PY, respectively. They also employed 7 scales of decomposition for OPF-MSSL-MR and 5 scales of decomposition for OPF-MSSL-PY. Such values have been empirically set.

[§]The authors employed the very same accuracy measure proposed by Papa et al.[4] that considers unbalanced datasets, which is often faced in land-cover image classification.

billing period, in kilowatts (kW);
(4) Reactive Energy (RE): energy that flows through the electric and magnetic fields of an AC system, in kilovolt-amperes reactive hours (kVArh);
(5) Power Transformer (PT): the power transformer installed for the consumers, in kilovolt-amperes (kVA);
(6) Power Factor (PF): the ratio between the consumed active and apparent power in a circuit. The PF indicates the efficiency of a power distribution system;
(7) Installed Power (P_{inst}): the sum of the nominal power of all electrical equipment installed and ready to operate at the consumer unit, in kilowatts (kW);
(8) Load Factor (LF): the ratio between the average demand ($D_{average}$) and maximum demand (D_{max}) of the consumer unit. The LF is an index that shows how the electric power is used in a rational way.

The commercial dataset contains 4,680 regular consumers (94.58%) and 268 irregular profiles, while the industrial dataset contains 2,984 samples (93.89%) that represent regular consumers, and 194 samples that denote irregular consumers. Notice all the aforementioned features are measured over one month.

The authors compared OPF against with the well-known k-means and a Gaussian Mixture Model (GMM). Tables 2 and 3 present the average accuracy and F-measure for each clustering technique using the parameters obtained in the previous step, respectively. Notice the most accurate results according to the Wilcoxon statistical test are in bold. As one can observe, OPF and k-means have obtained the best results considering both measures, followed by GMM. Such behavior is in agreement with the results obtained during the fine-tuning step. Although all techniques have obtained very close F-measure values considering B_c dataset (Table 3), OPF achieved an F-measure of 0.9747, k-Means obtained 0.9746 and GMM achieved 0.9706, which may explain the bolded results as being the best ones for OPF and k-Means. Notice OPF, k-means and GMM parameters have been optimized through cross-validation, i.e., we choose k_{max} (OPF) and the number of clusters (k-means and GMM) that optimized the classification accuracy over a validating set.

Technique	B_c	B_i
OPF	**81.57%±2.48**	**78.30%±3.11**
GMM	74.64%±3.90	74.80%±4.35
k-Means	**81.51%±3.71**	**77.88%±3.48**

Technique	B_c	B_i
OPF	**0.98±0.002**	**0.97±0.003**
GMM	0.97±0.003	0.97±0.005
k-Means	**0.98±0.003**	**0.97±0.004**

3.3. *Semi-supervised Learning with Optimum-Path Forest*

In this section, we described the work by Amorim et al.,[27] which presented a new semi-supervised OPF based on the one proposed by.[23] The authors addressed the task of classification in the context of binary and multi-class datasets with small and large numbers of features. The semi-supervised OPF proposed by Amorim et al.,[27] i.e., OPFSEMI$_{mst}$, was compared against with the semi-supervised version proposed by Amorim et al.[23] (OPFSEMI), traditional OPF, SVMs, Transductive SVMs (TSVMs) (as implemented in UniverSVM[32]), and SemiL — an efficient tool for solving large scale semi-supervised learning or transductive inference problems using Harmonic Gaussian method. A detailed comparison of the UniverSVM approach with other semi-supervised optimization schemes has been conducted by Collobert et al.[32] Their results demonstrated the superior performance of UniverSVM when compared to other related methods.

The experiments were conducted to evaluate different proportions between labeled ($\mathcal{D}_1^l$) and unlabeled ($\mathcal{D}_1^u$) samples for a fixed training set ($\mathcal{D}_1$) size. In such a case, the semi-supervised approaches employ the whole $\mathcal{Z}_1$ for the training step, while the supervised approaches shall use $\mathcal{Z}_1^l$ only (OPF and SVMs). The accuracy was measured as suggested in.[4] For this purpose, we used six datasets from various domains. Table 3.3 contains the number of samples, classes, and attributes of the seven datasets used in the experiments. The first four datasets are synthetic and publicly available. The last two (Cowhide and Parasites) were obtained from real applications. The Cowhide dataset is composed of five types of regions of interest in the Wet-Blue[¶] processing stage, namely: scabies, ticks, hot-iron, cut, and regions without defect.

Dataset	Samples	Attributes	Classes
Statlog[33]	2.310	19	7
Spambase[34]	4.601	57	2
Faces[35]	1.864	162	54
Pendigits[36]	10.992	16	10
Cowhide[37]	1.690	160	5
Parasites[38]	1.660	262	15

The sizes of $\mathcal{D}_1^l$ and $\mathcal{D}_1^u$ ranged from 1%–99% and 10%–90% to 50%–50% of $\mathcal{Z}_1$, as shown in Tables 5 and 6 for each dataset. These tables show the performance (mean accuracy and standard deviation) of each method over $\mathcal{D}_2$. Table 6 also shows the percentages of the propagation error of the unlabeled samples (*le.*) and the training time (*et.*) in seconds concerning the semi-supervised OPF approaches. The values in bold indicate the most accurate results.

[¶]Wet-Blue leather is an intermediate stage between untanned and finished leather.

		Cowhide data set—Classification rate				Spambase data set—Classification rate.			
$\mathcal{D}_1^l$	$\mathcal{D}_1^u$	**OPF**	**SVM**	**TSVM**	**SemiL**	**OPF**	**SVM**	**TSVM**	**SemiL**
1	99	82.71±0.038	85.19±0.035	83.14±0.085	83.81±0.007	58.76±0.042	60.75±0.026	60.37±0.082	58.59±0.085
10	90	90.88±0.107	85.55±0.033	84.25±0.041	81.61±0.016	65.89±0.017	64.15±0.066	**67.98±0.069**	62.20±0.102
20	80	93.19±0.038	85.86±0.085	89.48±0.019	80.14±0.012	66.13±0.038	**74.75±0.031**	70.13±0.082	66.31±0.078
30	70	93.59±0.018	86.45±0.050	86.48±0.059	81.58±0.042	67.54±0.096	**76.39±0.054**	73.41±0.014	70.23±0.066
40	60	94.29±0.031	86.09±0.083	86.68±0.097	82.63±0.075	68.99±0.044	**76.95±0.075**	69.84±0.007	71.34±0.014
50	50	95.77±0.069	87.76±0.040	90.04±0.057	91.36±0.096	70.16±0.099	**77.48±0.056**	71.03±0.013	71.85±0.099
		Statlog data set—Classification rate.				Pendigits data set—Classification rate.			
$\mathcal{D}_1^l$	$\mathcal{D}_1^u$	**OPF**	**SVM**	**TSVM**	**SemiL**	**OPF**	**SVM**	**TSVM**	**SemiL**
1	99	84.75±0.061	81.45±0.022	73.61±0.013	78.62±0.077	94.31±0.076	75.05±0.025	74.23±0.088	75.20±0.015
10	90	88.76±0.031	85.41±0.028	88.68±0.081	87.20±0.111	96.54±0.015	70.27±0.064	70.43±0.083	74.50±0.035
20	80	87.11±0.084	90.33±0.054	85.05±0.042	83.80±0.056	98.88±0.092	79.07±0.028	76.80±0.094	86.60±0.076
30	70	91.6±0.084	92.33±0.088	89.13±0.048	80.22±0.048	99.19±0.099	87.68±0.078	88.57±0.078	97.61±0.010
40	60	92.54±0.038	93.20±0.033	89.22±0.015	88.38±0.071	99.14±0.053	91.22±0.035	82.88±0.023	97.13±0.098
50	50	93.14±0.051	93.25±0.079	89.01±0.067	90.29±0.080	99.17±0.088	97.87±0.020	85.65±0.047	98.06±0.026
		Faces data set—Classification rate.				Parasites data set—Classification rate.			
$\mathcal{Z}_1^l$	$\mathcal{Z}_1^u$	**OPF**	**SVM**	**TSVM**	**SemiL**	**OPF**	**SVM**	**TSVM**	**SemiL**
1	99	80.77±0.089	75.13±0.019	68.45±0.070	79.25±0.064	88.09±0.092	78.15±0.035	71.78±0.016	82.52±0.058
10	90	85.47±0.038	71.31±0.051	77.61±0.051	83.81±0.021	96.11±0.015	94.45±0.029	91.84±0.108	88.25±0.012
20	80	92.95±0.072	80.81±0.002	84.38±0.088	84.37±0.082	97.26±0.084	98.56±0.047	89.32±0.007	84.33±0.073
30	70	95.43±0.015	82.24±0.017	80.34±0.052	87.49±0.041	98.00±0.114	98.41±0.068	94.22±0.035	88.11±0.111
40	60	97.15±0.13	90.24±0.013	84.35±0.036	90.38±0.044	97.93±0.039	97.79±0.013	95.85±0.061	86.66±0.019
50	50	97.48±0.023	97.13±0.008	90.04±0.059	93.61±0.094	98.42±0.031	**98.88±0.040**	94.56±0.008	93.42±0.064

		Cowhide data set—Classification rate						Spambase data set—Classification rate					
$\mathcal{D}_1^l$	$\mathcal{D}_1^u$	**OPFSEMI**	*le.*	*et.*	**OPFSEMI**$_{mst}$	*le.*	*et.*	**OPFSEMI**	*le.*	*et.*	**OPFSEMI**$_{mst}$	*le.*	*et.*
1	99	84.54±0.076	24.84	0.282	**84.54±0.069**	24.84	0.093	64.78±0.053	35.07	1.890	**65.22±0.061**	33.49	0.607
10	90	91.69±0.040	12.01	0.286	**92.44±0.015**	10.41	0.094	65.90±0.027	32.22	1.899	66.12±0.056	30.11	0.605
20	80	93.75±0.008	9.50	0.296	**94.79±0.053**	8.76	0.096	67.25±0.092	30.73	1.950	68.69±0.023	27.32	0.619
30	70	94.40±0.076	9.52	0.296	**95.82±0.053**	6.75	0.094	68.46±0.062	30.42	1.998	71.53±0.021	25.43	0.621
40	60	94.76±0.092	8.59	0.311	**96.16±0.016**	5.49	0.096	69.55±0.058	29.02	2.071	73.42±0.061	24.13	0.690
50	50	95.80±0.107	4.56	0.296	**96.68±0.069**	3.89	0.090	70.81±0.067	27.74	2.133	74.18±0.037	19.25	0.693
		Statlog data set—Classification rate						Pendigits data set—Classification rate					
$\mathcal{D}_1^l$	$\mathcal{D}_1^u$	**OPFSEMI**	*le.*	*et.*	**OPFSEMI**$_{mst}$	*le.*	*et.*	**OPFSEMI**	*le.*	*et.*	**OPFSEMI**$_{mst}$	*le.*	*et.*
1	99	85.56±0.013	29.10	0.294	**85.72±0.053**	29.10	0.093	95.66±0.015	7.58	6.957	**95.78±0.084**	7.54	2.141
10	90	91.73±0.107	25.01	0.299	**91.76±0.044**	18.78	0.094	98.27±0.053	5.89	7.168	**99.22±0.031**	1.05	2.212
20	80	91.94±0.099	24.83	0.293	**93.11±0.099**	16.80	0.091	99.10±0.076	1.33	7.084	**99.30±0.069**	0.95	2.162
30	70	91.91±0.046	21.76	0.312	**93.85±0.114**	15.38	0.095	99.28±0.046	1.07	7.214	**99.44±0.015**	0.85	2.155
40	60	93.15±0.081	17.81	0.310	**94.47±0.031**	15.09	0.095	98.56±0.033	2.62	7.328	**99.44±0.015**	0.73	2.151
50	50	93.96±0.014	16.71	0.316	**95.21±0.015**	14.11	0.091	98.61±0.053	2.67	7.469	**99.51±0.071**	0.57	2.142
		Faces data set—Classification rate						Parasites data set—Classification rate					
$\mathcal{D}_1^l$	$\mathcal{D}_1^u$	**OPFSEMI**	*le.*	*et.*	**OPFSEMI**$_{mst}$	*le.*	*et.*	**OPFSEMI**	*le.*	*et.*	**OPFSEMI**$_{mst}$	*le.*	*et.*
1	99	88.62±0.046	23.21	0.556	**89.15±0.053**	23.21	0.181	91.94±0.019	13.03	0.902	**92.01±0.053**	13.03	0.301
10	90	91.75±0.084	15.22	0.552	**93.35±0.031**	13.86	0.180	97.85±0.094	4.56	0.946	**97.94±0.038**	4.56	0.311
20	80	94.75±0.042	9.57	0.558	**96.12±0.015**	6.96	0.182	97.69±0.023	4.40	0.909	**97.82±0.046**	3.76	0.298
30	70	97.02±0.046	4.26	0.562	**98.14±0.038**	2.40	0.180	98.36±0.023	3.14	0.949	**98.43±0.094**	3.14	0.306
40	60	97.79±0.061	3.16	0.581	**98.38±0.137**	2.15	0.182	98.43±0.069	3.31	0.893	**98.45±0.094**	3.31	0.280
50	50	97.61±0.015	3.11	0.617	**98.61±0.053**	1.67	0.187	98.79±0.015	2.64	0.916	98.85±0.084	1.90	0.281

The authors verified that OPFSEMI$_{mst}$ approach was superior to OPFSEMI, SVMs, TSVMs and SemiL in the majority of cases. The SVMs has showed itself better than its semi-supervised counterpart TSVMs in most experiments, which is different from the results of OPFSEMI$_{mst}$ and OPF, where the semi-supervised version shows a gain in accuracy whenever it has been used. Both TSVMs and SemiL had difficulty in converging to a good classification rate based on a complex evaluation data, namely, the Cowhide and Parasites data sets. OPFSEMI$_{mst}$ shows itself

to be faster in its search for better accuracy even in situations with a few labeled samples. When the number of labeled samples was increased, there was an increase in the accuracy of TSVMs and SemiL, but still not better than OPFSEMI$_{mst}$. In most cases, even with a smaller amount of labeled samples, OPFSEMI$_{mst}$ manages to be superior than TSVMs and SemiL with a larger amount of labeled samples.

4. Conclusions

In this chapter, we presented a brief compilation of some recent research based on the Optimum-Path Forest classifier, as well as a literature review about OPF. The most recent works have focused on semi-supervised learning, anomaly detection and contextual classification.

We have observed a gain in popularity in the last years regarding the OPF classifier, since even more people are downloading the open-source library LibOPF and using it for several other applications. In regard to future works, we intend to combine OPF and deep learning techniques to asses its robustness over high-dimensional feature spaces.

Acknowledgments

The authors would like to thank all support given by several research foundations during the last years under the following grants: Capes, FAPESP #2009/16206-1, FAPESP #2014/16250-9, CNPq #451033/2015-9, #306166/2014-3, CNPq #470571/2013-6, CNPq #303182/2011-3 and CNPq #487032/2012-8.

References

1. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. (Wiley-Interscience, 2000). ISBN 0471056693.
2. C. Cortes and V. Vapnik, Support vector networks, *Machine Learning*. **20**, 273–297, (1995).
3. S. Haykin, *Neural Networks: A Comprehensive Foundation (3rd Edition)*. (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2007). ISBN 0131471392.
4. J. P. Papa, A. X. Falcão, and C. T. N. Suzuki, Supervised pattern classification based on optimum-path forest, *International Journal of Imaging Systems and Technology*. **19**(2), 120–131, (2009). ISSN 0899-9457.
5. J. P. Papa, A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares, Efficient supervised optimum-path forest classification for large datasets, *Pattern Recognition*. **45**(1), 512–520, (2012).
6. L. M. Rocha, F. A. M. Cappabianco, and A. X. Falcão, Data clustering as an optimum-path forest problem with applications in image analysis, *International Journal of Imaging Systems and Technology*. **19**(2), 50–68, (2009).
7. J. P. Papa and A. X. Falcão, A new variant of the optimum-path forest classifier., *International Symposium on Visual Computing*. **47**(1), 935–944, (2008).
8. K. A. Costa, L. A. Pereira, R. Y. Nakamura, C. R. Pereira, J. P. Papa, and A. X. Falcão, A nature-inspired approach to speed up optimum-path forest clustering and its

application to intrusion detection in computer networks, *Information Sciences*. **294**, 95–108, (2015). ISSN 0020-0255.

9. R. J. Pisani, R. Y. M. Nakamura, P. S. Riedel, C. R. L. Zimback, A. X. Falcão, and J. P. Papa, Toward satellite-based land cover classification through optimum-path forest, *IEEE Transactions on Geoscience and Remote Sensing*. **52**(10), 6075–6085, (2014).
10. F. A. Cappabianco, A. X. Falcão, C. L. Yasuda, and J. K. Udupa, Brain tissue mr-image segmentation via optimum-path forest clustering, *Computer Vision and Image Understanding*. **116**(10), 1047–1059, (2012). ISSN 1077-3142.
11. L. A. P. Júnior, K. A. P. Costa, and J. P. Papa. Fitting multivariate gaussian distributions with optimum-path forest clustering and its application for anomaly detection. In *12th International Conference on Applied Computing*, (2015). (accepted for publication).
12. C. R. Pereira, R. Y. M. Nakamura, K. A. P. Costa, and J. P. Papa, An optimum-path forest framework for intrusion detection in computer networks, *Engineering Applications of Artificial Intelligence*. **25**(6), 1226–1234, (2012).
13. R. Nakamura, L. Fonseca, J. Santos, R. Torres, X.-S. Yang., and J. Papa, Nature-inspired framework for hyperspectral band selection, *IEEE Transactions on Geoscience and Remote Sensing*. **52**(4), 2126–2137, (2014).
14. C. O. Ramos, A. N. Souza, J. P. Papa, and A. X. Falcão, A new approach for non-technical losses detection based on optimum-path forest, *IEEE Transactions on Power Systems*. **PP**(99), 1–9, (2010).
15. L. A. M. Pereira, R. Y. M. Nakamura, G. F. S. De Souza, D. Martins, and J. a. P. Papa, Aquatic weed automatic classification using machine learning techniques, *Computers and Electronics in Agriculture*. **87**, 56–63, (2012). ISSN 0168-1699.
16. C. Ramos, A. Souza, G. Chiachia, A. Falcão, and J. Papa, A novel algorithm for feature selection using harmony search and its application for non-technical losses detection, *Computers & Electrical Engineering*. **37**(6), 886–894, (2011).
17. D. Rodrigues, L. A. M. Pereira, R. Y. M. Nakamura, K. A. P. Costa, X.-S. Yang, A. N. Souza, and J. P. Papa, A wrapper approach for feature selection based on bat algorithm and optimum-path forest, *Expert Systems with Applications*. **41**(5), 2250–2258, (2014). ISSN 0957-4174.
18. J. Papa, A. X. Falcão, G. de Freitas, and A. Ávila, Robust pruning of training patterns for optimum-path forest classification applied to satellite-based rainfall occurrence estimation, *IEEE Geoscience and Remote Sensing Letters*. **7**(2), 396–400, (2010).
19. D. Osaku, R. Y. M. Nakamura, L. A. M. Pereira, R. J. Pisani, A. L. M. Levada, F. A. M. Cappabianco, A. X. Falcão, and J. P. Papa, Improving land cover classification through contextual-based optimum-path forest, *Information Sciences*. (2015). (accepted for publication).
20. D. R. Pereira, R. Y. M. Nakamura, Pisani, and J. P. Papa. Land-cover classification through sequential learning-based optimum-path forest. In *IEEE International Geoscience and Remote Sensing Symposium*, (2015). (accepted for publication).
21. L. A. M. Pereira, J. P. Papa, J. Almeida, R. d. S. Torres, and W. P. Amorim. A multiple labeling-based optimum-path forest for video content classification. In *Proceedings of the 2013 XXVI Conference on Graphics, Patterns and Images*, pp. 334–340, Washington, DC, USA, (2013). IEEE Computer Society.
22. G. B. Martins, L. C. S. Afonso, D. Osaku, J. Almeida, and J. P. Papa. Static video summarization through optimum-path forest clustering. In eds. E. Bayro-Corrochano and E. Hancock, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 8827, *Lecture Notes in Computer Science*, pp. 893–900. Springer

International Publishing, (2014).
23. W. P. Amorim, A. X. Falcão, and M. H. Carvalho. Semi-supervised pattern classification using optimum-path forest. In *27th SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 111–118, (2014).
24. C. Allène, J.-Y. Audibert, M. Couprie, and R. Keriven, Some links between extremum spanning forests, watersheds and min-cuts, *Image Vision Computing*. **28**(10), 1460–1471, (2010). ISSN 0262-8856.
25. D. Comaniciu, An algorithm for data-driven bandwidth selection, *IEEE Transaction on Pattern Analysis and Machine Intelligence*. **25**(2), 281–288, (2003).
26. J. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **22**(8), 888–905 (Aug, 2000).
27. W. P. Amorim, J. P. Papa, M. H. Carvalho, and A. X. Falcão, A novel semi-supervised learning approach based on optimum connectivity, *Pattern Recognition*. (2015). (submitted).
28. W. W. Cohen and V. R. Carvalho. Stacked sequential learning. In *Proceedings of the 19th international joint conference on Artificial intelligence*, pp. 671–676, San Francisco, CA, USA, (2005). Morgan Kaufmann Publishers Inc.
29. T. G. Dietterich. Machine learning for sequential data: A review. In *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, pp. 15–30, London, UK, (2002). Springer-Verlag.
30. C. Gatta, E. Puertas, and O. Pujol, Multi-scale stacked sequential learning, *Pattern Recognition*. **44**(10–11), 2414 – 2426, (2011). ISSN 0031-3203. Semi-Supervised Learning for Visual Content Analysis and Understanding.
31. F. Wilcoxon, Individual comparisons by ranking methods, *Biometrics Bulletin*. **1**(6), 80–83, (1945).
32. R. Collobert, F. Sinz, J. Weston, and L. Bottou, Large scale transductive svms, *J. Mach. Learn. Res.* **7**, 1687–1712 (Dec., 2006). ISSN 1532-4435.
33. C. Feng, A. Sutherland, R. King, S. Muggleton, and R. Henery. Comparison of machine learning classifiers to statistics and neural networks. In *Proceedings of the Third International Workshop in Artificial Intelligence and Statistics*, pp. 41–52, (1993).
34. L. Cranor and B. Lamacchia, Spam!, *Communications of the ACM*. **41**(8), 74–83 (Aug., 1998). ISSN 0001-0782.
35. Faces. biometrics database distribution. In *The Computer Vision Laboratory, University of Notre Dame*, (2011). URL `www.nd.edu/~cvrl/CVRL/Data_Sets.html`.
36. F. Alimoglu and E. Alpaydin. Methods of combining multiple classifiers based on different representations for pen-based handwriting recognition. In *Proceedings of the Fifth Turkish Artificial Intelligence and Artificial Neural Networks Symposium*, (1996).
37. W. P. Amorim, H. Pistori, M. C. Pereira, and M. A. C. Jacinto. Attributes reduction applied to leather defects classification. In *Graphics, Patterns and Images (SIBGRAPI), 2010 23rd SIBGRAPI Conference on*, pp. 353–359, (2010).
38. C. T. N. Suzuki, J. F. Gomes, A. X. Falcão, J. P. Papa, and S. H. Shimizu, Automatic segmentation and classification of human intestinal parasites from microscopy images, *IEEE Transactions on Biomedical Engineering*. **60**(3), 803–812, (2013).