Accepted version of paper published in International Conference on Pattern Recognition (ICPR), 2020.
DOI: https://doi.org/10.1109/ICPR48806.2021.9412596

Weakly Supervised Learning through Rank-based Contextual Measures

João Gabriel Camacho Presotto*, Lucas Pascotti Valem*, Nikolas Gomes de Sá*,
Daniel Carlos Guimarães Pedronette*, João Paulo Papa†
*Department of Statistics, Applied Math. and Computing, UNESP - São Paulo State University, Rio Claro - SP, Brazil
†School of Sciences, UNESP - São Paulo State University, Bauru - SP, Brazil

***Abstract*—Machine learning approaches have achieved remarkable advances over the last decades, especially in supervised learning tasks such as classification. Meanwhile, multimedia data and applications experienced an explosive growth, becoming ubiquitous in diverse domains. Due to the huge increase in multimedia data collections and the lack of labeled data in several scenarios, creating methods capable of exploiting the unlabeled data and operating under weakly supervision is imperative. In this work, we propose a rank-based model to exploit contextual information encoded in the unlabeled data in order to perform weakly supervised classification. We employ different rank-based correlation measures for identifying strong similarities relationships and expanding the labeled set in an unsupervised way. Subsequently, the extended labeled set is used by a classifier to achieve better accuracy results. The proposed weakly supervised approach was evaluated on multimedia classification tasks, considering several combinations of rank correlation measures and classifiers. An experimental evaluation was conducted on 4 public image datasets and different features. Very positive gains were achieved in comparison with various semi-supervised and supervised classifiers taken as baselines when considering the same amount of labeled data.**

***Index Terms*—classification, machine learning, weak supervision, rank correlation measure**

I. Introduction

Nowadays, the use of multimedia retrieval technologies in several domains is remarkable, ranging from medical to surveillance applications. Additionally, due to the several advances in technology and facilities for acquisition and storage, the generation of multimedia content has experienced a huge growth. Although very effective in several scenarios, retrieval and classification applications rely mainly on low-level features. However, often the direct use of low-level features is not able to properly represent data concepts in many applications, affecting the effectiveness of the results. It occurs mainly due to the semantic gap problem, which refers to the difficulty of mapping low-level features to high-level concepts [1].

In such scenarios, supervised learning approaches are commonly used for bridging the semantic gap, mainly supported by training data or user feedback. While it is indeed a robust solution, supervised learning approaches often require a vast amount of labeled data, especially the ones based on deep learning strategies. On the other hand, due to the quick and continuous growth of multimedia collections, the task of creating labeled sets has become increasingly laborious. Even for relevance-feedback approaches, in which the results are successively re-computed based on relevance judgments collected from the users, the continuous growth of collections represents a severe problem.

In fact, the large increase of data generation in recent years has created the need for more effective and efficient ways for training machine learning methods. One of the most common issues stands for the lack of labels due to the difficulties in manually labeling. Therefore, significant amounts of effort have been devoted to developing weakly supervised methods. Weakly Supervised Learning (WSL) denotes a wide range of approaches, covering a variety of studies which attempt to construct predictive models by learning with weak supervision [2], [3].

Most of these approaches are often designed to take advantage of the particular properties of weak supervision regimes, form of available data, and prior knowledge of the tasks [4]. Typically, there are three types of weak supervision [2]. The incomplete supervision, where a small subset of training data is given with labels and the remaining training data is unlabeled. The inexact supervision, where only coarse-grained labels are given. The last type is the inaccurate supervision, *i.e.*, when the provided labels are not always ground-truth.

This work focuses on incomplete supervision, where only a small amount of training data is available. Although there is no broadly accepted taxonomy in WSL [3], our approach is near to semi-supervised learning methods and analogous to active learning methods, in the sense that such methods are capable of dealing with incomplete supervision. However, different from active learning approaches, both our approach and semi-supervised methods perform without the use of an oracle [2]. Actually, semi-supervised classification methods are suitable to handle training sets with large amounts of unlabeled data and a small quantity of labeled data. In this scenario, how to exploit useful information encoded in the unlabeled data is a central challenge. One traditional approach used for this task is given by self-training [5] methods. In self-training, the classifier is initially trained on a small labeled set and subsequently retrained using its own predictions as additional labeled points [6]. Although self-training methods consist in one of the first semi-supervised learning approaches, research about effective ways to exploit the unlabeled data is still active and challenging, even in very recent works [7].

In this paper, we propose a novel rank-based model to exploit contextual information encoded in unlabeled data. In the last years, various unsupervised learning methods [8], [9] have been used to compute more effective affinity measures by exploiting the intrinsic manifold structure of multimedia datasets. More recently, rank-based approaches have been successfully proposed with the same goal [10], [11]. Rank statistics were also recently exploited for addressing the complex task of discovering new classes based on incomplete supervision [12]. In this scenario of crescent use of ranking information, we propose a weakly supervised approach capable of exploiting unlabeled data through a rank-based

model for classification tasks. The main idea consists in analyzing the manifold structure of the original training set in order to expand the labels to the unlabeled data. In our approach, rank correlation measures [13] are exploited to identify strong similarity relationships between images. Using this information, it is possible to assign classes to the unlabeled data, expanding the labeled training set so that it can be used by a supervised or semi-supervised classifier.

A broad experimental evaluation was conducted in order to assess the effectiveness of the proposed method. The experiments were performed on four public datasets considering different features. An additional contribution of the work regarding very recent rank-based approaches [12] consists in an experimental analysis of different rank correlation measures. The conducted experiments considered six rank measures evaluated in conjunction with various classifiers. Significant accuracy gains up to +60.72% were achieved, reaching superior results to supervised and semi-supervised baselines in most of scenarios.

The remaining of this paper is organized as follows. Section II presents our weakly supervised rank-based learning approach, and Section III presents the correlation measures considered. Section IV describes the experimental evaluation and Section V discusses conclusions and future work.

II. RANK-BASED WEAKLY SUPERVISED LEARNING

In most machine learning, computer vision, and retrieval applications, data objects are commonly represented as points in a high-dimensional space. In such scenarios, measuring the similarity/dissimilarity between data points remains a relevant and challenging research question, even for deep representations. Traditional distance measures, as the Euclidean distance, consider only pairs of objects. On the other hand, a ranking provides an inherent contextual representation which establish a relationship among all elements in each rank. Therefore, the main hypothesis of this work can be highlighted as:

- The contextual information encoded in ranked lists can be analyzed through rank correlation measures to identify strong similarity relationships between images;
- Once identified, strong similarity relationships can be used to expand small training sets.

Figure 1 illustrates a binary classification scenario in a two dimensional space. The unlabeled data is presented in grey borders, while labeled data is illustrated in blue or red borders. The green edges represent strong similarity relationships identified by unsupervised contextual analysis performed through rank correlation measures. The proposed model is formally defined in next sub-sections.

A. Formal Problem Definition

Let $\mathfrak{X} = \{x_1, x_2, \ldots, x_L, x_{L+1}, \ldots, x_N\}$ be a data collection, where each element x_i denotes a data item. The collection $\mathfrak{X}$ can be defined as a partially labeled dataset, where $\mathfrak{X}_L = \{x_i\}_{i=1}^{L}$ is the labeled data items subset and $\mathfrak{X}_U = \{x_i\}_{i=L+1}^{N}$ is the unlabeled data items subset. Usually, the volume of labeled data is much smaller than unlabeled data, i.e., $|\mathfrak{X}_L| \ll |\mathfrak{X}_U|$.

In general, the main objective of supervised learning is to assign labels to unlabeled data based on a model generated from the labeled data. Formally, let $\mathfrak{L} = \{1, \ldots, C\}$ be a set which contains the labels of the dataset. Let $y : \mathfrak{X} \rightarrow \mathfrak{L}$ be a set which associates each $x_i \in \mathfrak{X}$ to its label $y(x_i)$ in the classification results. Therefore, a supervised learning task can be defined as the estimation of a function $y(x_i)$ for each unlabeled data item $x_i \in \mathfrak{X}_U$ based on the learned model.

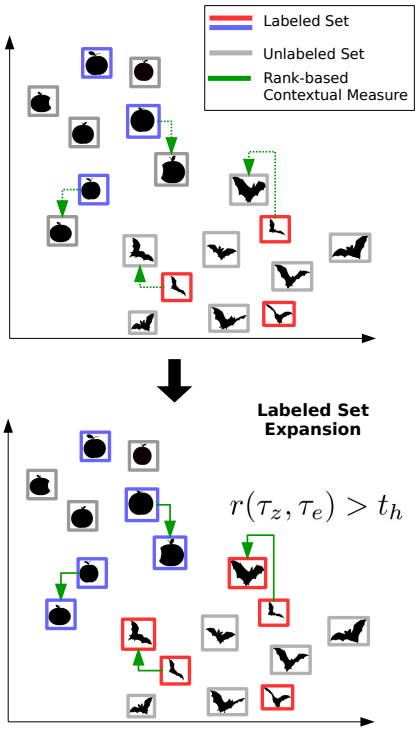


Fig. 1: Weakly supervised learning based on contextual rank measures, represented by green lines and function $r(\tau_z, \tau_e)$.

The main objective of this work is to identify a sub-set of the unlabeled data which can be aggregated to a weakly labeled set through a rank-based estimation of the labeling function y. Formally, let $\mathfrak{X}_E \subset \mathfrak{X}_U$ be an estimated label set. A weakly labeled set $\mathfrak{X}_W$ is defined as $\mathfrak{X}_W = \mathfrak{X}_L \cup \mathfrak{X}_E$. For each $x_e \in \mathfrak{X}_E$, a rank-based labeling function $e(x_i)$ is used to estimate $y(x_e)$. Subsequently, the expanded labeled set $\mathfrak{X}_W$ is used for training the supervised learning algorithm. The central question, discussed in next sub-section, is how to define the sub-set $\mathfrak{X}_E$ and the labeling estimation function $e(x_e)$.

B. Labeled Set Expansion

The labeled set expansion is defined on ranking information obtained from the data collection $\mathfrak{X}$. Let v_{x_i} be a feature vector defined in $\mathbb{R}^n$, which represents the data item $x_i \in \mathfrak{X}$. Let d: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a distance function, which computes the distance between two data items according to their corresponding feature vectors. Formally, the distance between two data items x_i and x_j is given by $d(v_{x_i}, v_{x_j})$.

Based on the distance function d, a ranked list τ_q can be computed in order to obtain the most similar data items of a given data item x_q. The ranked list τ_q=(x_1, x_2, …, x_M) can be formally defined as a permutation of the collection $\mathfrak{X}$. For a permutation τ_q, we interpret $\tau_q(i)$ as the position (or rank) of the data item x_i in the ranked list τ_q. We can say that, if x_i is ranked before x_j in the ranked list of x_q, that is,

$\tau_q(i) < \tau_q(j)$, then $d(v_q, v_i) \leq d(v_q, v_j)$. A ranked list τ_i can be computed for every data item $x_i \in \mathfrak{X}$ in order to obtain a set $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_M\}$ of ranked lists.

A rank correlation measure can be defined as a function r: $\mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$ which compares two ranked lists by assigning a real number to them. To highly correlated ranked lists higher values of correlation are expected to be assigned. The analyzed hypothesis is that a high rank correlation indicates a strong similarity measure, which can be used to expand the labeled set. Let $x_l \in \mathfrak{X}_L$ be a labeled item and let $x_u \in \mathfrak{X}_U$ be an unlabeled item. Therefore, if the rank correlation between τ_l and τ_u overcomes a given threshold t_h only for labeled items of the same class, the labeled set is expanded to incorporate x_u. More formally, we can define the expanded labeled set $\mathfrak{X}_E$ as:

$$\begin{aligned}\mathfrak{X}_E = \{x_e : x_e \in \mathfrak{X}_U, x_l \in \mathfrak{X}_L \wedge r(\tau_l, \tau_e) > t_h \wedge \\ \nexists x_z \in \mathfrak{X}_L | r(\tau_z, \tau_e) > t_h \wedge y(x_l) \neq y(x_z)\}.\end{aligned} \quad (1)$$

The estimated class of the element x_e is the same of the labeled items x_l, therefore:

$$e(x_e) = y(x_l). \quad (2)$$

The process of labeled set expansion is also illustrated in Figure 1. Labeled items are represented in different colors for each class (blue and red) and the rank correlation measure $r(\tau_z, \tau_e)$ as a green line. Finally, the expanded labeled set is provided to a supervised or semi-supervised classifier.

The computational complexity of label expansion is closely associated to the amount of labeled data. For each labeled item (L), a rank measure is computed for each unlabeled item ($N - L$). The complexity of rank measures varies, but is constrained to a constant k, and therefore $O(1)$. The total complexity is $O(L(N-L))$, but can be reduced to $O(L)$ by computing the rank measure only in top rank positions of each labeled item.

III. Contextual Rank Correlation Measures

Once defined the process of labeled set expansion, a fundamental issue consists in how to compute the rank correlation measure. In this section, we discusses different possibilities for defining the rank correlation measure $r(\cdot, \cdot)$. Distinct approaches were considered by the measures, such as the intersection at top-k positions (RBO and Intersection Measure), difference between rank positions (Kendall τ), and set operations (Jaccard and Jaccard$_k$). The notation $\mathcal{N}(q, k)$ used along the section refers to the neighborhood set of size k, which contains the k-nearest neighbors of given data item x_q.

A. Intersection Measure

Intersection measure [14] computes the amount of overlap between two top-k lists τ_i and τ_j taking in consideration the overlap in all depths, from 1 to k. That is, for each depth $d \in \{1 \ldots k\}$, the overlap at d is calculated, then each one of those overlaps are averaged in order to compute a similarity measure. To the first positions of the top k lists, Intersection Measure gives a higher weight, since them are considered several times. Next, in Equation 3, a formal definition of Intersection measure r_i is presented:

$$r_i(\tau_i, \tau_j) = \frac{\sum_{d=1}^{k} \mid \mathcal{N}(i, d) \cap \mathcal{N}(j, d) \mid}{k}. \quad (3)$$

B. Jaccard

The Jaccard coefficient [15], given two non-empty sets of depth k, computes the probability that an element of at least one of the two sets is an element of both. Equation 4 defines the Jaccard coefficient:

$$r_j(\tau_i, \tau_j) = \frac{|\mathcal{N}(i, k) \cap \mathcal{N}(j, k)|}{|\mathcal{N}(i, k) \cup \mathcal{N}(j, k)|}. \quad (4)$$

C. Jaccard$_k$

Since the traditional Jaccard coefficient considers only a depth k in a ranked list in its computation, it ignores the information provided by the first positions of the rank with a depth smaller than k. In this scenario, Jaccard$_k$ [16], an accumulated Jaccard score, was proposed. The measure considers depths from 1 to k assigning higher weights to the top positions. The measure is defined as follows:

$$r_k(\tau_i, \tau_j) = \frac{\sum_{d=1}^{k} r_j(\tau_i, \tau_j, k)}{k}. \quad (5)$$

D. Kendall τ

A traditional distance measure between permutations, Kendall's τ, is computed based on the number of exchanges needed in a bubble sort to convert one permutation to the other [14]. Kendall's τ measure is defined as follows:

$$r_\tau(\tau_i, \tau_j) = \frac{(k \times (k-1))}{2 \times \sum_{x,y \in \mathcal{N}(i,k) \cup \mathcal{N}(j,k)} \bar{K}_{x,y}(\tau_i, \tau_j)}, \quad (6)$$

where $\bar{K}_{x,y}(\tau_i, \tau_j)$ is a function that determines if two images img_x and img_y are in the same order in top-k ranked lists τ_i and τ_j. $\bar{K}_{x,y}(\tau_i, \tau_j)$ can be defined as follows:

$$\bar{K}_{x,y}(\tau_i, \tau_j) = \begin{cases} 0 & \text{if } (\tau_i(x) \leqslant \tau_i(y) \wedge \tau_j(x) \leqslant \tau_j(y)), \\ 0 & \text{if } (\tau_i(x) \geqslant \tau_i(y) \wedge \tau_j(x) \geqslant \tau_j(y)), \\ 1 & \text{otherwise.} \end{cases}$$

E. Rank-Biased Overlap (RBO)

Rank-Biased Overlap (RBO) [17], similar to Intersection measure, also considers the overlap of two ranked lists at increasing depths. The difference between the two measures is that RBO uses a parameter that determines the probability of considering the overlap at the next depth. Then, the measure is computed based on those probabilities. RBO measure is defined in Equation 7:

$$r_r(\tau_i, \tau_j) = (1-p) \sum_{d=1}^{k} p^{d-1} \frac{|\mathcal{N}(i, k) \cap \mathcal{N}(j, k)|}{d}, \quad (7)$$

where $p \in [0, 1]$ is used to compute the probability at depth d.

F. Spearman

Spearman's metric is a non-parametric measure that evaluates the relationship between two variables. Frequently denoted by the letter ρ, the metric can be seen as the L1 distance between two permutations [14]. The measure is defined formally as follows:

$$r_\rho(\tau_i, \tau_j) = \frac{k \times (k+1)}{\sum_{x,y \in \mathcal{N}(i,k) \cup \mathcal{N}(j,k)} |\tau_i(j) - \tau_j(i))|}, \quad (8)$$

where k is the depth of the ranked lists being analyzed.

IV. EXPERIMENTAL EVALUATION

This section presents the experimental evaluation conducted for assessing the accuracy of the proposed approach. Section IV-A presents the datasets and features considered. Section IV-B presents an analysis of rank correlation measures and thresholds. Finally, Section IV-C discusses the experiments conducted using the proposed weakly supervised approach in classification tasks.

A. Datasets and Features

We considered four different public image datasets with size ranging from 1,360 to 70,000 images, considering different features. The datasets are described as follows:

- **MPEG-7** [18]: there are 1,400 shape images divided into 70 classes. This dataset is commonly used on evaluation of unsupervised post-processing methods for shape retrieval. Two features were used: ASC - Aspect Shape Context [19] and CFD - Contour Features Descriptor [20][1].
- **Flowers** [21]: different images of 17 flower species with 80 images of each presenting pose and light variations. This dataset is distributed by the University of Oxford. Color and deep features were considered: ACC - Auto Color Correlogram [22] and CNN-Resnet [23].
- **Corel5k** [24]: the dataset includes diverse scene content such as fireworks, bark, microscopy images, tiles, trees, and others. It is composed of 50 categories with 100 images for each class. The same features used for Flowers dataset was considered: ACC [22] and CNN-Resnet [23].
- **MNIST** [25]: handwritten digit database composed by 60,000 training and 10,000 test images of ten different classes of numbers from 0 to 9. For this dataset, we used CNN-Resnet [23] to extract features.

B. Labeled Set Expansion Evaluation

This section presents an experimental analysis considering different rank correlation measures and its respective values of threshold to obtain an adequate labeled set expansion. High values of threshold leads to small or insignificant expansions. On the other hand, as its values decreases, the number of images contained in the expanded set increases. However, it tends to incorporate incorrect images in this set as well, which can be especially harmful to the accuracy results. A visual analysis of this scenario is showed in Figure 2. A ranked list is illustrated considering incremental values of thresholds. As we can observe, the first top lines provide small set expansions, while last lines, which consider higher thresholds, could include incorrect images.

1) Effectiveness Measures: In fact, the trade-off between how precise and how comprehensive are the results is well modeled by the relationship between Precision and Recall measures. While precision evaluates the number of items labeled correctly, recall analyzes the images for which the labels have not been expanded, but could have been.

Once Precision and Recall measures encode distinct, although relevant information, we consider both measures. In order to combine both information in a single numeric measure, we use F-Measure and F-Beta (F_β). F-Measure combines precision and recall through the harmonic mean, while F-Beta allows to adjust the weight assigned to precision or recall. For values of β smaller than 1, more relevance is given to precision, while values greater than 1 assigns more relevance to recall. For all the experiments we considered $\beta = 0.15$, giving more importance to precision, with the objective of being more strict to false positives and allow possible false negatives.

[1]For this dataset, we have used the distances to other images as features.

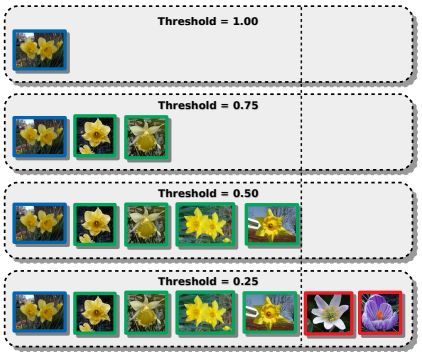


Fig. 2: Visual results for a correlation measure with different thresholds on Flowers dataset.

2) Labeled Set Expansion Results: The proposed training set expansion relies on the idea that the correlation measures applied to the ranked lists are able to find new elements with high confidence to be used as part of the training set. Firstly, with the objective of analyzing the behavior of each measure, an experiment was conducted for all the considered datasets, where the threshold t_h is varied to evaluate the impact in the effectiveness measures. Figure 3 reports the results for the Corel5K dataset. Analogous experiments were conducted for MPEG-7 and Oxford Flowers datasets.

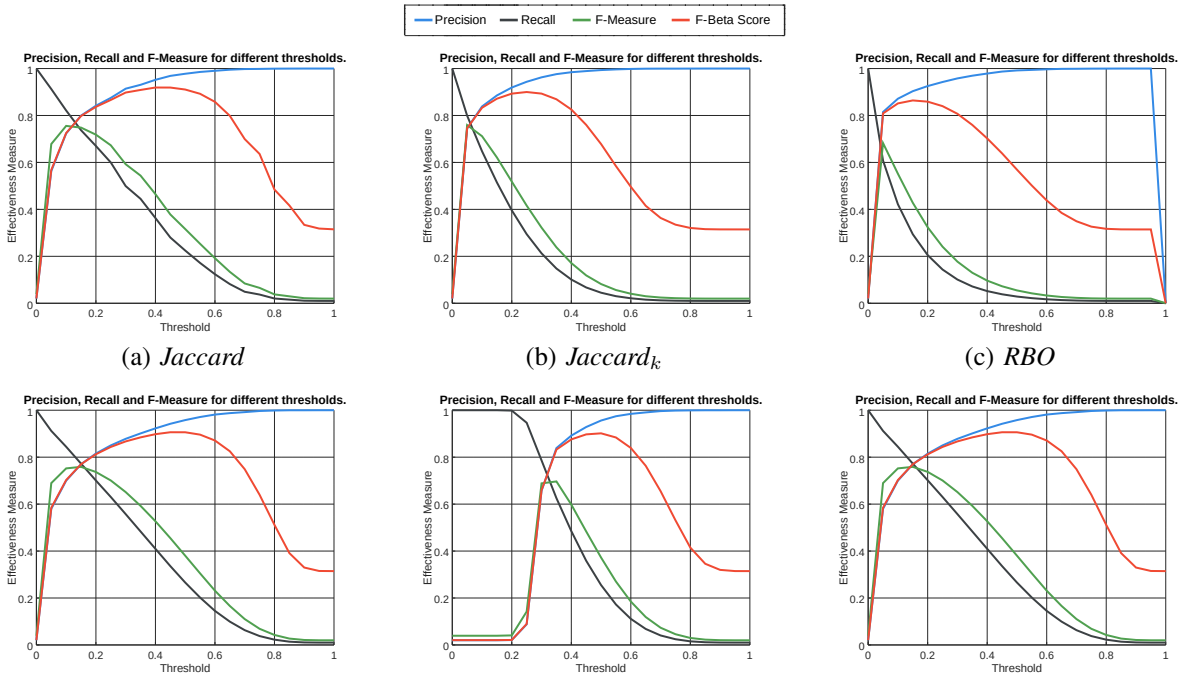
For the sake of visualization, Table I presents the threshold that provided the best value of F-Beta for each dataset, descriptor, and correlation measure. The threshold value obtained for the whole dataset is very correlated to the value obtained for each fold (the current training set). An experiment was conducted in order to compare the such values. Figure 4 illustrates the comparison, considering the RBO rank correlation measure on different datasets. This approach can be exploited for automatically defining the threshold value in real-world situations.

TABLE I: Maximum value of F-Beta and its respective threshold for each correlation measure.

		MPEG-7		Flowers		Corel5k	
		ASC	CFD	ACC	Resnet	ACC	Resnet
Intersection	**F-Beta**	0.963	0.969	0.483	0.858	0.68	0.906
	Threshold	0.7	0.6	0.15	0.15	0.5	0.45
Jaccard	**F-Beta**	0.977	0.969	0.506	0.88	0.71	0.919
	Threshold	0.75	0.6	0.4	0.45	0.4	0.4
Jaccard$_k$	**F-Beta**	0.961	0.966	0.501	0.86	0.66	0.9
	Threshold	0.4	0.35	0.35	0.3	0.3	0.25
Kendalτ	**F-Beta**	0.953	0.961	0.504	0.865	0.668	0.902
	Threshold	0.65	0.55	0.5	0.55	0.5	0.5
RBO	**F-Beta**	0.953	0.955	0.462	0.804	0.561	0.864
	Threshold	0.35	0.3	0.4	0.2	0.3	0.15
Spearman	**F-Beta**	0.963	0.969	0.51	0.872	0.68	0.906
	Threshold	0.7	0.6	0.5	0.55	0.5	0.45

C. Weakly-Supervised Classification

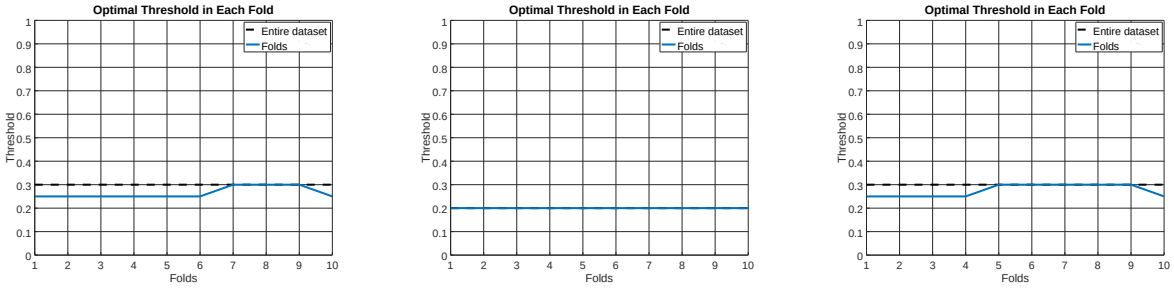
This section presents an evaluation of the proposed weakly supervised learning approach. Different classifiers were considered in the experiments which utilized the thresholds pre-



(a) *Jaccard* (b) *Jaccard*$_k$ (c) *RBO*

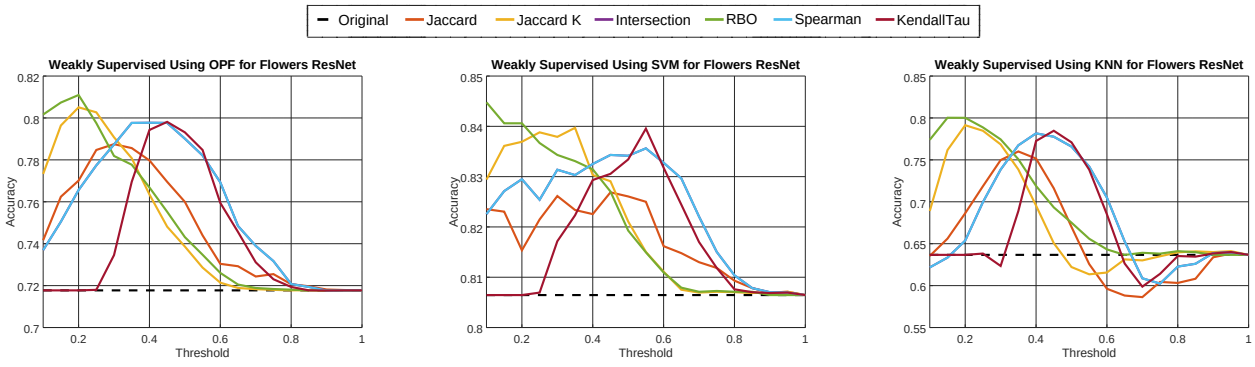
(d) *Spearman* (e) *Kendall* τ (f) *Intersection*

Fig. 3: Precision, Recall, and F-Measure for different rank correlation measures and thresholds on Corel5k with Resnet feature.



(a) *MPEG-7 with CFD* (b) *Flowers with ResNet* (c) *Corel5K with ACC*

Fig. 4: Comparison between threshold values estimated to whole dataset and for each fold.



(a) *OPF-WS* (b) *SVM-WS* (c) *kNN-WS*

Fig. 5: Accuracy using OPF, SVM, and kNN considering training set expansion with different rank correlation measures.

sented in Table I to expand the training set. For MNIST [25] dataset, only the RBO measure was considered and the threshold used is $t_h = 0.05$. Sections IV-C1 and IV-C2 presents an overview of the supervised and semi-supervised methods used in our experiments, and Section IV-C3 presents the results obtained by the proposed weakly supervised method.

1) Supervised Methods: Aiming at evaluating the proposed approach in different scenarios, the method was tested consid-

ering different methods, supervised and semi-supervised. The supervised classifiers are briefly described in the following.

- **Optimum Path Forest:** The Optimum Path Forest (OPF) [26] builds a graph where each element is represented by a node and the edges are weighted by their distances. It aims at finding the optimum path between the nodes for classifying them into a given class/category.
- **Support Vector Machines:** The Support Vector Machines (SVM) [27] aim at finding the best surface that properly separates the data into the correct subsets according to a ground truth. That is, the decision boundary created by the SVM is placed in a position where the majority of samples from each class are separated from each other. The model built by the SVM is a non-probabilistic linear classifier, in other words, it assigns new samples to a single class[2].
- **kNN:** We also considered the traditional k-Nearest Neighbors (kNN) classifier, which assign labels to the elements according to the nearest k elements. The value of $k = 20$ was used in the experiments.

2) Semi-Supervised Methods: In addition to the utilized supervised methods, three semi-supervised methods were used in evaluation: LDS-GNN [29], Label Spreading [30], and Pseudo- Label [31]. The first was used similarly to the supervised methods, evaluating its performance with and without the training set expansion. The latter was used as a baseline to our proposed method.

- **LDS-GNN:** Learning Discrete Structures for Graph Neural Networks (LDS-GNN) [29] is a recent semi-supervised classifier that learns a discrete and sparse dependency structure between data points while simultaneously training the parameters of a GCN. For all the experiments we considered the implementation provided by the authors and the default parameters of the method.
- **Label Spreading:** Label Spreading [30] is a semi-supervised algorithm similar to Label Propagation [32] that uses an affinity matrix based on the normalized graph Laplacian instead of a non-normalized graph Laplacian. The method propagate labels from labeled nodes to their neighbors considering their proximity[3]. We used this method as a baseline to the proposed method.
- **Pseudo-Label:** Pseudo-Label [31] is a semi-supervised learning method originally used for deep neural networks. The central idea behind the technique is to use the model probabilities to assign labels to unlabeled data. The model is then re-trained using both labeled data and the pseudo-labeled data. Although this method was originally proposed for deep neural networks, it can also be used with several classifiers. In our experiments, we used a Pseudo-Label implementation publicly available[4] and the Logistic Regression with Stochastic Gradient Descent (SGD) training as the classifier[5].

[2]We used the scikit-learn [28] implementation of SVM, we also considered a polynomial kernel with degree 2, $\gamma = 0.001$ and the cost of misclassification (C) equals to 10.

[3]We also used the scikit-learn [28] implementation of Label Spreading, considering a RBF kernel with $\alpha = 0.4125$, $\gamma = 0.1$ and $max\_iter = 100$.

[4]https://github.com/anirudhshenoy/pseudo_labeling_small_datasets

[5]We used sklearn.linear_model.SGDClassifier with loss="log", and $\alpha = 0.00001$

3) Classifications Results: This section presents the results of the the proposed weakly supervised approach in classification tasks. For evaluation purposes, we split each dataset in 10 folds. On each evaluation, we consider the proportion of 10%/90% for training/test sets. The reported results are the mean accuracy obtained between all 10 folds.

Firstly, an experiment was conducted to evaluate the impact of the rank measure and the threshold value on classification accuracy. Figure 5 presents the classification results for different correlation measure, classifiers and thresholds considering the Flowers dataset the CNN-Resnet feature. The original classification accuracy without the proposed weakly supervised approach is reported by the black dashed line. These results reveal two interesting things: *(i)* the majority of the weakly supervised results are superior to the original approach, presenting significant gains and robustness to threshold variations; *(ii)* the thresholds that provide the best accuracy results are very close to what has been reported in the analysis conducted and reported in Table I.

In the following, we evaluated the different combinations of classifiers and rank measures. For each classifier, we also report the accuracy of the classifier in isolation and the results obtained by our method. Table II presents the results for each dataset, descriptor, and similarity measure considering the OPF classifier. The results were reported for the threshold that achieved the highest F-Beta in each case. Notice that we achieved positive gains in most cases, evidencing the accuracy of our approach in different scenarios. The Jaccard$_k$, Kendallτ, and RBO measures provided the best results in the majority of cases. The Intersection showed results with loss for the Flowers dataset and seems to be the less effective one.

Similarly, Table III present the results considering the SVM classifier. As we can observe, our approach also achieved gains for these classifiers, evincing the robustness of our method. However, it showed some difficulty in providing gains for the ACC features, probably due to the fact that it has a lower accuracy, which can cause more false positives in the labeled set expansion.

Table IV shows the results obtained considering the kNN classifier. We can observe that the classifier does not perform well on the non-weakly supervised scenario in comparison with OPF and SVM, however, expanding the training set achieved fairly positive gains in all cases.

Table V presents the results obtained by the LDS-GNN semi-supervised classifier. Despite the low accuracy obtained in MPEG-7 [18] dataset, the classifier achieved fairly positive results in Flowers [21] and Corel5k [24] datasets. The poor accuracy obtained in MPEG-7 [18] dataset can be explained by the usage of distances to other images as features.

Finally, we evaluated the results achieved by the proposed Weakly Supervised approach in comparison with supervised and semi-supervised classifiers. Table VI presents the results. Label Spreading [30] accuracy is computed over the amount of correct label predictions it does considering 10% of the dataset as labeled. We can observe that the proposed method achieved the best accuracy results for most of the datasets and features, except for Corel5k [24] with ACC [22] and MNIST [25] dataset. The best average accuracy considering all datasets and features were achieved by our proposed method and the SVM classifier.

TABLE II: Accuracy for each dataset and measure before and after our weakly supervised approach using OPF [26].

		MPEG-7		Flowers		Corel5k		Mean
		ASC	CFD	ACC	Resnet	ACC	Resnet	
OPF [26]		82.95%	67.75%	30.54%	71.77%	40.21%	83.56%	62.80%
Intersection	**WS-OPF**	85.56%	81.28%	30.69%	75.05%	41.69%	89.11%	67.23%
	Gain	+2.6%	+13.52%	+0.16%	+3.28%	+1.48%	+5.55%	+4.43%
Jaccard	**WS-OPF**	84.45%	77.56%	31.2%	76.95%	41.15%	88.44%	66.63%
	Gain	+1.5%	+9.81%	+0.66%	+5.18%	+0.94%	+4.88%	+3.83%
Jaccard$_k$	**WS-OPF**	**86.74%**	81.63%	31.97%	79.08%	**41.92%**	89.19%	68.42%
	Gain	**+3.79%**	+13.88%	+1.43%	+7.3%	**+1.71%**	+5.64%	+5.63%
Kendallτ	**WS-OPF**	85.67%	**82.63%**	**32.12%**	78.5%	41.77%	88.84%	68.26%
	Gain	+2.71%	**+14.88%**	**+1.58%**	+6.72%	+1.56%	+5.29%	+5.46%
RBO	**WS-OPF**	**86.75%**	82.2%	30.62%	**81.09%**	41.5%	**89.42%**	**68.60%**
	Gain	**+3.79%**	+14.44%	+0.08%	**+9.32%**	+1.29%	**+5.87%**	**+5.80%**
Spearman	**WS-OPF**	85.56%	81.28%	31.91%	78.21%	41.69%	89.11%	67.96%
	Gain	+2.6%	+13.52%	+1.37%	+6.44%	+1.48%	+5.55%	+5.16%

TABLE III: Accuracy for each dataset and measure before and after our weakly supervised approach using SVM [27].

		MPEG-7		Flowers		Corel5k		Mean
		ASC	CFD	ACC	Resnet	ACC	Resnet	
SVM [27]		83.12%	68.56%	37.5%	80.65%	45.27%	88.33%	67.24%
Intersection	**WS-SVM**	86.24%	83.44%	**37.75%**	82.71%	45.3%	91.18%	71.10%
	Gain	+3.12%	+14.89%	**+0.25%**	+2.07%	+0.03%	+2.85%	+3.87%
Jaccard	**WS-SVM**	85.16%	79.61%	35.88%	82.69%	44.94%	91.15%	69.91%
	Gain	+2.04%	+11.06%	-1.62%	+2.04%	-0.33%	+2.82%	+2.67%
Jaccard$_k$	**WS-SVM**	**87.15%**	83.62%	37.64%	83.79%	**45.6%**	**91.22%**	**71.50%**
	Gain	**+4.03%**	+15.06%	+0.14%	+3.15%	**+0.33%**	**+2.89%**	**+4.27%**
Kendallτ	**WS-SVM**	86.39%	**84.44%**	36.36%	83.98%	45.07%	91.23%	71.25%
	Gain	+3.27%	**+15.88%**	-1.14%	+3.33%	-0.2%	+2.9%	+4.01%
RBO	**WS-SVM**	86.94%	83.95%	36.66%	**84.06%**	44.89%	90.86%	71.23%
	Gain	+3.83%	+15.4%	-0.84%	**+3.42%**	-0.38%	+2.54%	+4.00%
Spearman	**WS-SVM**	86.24%	83.44%	36.48%	83.56%	45.3%	91.18%	71.03%
	Gain	+3.12%	+14.89%	-1.02%	+2.92%	+0.03%	+2.85%	+3.80%

TABLE IV: Accuracy for each dataset and measure before and after our weakly supervised approach using kNN.

		MPEG-7		Flowers		Corel5k		Mean
		ASC	CFD	ACC	Resnet	ACC	Resnet	
kNN		13.92%	12.39%	28.47%	63.67%	34.05%	76.8%	38.22%
Intersection	**WS-kNN**	71.93%	65.31%	28.88%	63.34%	36.97%	87.51%	58.99%
	Gain	+58.01%	+52.92%	+0.41%	-0.33%	+2.92%	+10.71%	+20.77%
Jaccard	**WS-kNN**	61.6%	58.6%	30.49%	71.61%	35.06%	85.77%	57.19%
	Gain	+47.68%	+46.21%	+2.02%	+7.94%	+1.01%	+8.97%	+18.97%
Jaccard$_k$	**WS-kNN**	**74.64%**	65.94%	32.6%	76.85%	37.41%	87.72%	62,53%
	Gain	**+60.72%**	+53.56%	+4.13%	+13.18%	+3.36%	+10.92%	+24.31%
Kendallτ	**WS-kNN**	69.67%	**66.67%**	32.82%	73.85%	37.66%	86.79%	61,24%
	Gain	+55.75%	**+54.28%**	+4.35%	+10.18%	+3.61%	+9.98%	+23.03%
RBO	**WS-kNN**	72.02%	64.93%	32.4%	**80.02%**	**40.04%**	**89.01%**	**63,07%**
	Gain	+58.1%	+52.54%	+3.93%	**+16.35%**	**+5.99%**	**+12.21%**	**+24.85%**
Spearman	**WS-kNN**	71.93%	65.31%	**32.98%**	74.2%	36.97%	87.51%	61,48%
	Gain	+58.01%	+52.92%	**+4.51%**	+10.53%	+2.92%	+10.71%	+23.27%

TABLE V: Accuracy for each dataset and measure before and after our weakly supervised approach using LDS-GNN [29].

		MPEG-7		Flowers		Corel5k		Mean
		ASC	CFD	ACC	Resnet	ACC	Resnet	
LDS-GNN [29]		2.55%	5.14%	28.69%	55.69%	24.66%	60.01%	29.46%
Intersection	**WS-LDS**	5.25%	17.48%	27.67%	50.06%	39.42%	85.43%	37.55%
	Gain	+2.7%	+12.34%	-1.02%	-5.63%	+14.76%	+25.42%	+8.10%
Jaccard	**WS-LDS**	**5.1%**	17.75%	42.17%	73.9%	35.56%	81.36%	42.64%
	Gain	**+2.55%**	+12.61%	+13.48%	+18.21%	+10.9%	+21.35%	+13.18%
Jaccard$_k$	**WS-LDS**	5.09%	17.02%	40.3%	79.93%	40.13%	85.82%	44.72%
	Gain	+2.54%	+11.88%	+11.61%	+24.24%	+15.47%	+25.81%	+15.26%
Kendallτ	**WS-LDS**	4.86%	16.74%	**46.03%**	74.75%	40.97%	83.09%	44.41%
	Gain	+2.31%	+11.6%	**+17.34%**	+19.06%	+16.31%	+23.08%	+14.95%
RBO	**WS-LDS**	4.8%	17.01%	41.96%	**85.86%**	**46.32%**	**88.8%**	**47.46%**
	Gain	+2.25%	+11.87%	+13.27%	**+30.17%**	**+21.66%**	**+28.79%**	**+18.00%**
Spearman	**WS-LDS**	4.91%	**17.81%**	41.0%	76.39%	38.88%	85.77%	44.13%
	Gain	+2.36%	**+12.67%**	+12.31%	+20.7%	+14.22%	+25.76%	+14.67%

TABLE VI: Weakly supervised results in comparison with supervised and semi-supervised classifiers in isolation. Weakly supervised results consider the best rank measure in Tables II to V and RBO for MNIST dataset. Label Spreading [30] and Pseudo-Label [31] are reported as additional baselines.

		MPEG-7 [18]		Flowers [21]		Corel5k [24]		MNIST [25]	Mean
		ASC [19]	CFD [20]	ACC [22]	Resnet [23]	ACC [22]	Resnet [23]	Resnet [23]	
Supervised	**kNN**	13.92%	12.39%	28.47%	63.67%	34.05%	76.8%	89.04%	45.48%
	OPF [26]	82.95%	67.75%	30.54%	71.77%	40.21%	83.56%	88.71%	66.50%
	SVM [27]	83.12%	68.56%	37.5%	80.65%	45.27%	88.33%	84.89%	69.70%%
Semi-Supervised	**Label Spreading [30]**	84.94%	71.90%	33.37%	72.65%	**46.52%**	82.32%	70.08%	65.97%
	LDS-GNN [29]	2.55%	5.14%	28.69%	55.69%	24.66%	60.01%	-	29.46%
	Pseudo-Label [31]+SGD	20.26%	19.39%	28.8%	80.89%	32.52%	87.35%	**92.21%**	51.63%
Proposed Weakly Supervised	**WS-KNN**	74.64%	66.67%	32.98%	80.02%	40.04%	89.01%	89.81%	67.60%
	WS-OPF	86.75%	82.63%	32.12%	81.09%	41.92%	89.42%	89.37%	71.9%
	WS-SVM	**87.15%**	**84.44%**	37.75%	84.06%	45.6%	**91.22%**	86.96%	**73.88%**
	WS-LDS	5.1%	17.81%	**46.03%**	**85.86%**	46.32%	88.8%	-	48.32%

V. Conclusions

In this work, we have presented a rank-based model applied to scenarios of weakly supervised learning. Our approach innovates by considering ranked list contextual information to analyze manifold information and decide which data samples can be included in an expanded labeled set. The proposed method was evaluated on four datasets, considering different features, various rank correlation measures, and classifiers. The obtained results indicated very positive accuracy gains in most of scenarios. As future work, we intend to incorporate the analysis F-beta for each fold as an automatic strategy for threshold definition. We intend to investigate the automatic choice of the rank measure and the use of other deep learning methods (CNN-Resnet and others) as final classifiers. We also intend to conduct experiments in large-scale datasets.

Acknowledgements

The authors are grateful to the São Paulo Research Foundation - FAPESP (2019/04754-6, 2020/11366-0, 2019/11104-8, 2017/25908-6, 2018/15597-6, 2014/12236-1), the Brazilian National Council for Scientific and Technological Development - CNPq (308194/2017-9, 307066/2017-7, and 427968/2018-6), and Petrobras (2017/00285-6).

References

[1] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," vol. 40, no. 1, 2007.

[2] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2018.

[3] V. Cheplygina, M. de Bruijne, and J. P. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Medical Image Analysis*, vol. 54, pp. 280 – 296, 2019.

[4] K. Li and J. Malik, "Bandit Label Inference for Weakly Supervised Learning," pp. 1–10, 2015. [Online]. Available: http://arxiv.org/abs/1509.06807

[5] I. Triguero, S. García, and F. Herrera, "Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study," *Knowl. Inf. Syst.*, vol. 42, no. 2, pp. 245–284, 2015.

[6] O. Chapelle, B. Schlkopf, and A. Zien, *Semi-Supervised Learning*, 1st ed. The MIT Press, 2010.

[7] K. Sun, Z. Lin, and Z. Zhu, "Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes," in *AAAI Conference on Artificial Intelligence, (AAAI-20)*, 2020.

[8] M. Donoser and H. Bischof, "Diffusion Processes for Retrieval Revisited," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1320–1327.

[9] S. Bai, X. Bai, Q. Tian, and L. J. Latecki, "Regularized diffusion process on bidirectional context for object retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 5, pp. 1213–1226, 2019.

[10] D. C. G. Pedronette, F. M. F. Gonçalves, and I. R. Guilherme, "Unsupervised manifold learning through reciprocal knn graph and connected components for image retrieval tasks," *Pattern Recognition*, vol. 75, pp. 161 – 174, 2018.

[11] D. C. G. Pedronette, L. P. Valem, J. Almeida, and R. da S. Torres, "Multimedia retrieval through unsupervised hypergraph-based manifold ranking," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5824–5838, 2019.

[12] K. Han, S. Rebuffi, S. Ehrhardt, A. Vedaldi, and A. Zisserman, "Automatically discovering and learning new visual categories with ranking statistics," in *8th International Conference on Learning Representations, ICLR 2020*, 2020.

[13] L. P. Valem, C. R. D. Oliveira, D. C. G. Pedronette, and J. Almeida, "Unsupervised similarity learning through rank correlation and knn sets," *ACM TOMCCAP*, vol. 14, no. 4, pp. 80:1–80:23, 2018.

[14] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," in *SODA'03*, 2003, pp. 28–36.

[15] M. Levandowsky and D. Winter, "Distance between sets," *Nature*, vol. 243, pp. 34 – 35, 1971.

[16] C. Y. Okada, D. C. G. Pedronette, and R. da S. Torres, "Unsupervised distance learning by rank correlation measures for image retrieval," in *ACM on International Conference on Multimedia Retrieval*, ser. ICMR '15, 2015, pp. 331–338.

[17] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Transactions on Information Systems*, vol. 28, no. 4, pp. 20:1–20:38, 2010.

[18] L. J. Latecki, R. Lakmper, and U. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *CVPR*, 2000, pp. 424–429.

[19] H. Ling, X. Yang, and L. J. Latecki, "Balancing deformability and discriminability for shape matching," in *ECCV*, 2010, pp. 411–424.

[20] D. C. G. Pedronette and R. da S. Torres, "Shape retrieval using contour features and distance optmization," in *VISAPP*, 2010.

[21] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1447–1454.

[22] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *CVPR*, 1997, pp. 762–768.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[24] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, no. 1, pp. 188 – 198, 2013.

[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[26] 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*, vol. 19, no. 2, pp. 120–131, 2009.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

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

[29] L. Franceschi, M. Niepert, M. Pontil, and X. He, "Learning discrete structures for graph neural networks," in *Thirty-sixth International Conference on Machine Learning (ICML 2019)*, 2019.

[30] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems 16*. MIT Press, 2004, pp. 321–328.

[31] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, vol. 3, no. 2, 2013.

[32] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," 2002.