Learning Visual Representations with Optimum-Path Forest and its Applications to Barrett’s Esophagus and Adenocarcinoma Diagnosis

Luis A. de Souza Jr. · Luis C. S. Afonso · Alanna Ebigbo · Andreas Probst · Helmut Messmann · Robert Mendel · Christian Hook · Christoph Palm · João P. Papa

Received: date / Accepted: date

Abstract Considering the rose number of the Barret’s esophagus (BE) number in the last decade, and its expectation of continue increasing, methods that can provide an early diagnosis of dysplasia in BE diagnosed patients may provide a high probability of cancer remission. The limitations related to traditional methods of BE detection and management encourage the creation of computer-aided tools to assist in this problem. In this work, we introduce the unsupervised Optimum-Path Forest (OPF) classifier for learning visual dictionaries in the context of Barrett’s esophagus (BE) and automatic adenocarcinoma diagnosis. The proposed approach was validated in two datasets (MICCAI 2015 and Augsburg) using three different feature extractors (SIFT, SURF, and the not yet applied to the BE context A-KAZE), as well as five supervised classifiers, including two variants of the OPF, Support Vector Machines with Radial Basis Function and Linear kernels, and a Bayesian classifier. Concerning MICCAI 2015 dataset, the best results were obtained using unsupervised OPF for dictionary generation using supervised OPF for classification purposes and using SURF feature extractor with accuracy nearly to 78% for distinguishing BE patients from adenocarcinoma ones. Regarding the Augsburg dataset, the most accurate results were also obtained using both OPF classifiers but with A-KAZE as the feature extractor with accuracy close to 73%. The combination of feature extraction and bag-of-visual-words techniques showed results that outperformed others obtained recently in the literature, as well as we highlight new advances in the related research area. Reinforcing the significance of this work, to the best of our knowledge, this is the first one that aimed at addressing computer-aided BE identification using bag-of-visual-words and OPF classifiers, being this application of unsupervised technique in the BE feature calculation the major contribution of this work. It is also proposed a new BE and adenocarcinoma description using the A-KAZE features, not yet applied in the literature.

Keywords Barrett’s esophagus · optimum-path forest · machine learning · adenocarcinoma · image processing

1 Introduction

Pattern classification has been paramount in the last decades, mainly due to the increasing number of applications that require some intelligent-decision-making mechanism. The standard pipeline adopted for so many years follows a robust but straightforward workflow: (i) feature extraction, (ii) model learning, and (iii) classification outcomes. The former step can be performed us-
ing handcrafted features or information learned through deep learning approaches. In this latter case, one may not know what kind of information the model is learning, since the set of outcome values that minimizes some loss function is the one employed in the model learning step. Handcrafted features require a more knowledgeable personnel, which is usually in charge of selecting and extracting features that matter when performing pattern classification.

Describing images using their most important information, the so-called “points of interest” (PoIs) or key points, has been an active area of interest by many researchers worldwide. Notable approaches have been proposed in the literature to compute those points, which somehow aim at capturing subtle information that is less variant to geometric transformations such as rotation, and translation, among others. Scale-Invariant Feature Transform (SIFT) [17], Speeded-Up-Robust-Features (SURF) [4], and Accelerated-KAZE features (A-KAZE) [2] are some examples.

However, the main problem related to the mentioned approaches concern the final feature vector. Since the number of PoIs may vary from one image to another, the feature vectors used to represent the images shall have different dimensions. To overcome this issue, an additional step called “quantization” is required (some works refer to this step as the “codebook generation” [4]). In a nutshell, given a training set composed of PoIs extracted from all training images, we can build a “bag” (i.e., a visual dictionary) with the most representative PoIs (from now on called “visual words”). Further, for each training image and each of its PoIs, we can find the “closest” visual word in the bag and build up a histogram that stores the number of times a visual word is nearest to each PoI from the training images. Therefore, the final feature vector of each training image will be that histogram with a dimensionality that corresponds to the number of visual words (i.e., the size of the dictionary or bag). Essentially, that is the main reason such approaches are usually referred to as “bag-of-visual-words” (BoVW) [6].

Bag-of-visual-words have been widely used in the literature for a number of different purposes, such as video-based action recognition [25], retinal health diagnosis [14], and perivascular spaces categorization in brain data [10], among others. Nonetheless, one still has two problems to face regarding the BoVW approach: (i) how to find out the most representative visual words, and (ii) how to establish a proper bag size, i.e., the number of visual words. Notice that both issues are pretty much crucial since they are in charge of the feature vector composition and dimensionality.

To cope with the first issue, i.e., finding out the most representative visual words, two approaches are commonly used: (i) random sampling and (ii) clustering. The former randomly selects a given number of visual words to compose the bag. On the other hand, clustering-based approaches make use of some unsupervised learning algorithm (usually k-means) to group the visual words, and the most representative ones (i.e., centroids) are elected to compose the dictionary [1]. However, randomly choosing visual words does not lead to good results, and the usage of certain unsupervised learning algorithms turns out to be a problem since most of them require the number of clusters (i.e., the bag size) beforehand.

Therefore, clustering techniques that do not require a priori information about the data are usually preferred. Among several techniques that have been proposed in the literature, one is gaining attention daily due to its effectiveness and efficiency in different research areas. The Optimum-Path Forest (OPF) is a framework for the design of pattern classifiers based on graph partition. In short, OPF-based classifiers work on a reward-competition process, in which previously selected samples called “prototypes” try to conquer other samples by offering them optimum-path costs. Once a sample is conquered by another one, it receives its label and a “mark” (i.e., a predecessor map) that reveals its conqueror. The Optimum-Path Forest framework comprises supervised [21,20,22], unsupervised [28], and semi-supervised [3] versions that have been widely employed in a number of applications, from remote sensing [27,19] to human intestinal parasites identification [35], just to cite a few.

One particular strength of unsupervised OPF concerns the fact it does not require the number of clusters beforehand, i.e., it finds clusters on-the-fly. Such feature is quite interesting in the context of BoVW generation since we skip the problem of choosing suitable bag sizes. As far as we are concerned, only two works attempted at using OPF in the context of BoVW: (i) Papa and Rocha [23] evaluated the supervised OPF for image categorization using visual words, and further (ii) Afonso et al. [1] studied the impact of using unsupervised OPF for learning proper visual dictionaries.

We are particularly interested in the application of such technique for the recognition of Barrett’s esophagus (BE), which happens to be a side effect of some reflux diseases. BE comprises a very severe and growing disease in the last decades, and since BE is often not identified properly at the early stages, it may evolve to a more aggressive version, and even to cancer. However, the early diagnosis of dysplastic tissue in BE diagnosed patients may provide very high rates of remission after
the treatment [7,31,36]. There are several endoscopic
techniques for the BE diagnosis and detection, such as
chromoendoscopy and narrow-band imaging, but the
human screening for the injured region definition is still
often misclassified by endoscopists, once the region does
not present enough goblet cells in biopsy or the experts
refuse to use the recommended procedure for extensive
biopsies [32]. Moreover, computer-assisted diagnosis
may bring precision and accuracy to the BE screening
and evaluation, once this task can be very influenced
by the human factor [35,37,36]. To the best of our
knowledge, only one very recent work coped with BE
identification using OPF. Souza et al. [35] introduced
the supervised OPF for Barrett’s esophagus automatic
identification using features based on BoVW. The au-
thors considered both random- and $k$-means-based
sampling strategies to build the visual dictionaries and then
used OPF for classification purposes.

Some works that dealt with endoscopic image analy-
sis can be referred to as well, but that is still an emerging
area of research [37]. Seibel et al. [30] developed
a low-cost but high-performance technology to assist
the diagnosis of BE and esophageal cancer. However,
their primary contributions rely on hardware advances
rather than software ones. The work presented by van
der Sommen [34] aimed at using machine learning tech-
niques to detect early neoplasia in Barrett’s esophagus,
and Swager et al. [39] addressed the very same context
mentioned above but using volumetric laser endoscopy
images.

Klomp et al. [13] proposed new features for
computer-aided Barrett’s esophagus identification, and
Hassan and Haque [11] used endoscopy videos obtained
from wireless capsules to assess gastrointestinal hemor-
hages. Later on, Seguí et al. [29] used the same source
of images (i.e., wireless capsules) together with Deep
Convolutional Neural Networks for intestine motility
characterization. Mendel et al. [18] started the study
of deep learning application to the BE and adenocarci-
noma evaluation problem.

The major problems around the computer-assisted
systems developed for the BE and adenocarcinoma eval-
uation are related to the type of technique to provide a
correct description of the injured areas and which clas-
sification techniques should be designed for the prob-
lem. These problems are related to all proposed works,
and considering the high potential in this research area,
new ways to describe the injured areas (that are very
similar), and the evaluation of different classifiers can
deliver important and substantial improvements to the
precision and correct differentiation of both.

As one can observe, Barrett’s esophagus automatic
identification using machine learning techniques presents
a growing interest in the last years. Therefore, there is
plenty of room for new works that employ techniques
that were not considered in such a context. In this work,
we extended and outperformed the approach proposed
by Souza et al. [35] by learning proper visual dictionar-
ies using unsupervised OPF, as well as we introduced a
variant of supervised OPF (OPF$_{knn}$) proposed by Papa
et al. [22] in the context of BE identification. The OPF
was never applied to such problem in the visual learn-
ing step, and this could deliver, besides the novelty in
the feature vector calculation, a new way to evaluate
the key points provided by the feature extraction tech-
niques. Last but not least, we introduce the A-KAZE
feature extraction technique for the calculation of the
key points, for comparison with SURF and SIFT, previ-
osly adopted for the BE and adenocarcinoma differen-
tiation context [37]. The results presented in this paper
are close to some state-of-the-art recognition rates [13],
and it features recent advances to BE automatic iden-
tification by means of machine learning and computer
vision. Therefore, the main contributions of this paper
are five-fold:

- to extend and outperform the recent results obtained
  by Souza et al. [35] in which the evaluation of BE
  and adenocarcinoma context were performed using:
  (i) SURF and SIFT techniques for key points calcu-
  lation, (ii) $k$-means and random techniques for the
  bag-of-visual-words calculation, and (iii) OPF and
  SVM classifiers for the classification task;
- to introduce OPF$_{knn}$ [22] for BE and adenocarci-
noma automatic diagnosis, considering that Souza
  et al. [35] employed only the complete graph ver-
  sion of OPF classifier for the classification task;
- to introduce A-KAZE features for the aforemen-
tioned context, once such technique has been largely
  applied in the literature for image description and
  retrieving;
- to extend the work by Afonso et al. [1] with a more
  robust evaluation of the unsupervised OPF for learn-
  ing visual dictionaries;
- to introduce a new representation of feature extrac-
tion techniques (such as SURF and SIFT) based on
  their most representative words in the feature space
  using the OPF clustering technique.

The remainder of this paper is organized as follows.
Sections 2 to 4 present a theoretical background of un-
supervised OPF and the methodology adopted in this
work, respectively. Section 5 discusses the experiments,
and Section 6 states conclusions and future works.
2 Unsupervised Learning with Optimum-Path Forest

In this section, we briefly present the theoretical background related to unsupervised OPF, which is used to learn proper visual dictionaries.

Let $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ be an unlabeled dataset such that $x_i \in \mathbb{R}^m$ stands for a feature vector extracted from some sample (i.e., images in our case) related to the problem to be addressed. Additionally, let $G = (\mathcal{D}, \mathcal{A}_k)$ be a graph derived from that dataset, which means $\mathcal{D}$ denotes the set of graph nodes (i.e., vertices) and $\mathcal{A}_k$ stands for a $k$-nearest neighbors adjacency relation.

In a nutshell, the OPF working mechanism is based on a reward-competition problem, where some samples called “prototypes” employ a competitive process among themselves to conquer the other samples from the dataset $\mathcal{D}$. Such competition ends up partitioning $\mathcal{D}$ into optimum-path trees (OPTs), which are rooted at each prototype node. It is worth mentioning that a sample that belongs to a given OPT is more “strongly connected” to the root and samples of that tree than to any other in the forest (i.e., a collection of all trees in the graph).

At a glance, the whole process can be summarized in the following steps:

1. To establish a proper neighborhood size and build up $\mathcal{A}_k$ (i.e., to find out “suitable” $k$ values);
2. To elect the prototypes and Learning Visual Representations with Optimum-Path Forest and its Applications to Barrett’s esophagus and Adenocarcinoma Diagnosis
3. To start the competition process.

Concerning step 1), a number of different approaches to cope with the task could be considered. Rocha et al. [28] proposed to compute the best value of $k$ (i.e., the neighborhood size), say that $k^*$, as the one that minimizes the normalized graph cut, which is a measure that considers both the dissimilarity between clusters as well as the similarity within the groups of samples [33].

Soon after computing $k^*$, the next move concerns finding the prototypes (i.e., step 2), also known as the “roots of the trees”. Such essential samples are in charge of ruling the competition process that ends up partitioning the graph into OPTs (i.e., clusters). Those samples will be used as the visual words to compose the final dictionary, as further discussed.

The supervised OPF proposed by Papa et al. [21] elects the prototypes as the nearest samples from different classes, which can be accomplished by computing a Minimum Spanning Tree (MST) over the training graph. Then, the samples from different classes that are connected in the MST are marked as prototypes. However, unsupervised OPF does not make use of labeled datasets, which motivated Rocha et al. [28] to elect the prototypes as the samples that are located at the center of the clusters. Such samples can be computed by assigning a density score $\rho(x_i)$ for each dataset sample $x_i \in \mathcal{D}$. That score is computed using a probability density function (pdf) given by a Gaussian distribution considered in the neighborhood of each sample as follows:

$$\rho(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{\forall x_j \in A_k(x_i)} \exp \left(-\frac{d(x_i, x_j)}{2\sigma^2}\right),$$

where $i \neq j$ and $\sigma = d_{\text{max}}/3$. In this case, $d_{\text{max}}$ stands for the maximum arc-weight in $G$. Using such formulation, $\rho(x_i)$ considers all adjacent nodes for the probability computation purposes since a Gaussian function covers 99.7% of the samples within $d(x_i, x_j) \in [0, 3\sigma]$.

After computing Equation (1) for all nodes, the competition process among samples can take place. Each density value will be used to populate a priority queue, where the idea of the unsupervised OPF algorithm is to end up maximizing the cost of each sample, and thus partitioning the graph.

The definition of “cost” is based on paths on graphs, i.e., a sequence of adjacent samples with no cycles. Let $\pi_{x_i}$ be a path with terminus at sample $x_i$ and starting from some root $\mathcal{R}(x_i)$, where $\mathcal{R}$ stands for the set of prototype samples. Additionally, let $\pi_{x_i} = (x_i)$ be a trivial path (i.e., a path composed of a single sample) and $\pi_{x_i} \cdot (x_i, x_j)$ the concatenation of $\pi_{x_i}$ and the arc $(x_i, x_j)$ such that $i \neq j$.

The OPF algorithm assigns to each path $\pi_{x_i}$ a value $f(\pi_{x_i})$ given by a connectivity function $f : \mathcal{X} \rightarrow \mathbb{R}$. In this context, a path $\pi_{x_i}$ is considered optimum if $f(\pi_{x_i}) \geq f(\pi_{x_j})$ for any other path $\pi_{x_j}$. Such sort of functions are known as “smooth functions”, and they figure important constraints that ensure the theoretic correctness of the OPF algorithm [8].

Among different path-cost functions that have been proposed in the literature, unsupervised OPF employs the following formulation for $\forall x_i, x_j \in \mathcal{D}$ such that $i \neq j$:

$$f((x_i)) = \begin{cases} \rho(x_i) & \text{if } x_i \in \mathcal{R} \\ \rho(x_i) - \delta & \text{otherwise,} \end{cases}$$

and

$$f(\pi_{x_i} \cdot (x_i, x_j)) = \min\{f(\pi_{x_i}), \rho(x_j)\},$$

where $\delta = \min\{\rho(x_i) \in A_k(x_i); |\tau(t) - \rho(s)| \rho(t) - \rho(s)\}$. In a nutshell, $\delta$ stands for the smallest quantity required...
to avoid plateaus in the regions nearby the prototypes (i.e., areas with the highest density).

Among all possible paths $\pi_{x_i}$ from the maxima of the pdf, the method assigns to sample $x_i$ a final path whose minimum density value along it is maximum. Such final path value is represented by a cost map $C$, as follows:

$$C(x_i) = \max_{\forall \pi_{x_j} \in (D, A_k), i \neq j} \{f(\pi_{x_j} \cdot \langle x_j, x_i \rangle)\}. \quad (4)$$

The OPF algorithm maximizes the connectivity map $\hat{C}(x_i), \forall x_i \in D$, by computing an optimum-path forest over the dataset. Such forest is encoded as a predecessor map $P$ with no cycles that assigns to each sample $x_i \notin R$ its predecessor $P(x_i)$ in the optimum path from $R$, or a marker nil when $x_i \in R$.

Figures 1 to 3 depict a toy example concerning the unsupervised OPF working mechanism. Figures 1a and 1b illustrate an unlabeled dataset and its 3-nearest neighbors graph, respectively (we assume $k = 3$ to explain step 1). For the sake of visualization purposes, we assigned the same color to each graph node and the arcs corresponding to its 3-nearest neighbors.

![Fig. 1](a) Toy example: (a) unlabeled dataset and its (b) 3-nearest neighbors graph.

Notice the arcs are also weighted by the distance (e.g., Euclidean distance) among their corresponding nodes. One can observe that some arcs and their weights are double-colored, which means their corresponding nodes share the very same 3-neighborhood.

Figure 2 illustrates the density computation step to further elect the prototypes (i.e., step 2). Therefore, given the arc-weights depicted in Figure 1b, we can use Equation 1 to compute $\rho(x_i), \forall x_i \in D$. Notice the density values are computed over the adjacency relation encoded by $A_k$. One can realize that the samples located at the center of the clusters tend to be the ones with the highest value of $\rho$ since they are connected by smaller arc-weights.

The density values are then stored in a priority queue (i.e., a max-heap) that pops out the sample $x_i$ with the highest $\rho(x_i)$. Concerning the toy example depicted in Figure 2, the first sample to come out of the queue is either ‘H’ or ‘A’ since both have the highest densities. Suppose ‘H’ has been added first to the queue. Since it has no predecessor, it is added to the set $R$ and assigned $f(H) = \rho(H) = 0.66$ according to Equation 2.

Further, the competition process (i.e., step 3) takes place. In short, sample ‘H’ evaluates its neighbors ‘I’, ‘J’, and ‘K’ to offer better costs to them (i.e., costs that are greater than the ones they have already). Therefore, one has $f(H \cdot \langle H, I \rangle) = \min \{0.66, 0.63\} = 0.63$, $f(H \cdot \langle H, J \rangle) = \min \{0.66, 0.65\} = 0.65$, and $f(H \cdot \langle H, K \rangle) = \min \{0.66, 0.65\} = 0.65$. Since the costs offered by ‘H’ are greater or equal than the costs of its neighbors, they are conquered by sample ‘H’. Such process is encoded by the aforementioned predecessor map $P$, i.e., after this first move of sample ‘H’, one has that $P(I) = H$, $P(J) = H$, and $P(K) = H$.

The next sample to start the competition process is sample ‘A’, and the very same process mentioned earlier is repeated until all samples have played in the competition process. The resulting optimum-path forest is depicted in Figure 3. Notice one can obtain a different number of clusters based on the value of $k_{max}$. In this toy example, we obtained two clusters, which are labeled with the same color of its prototype/root of the tree (i.e., the dashed nodes ‘A’ and ‘H’).
The unsupervised OPF algorithm finds the number of the clusters on-the-fly, which means there is no need to have such information beforehand. The only parameter that needs to be set is the \( k_{\text{max}} \), which constrains the search for suitable neighborhood sizes. One can observe that the knowledge required to set \( k_{\text{max}} \) is considerably lower than the one needed to set the number of clusters used by \( k \)-means, for instance. Such skill makes OPF pretty much attractive to the application addressed in this paper, as discussed in the next section.

3 Barrett’s Esophagus

The BE disease is known as the replacement of squamous cells by columnar cells in the esophagus. This process is a result of a complication of gastroesophageal reflux disease, being able to progress into esophageal cancer \[7,31\].

The incidence of BE and Barrett’s adenocarcinoma in the western population of the world has risen significantly in the past decade. Their close association with the metabolic syndrome suggest growth in the next years \[15,7,16\]. The early diagnosis of Esophageal adenocarcinoma in BE diagnosed patients is critical for remission and justifies the necessity of robust surveillance, detection, and characterization. However, the detection of dysplastic tissues and their characterization of abnormalities within BE-diagnosed patients can be challenging, especially for manual evaluation made by endoscopists. Even considering the dangerousness of the disease, when detected at the early stages, the disease can be treated with very high rates of remission (93% after 10 years) \[7,31,26\].

The esophagus mucosa is composed of squamous cells (similar to the skin or mouth cells), with a whitish-pink color surface, while the gastric mucosa goes sharply from salmon-pink to red \[7,31\]. The point in which the stomach and the stomach meet is called squamo-columnar junction or “Z-line”. BE’s mucosa may extend upward in a continuous pattern, changing the Z-line position \[7,31,26\]. Figure 4 shows the two cases in which patients can present long-segment of BE and short-segment of BE in a Z-line variation.

4 Methodology and Proposed Approach

In this section, we present the proposed approach and the methodology adopted to cope with the problem of Barrett’s esophagus automatic identification using bag-of-visual-words. First, the proposed method is defined, followed by the datasets used for the experiments, adopted classifiers and experimental delineation.

4.1 Proposed Method

As mentioned earlier, one of the leading contributions of this work is to evaluate the robustness of the OPF clustering for learning visual dictionaries. To fulfill that purpose, we considered three distinct feature descriptors based on key point extraction from images: (i) SIFT, (ii) SURF, and (iii) A-KAZE. Although any other approaches could be used, we opted to employ these mainly because they are well known and widely considered in the literature of bag-of-visual-words for both image classification and retrieval, but any other techniques could be applied considering the generalization of the learning visual dictionaries.

For the initial step of the proposed model, given a set of training images, it is needed to build a bag of key points extracted from them. In hands of a feature extraction technique for the description of an image (being SURF, SIFT or A-KAZE, in this specific case), the model aims to provide the most discriminative key points in the feature dimension based on the
entire feature domain. Therefore, taking into account the key points of the entire dataset, a clustering algorithm can be used to group the key points into clusters that share similar properties for choosing the “best key point” from each cluster and use it as the representative of that group. Such samples will compose the final bag-of-visual-words. The main contribution of this work is the calculation of such most representative key points from clusters (as we use to call “prototypes”) by using the OPF clustering technique. After obtaining the bag-of-visual-words, the last step of the model is known as “quantization” and computes the new representation for both training and testing images. For each image, it is computed the frequency of each visual word from the bag in the given image by finding the most similar visual word to each key point based on a distance metric. The outcome of that process is a histogram (feature vector) where each bin has the number of key points that are similar to its corresponding visual word. Notice that the representation of both training and testing images are computed based on the same bag. Finally, in hands of the feature vectors, each image of the evaluated dataset shows the exact same number of features for its description, but with the calculation based on the entire feature space domain. The training and testing may be conducted as the final step of the model.

In this work, we propose to cluster the dataset of key points using the OPF technique presented in the previous section and then use the prototypes to compose the bag-of-visual-words. As aforementioned, the prototypes are located in the regions of highest densities, which means they are pretty much suitable to describe the clusters. Another decisive point about OPF concerning other optimization-based clustering techniques relates the fact of not being attracted to local optima, such as $k$-means or $k$-medoids, for instance, which are widely used for learning dictionaries due to their simplicity and low computational cost.

As mentioned earlier, OPF finds the clusters on-the-fly, i.e., the clustering process is dynamic, and the forest configuration can change until the last sample finishes the conquering process. Instead of varying the size of the dictionary, one can change the value of $k_{\text{max}}$ and then may find the different number of clusters. The cluster calculation comprises one of the most important steps of the proposed method. The prototype computation is performed in an unsupervised process, turning the calculation of the feature vectors based only on the key points themselves, and providing a high generalization for this task. The problem of a different number of key points for each image can be solved using this bag-of-visual-words approach, proposing a consistent way of regular description for images evaluated by feature extraction techniques. Figure 5 depicts the pipeline adopted in this work.

Since the images are colored, a gray-scale normalization is applied to the images so that the key points can be extracted. Later, such PoIs are then mapped onto a feature space for clustering purposes. An example of the outcome of the clustering process is depicted at the bottom of Figure 5. Each color stands for a different group and the dashed nodes represent the prototypes selected by OPF to be part of the visual dictionary. As aforementioned, a histogram is built upon the training PoIs and the visual words for the further design of the final set of handcrafted features. For the selected visual words, an evaluation of their appearance is performed in the PoIs of each dataset image aiming to calculate the final cumulative histogram that represents each feature vector, with dimension depending on the number of visual words generated in the clustering calculation of the bag.

4.2 Datasets

An in-depth analysis concerning the robustness of the proposed approach is provided through two datasets. The first dataset comprises a set of images from a benchmark dataset provided at the “MICCAI 2015 EndoVis Challenge” was considered, hereinafter called “MICCAI 2015” dataset, which aimed at differentiating Barrett’s esophagus from cancerous images. Such dataset is composed of 100 endoscopic pictures of the lower esophagus captured from 39 individuals, 22 of them being diagnosed with early-stage Barrett’s, and 17 showing signs of esophageal adenocarcinoma. Each patient has several endoscopic images available, ranging from one to a maximum of eight. The database comprises a total of 50 images displaying cancerous tissue areas as well as 50 images showing dysplasia without signs of cancer. Suspicious lesions observed in the cancerous images had been delineated individually by five endoscopy experts.

Additionally, a dataset provided by the Augsburg Klinikum, Medizinische Klinik III was also used for the experiments. Such dataset is composed of 76 endoscopic images (esophagus) captured from different patients with adenocarcinoma (34 samples) and BE (42 samples). The images were annotated (manual segmentation of the adenocarcinoma’s and Barrett’s area, respectively) by an expert from the Augsburg Klinikum, and the diagnosis was provided using biopsy. Since we are dealing with a classification problem, the annota-

1. [https://endovissub-barrett.grand-challenge.org/](https://endovissub-barrett.grand-challenge.org/)
tions provided by the experts were not considered in our work.

Figure 6 depicts some examples of the MICCAI 2015 dataset positive for cancer (i.e., negative for BE) and their respective delineations performed by five experts. However, we are not working with the delineation information since we compute the PoIs for the whole image. One could use the information about the delineated regions to extract PoIs from that areas only, which could guarantee that pure adenocarcinoma PoIs are computed, but the problem still concerns the fact that delineations are not available to all real-world images.

Figure 7 displays some images positive for cancer from Augsburg dataset. In this case, we have only one delineation per image. Once again, such information is not used in this work since we are interested mostly in the differentiation of Barrett’s esophagus and adenocarcinoma rather than its segmentation.

4.3 Adopted Classifiers

We considered different supervised pattern recognition techniques to assess the robustness of unsupervised OPF for learning visual dictionaries:

- \( \text{OPF}_{\text{cpl}} \): supervised OPF with complete graph proposed by Papa et al. [21][20];
- \( \text{OPF}_{\text{knn}} \): supervised OPF with \( k \)-nn graph proposed by Papa et al. [22];
Fig. 6 Some examples of images positive for cancer and their respective delineations (MICCAI 2015 dataset).

Fig. 7 Some examples of images positive for cancer and their respective delineations (Augsburg dataset).

– SVM-RBF: Support Vector Machines with Radial Basis Function kernel and parameters optimized by cross-validation [5];
– SVM-Linear: Support Vector Machines with Linear kernel and parameters optimized by cross-validation [5];
– Bayes: standard Bayesian classifier.

Regarding the OPF-based classifiers, we used the LibOPF [24], which is an open-source library that implements both the supervised as well as the unsupervised versions of the OPF used in this work. With respect to the Bayesian classifier, we employed our own implementation.

4.4 Experimental Delineation

To compose the set of experiments, we considered three different sizes for the dictionaries: 100, 500, and 1,000. The main idea is to evaluate the robustness of the techniques used in this work under different scenarios. As we shall discuss later, the usage of dictionaries with 500 visual words seemed to achieve better results, as stated in a previous work [35], which motivated us to set $k_{max} = 500$ for this one. However, this not implies in constraining OPF to find exactly 500 clusters, just to limit the size of the neighborhood of each sample to be 500. Regarding OPF$_{knn}$, its parameter $k$ is fine-tuned within the range $[1, 500]$, and the value that maximized the accuracy over the training set was used.

Regarding the experimental validation, it was considered a cross-validation approach with 20 runs and using 70% of the dataset for training purposes, as well as the remaining 30% for classification. Moreover, the experimental results were assessed using a statistical analysis using the Wilcoxon signed-rank test with confidence as of 5% [40]. All experiments were conducted on an 8GB-memory computer equipped with an Intel Core i5 - 2.30 GHz processor. Additionally, we employed the OpenCV [14] implementation for feature extraction using SIFT, SURF, and A-KAZE.

5 Experimental Results

In this section, we present the experiments used to evaluate the proposed approach. Five supervised classifiers were considered to discriminate between samples positive and negative to adenocarcinoma: OPF$_{cpl}$, OPF$_{knn}$, SVM-RBF, SVM-Linear, and Bayesian classifier (hereinafter called Bayes). For all the classifiers adopted for
such evaluation, there was no need for setting any parameter, as long as they were used in the default set. The same experimental protocol was applied to all techniques using cross-validation, i.e., three distinct feature representations were considered (SURF, SIFT, and A-KAZE, with the metric threshold of all sets is default), and with different bag sizes (i.e., 100, 500 and 1,000 visual words). The results are presented and discussed considering each dataset individually.

A statistical evaluation using the signed-rank Wilcoxon test [10] was used for comparison purposes as follows:

1. For each dictionary generation approach (i.e., clustering by $k$-means, random or unsupervised OPF), it was verified the classification results and the best ones were highlighted in bold. Statistically similar results were highlighted in bold.
2. For each feature extractor (i.e., A-KAZE, SIFT, and SURF), the best statistical results were underlined.
3. Additionally, the best results among all configurations were marked with a ‘⋆’ symbol.

This very same procedure was adopted to both datasets.

In this work, we used the following accuracy rate:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100,$$

where $TP$ and $TN$ stand for the true positives and true negatives, respectively, and $FN$ and $FP$ denote the false negatives and false positives, respectively. In a nutshell, the above equation computes the ratio between the number of correct classifications (i.e., $TP + TN$) and the size of the dataset (i.e., all correct and wrong classifications).

### 5.1 MICCAI 2015 Dataset

Tables 1, 2, and 3 present the results related to A-KAZE, SURF, and SIFT descriptors, respectively, concerning MICCAI 2015 dataset. Regarding the A-KAZE results presented in Table 1, one can draw the following conclusions: (i) OPF$_{cpl}$ obtained the best results for all dictionary generation techniques, and (ii) OPF clustering achieved the best results (77.6% of recognition rate with 1,000 visual words) for BE recognition among all configurations, although being statistically similar to $k$-means with OPF$_{cpl}$ with 500 and 1,000 visual words as well.

The average number of PoIs used for training and test sets concerning A-KAZE feature extractor were 16,024 and 6,868, respectively, taking an average computational load of 4.05 minutes. A training set composed of around 16,000 visual words is enough to support the sizes of the dictionaries we used to build the feature vector of each image, i.e., 100, 500, 1,000. Larger dictionaries may not be interesting since there will be numerous small-sized clusters, which means less spatial information about the visual words is captured.

Table 2 presents the results concerning the SURF feature extractor. Once again, OPF$_{cpl}$ achieved the best classification results regarding all dictionary generation approaches, and OPF clustering allowed the best results among all, i.e., it could learn better dictionaries for image representation. In this context, a dictionary of size 500 computed by $k$-means also achieved the best recognition rates according to the statistical test. The average number of PoIs used for training and test sets concerning SURF feature extractor were 14,411 and 6,189, respectively, taking an average computational load of 13.77 minutes.

### Table 1 Mean accuracy results using A-KAZE features with 100, 500, and 1,000 visual words.

| Dictionary | 100     | 500     | 1000    |
|------------|---------|---------|---------|
| $k$-means  | OPF$_{cpl}$ 73.6% | ⋆76.3% | ⋆77.2% |
|            | SVM-RBF | 60.7%   | 65.9%   | 66.1%   |
|            | SVM-Linear | 58.5% | 63.0%   | 67.4%   |
|            | Bayes   | 56.8%   | 60.0%   | 60.9%   |
| Random     | OPF$_{cpl}$ 59.3% | 63.9%   | 70.3%   |
|            | SVM-RBF | 58.3%   | 61.7%   | 62.3%   |
|            | SVM-Linear | 62.1% | 65.6%   | 63.7%   |
|            | Bayes   | 55.3%   | 62.2%   | 61.1%   |
| OPF clustering | OPF$_{cpl}$ 72.3% | 73.1%   | 77.6%   |
|            | SVM-RBF | 61.9%   | 65.1%   | 70.9%   |
|            | SVM-Linear | 55.8% | 60.5%   | 66.8%   |
|            | Bayes   | 55.8%   | 58.0%   | 61.3%   |

### Table 2 Mean accuracy results using SURF Features and 100, 500, and 1,000 visual words.

| Dictionary | 100     | 500     | 1000    |
|------------|---------|---------|---------|
| $k$-means  | OPF$_{cpl}$ 70.0% | 74.8%   | 73.6%   |
|            | SVM-RBF | 64.1%   | 66.0%   | 65.1%   |
|            | SVM-Linear | 63.6% | 64.8%   | 62.6%   |
|            | Bayes   | 56.4%   | 56.9%   | 57.4%   |
| Random     | OPF$_{cpl}$ 60.5% | 70.2%   | 66.1%   |
|            | SVM-RBF | 58.0%   | 58.4%   | 61.8%   |
|            | SVM-Linear | 61.0% | 63.4%   | 62.1%   |
|            | Bayes   | 53.5%   | 58.5%   | 56.5%   |
| OPF clustering | OPF$_{cpl}$ 69.4% | 75.4%   | 77.1%   |
|            | SVM-RBF | 67.5%   | 71.8%   | 70.9%   |
|            | SVM-Linear | 65.1% | 66.9%   | 66.7%   |
|            | Bayes   | 53.5%   | 56.8%   | 57.1%   |

One can observe that SVM did not obtain proper recognition rates in both situations, i.e., A-KAZE and SURF feature extractors. One possible reason concerns
the number of training samples, which is usually lower than the number of features. Therefore, SVM will map samples to a lower-dimensionality feature space instead of a higher one, thus neglecting the assumption of linearity in higher-dimensionality spaces.

Table 3 presents the results considering the SIFT feature extractor. Once again, OPF_{cpl} achieved the best results so far, with OPF_{knn} and SVM-RBF being statistically similar for k-means with 1,000 words and a random generation of dictionaries with 1,000 words. However, the best global results were achieved using OPF clustering with OPF_{cpl} with 500 and 1,000 visual words, outperforming by far the other results with SIFT feature extractor. The average number of Pols used for training and test sets concerning SIFT feature extractor were 28, 137, and 12, 059, respectively, taking an average computational load of 5.95 minutes.

Table 3 Mean accuracy results using SIFT Features and 100, 500, and 1,000 visual words.

| Dictionary | 100 | 500 | 1000 |
|------------|-----|-----|------|
| k-means    |     |     |      |
| OPF_{cpl}  | 68.3% | 72.3% | 71.4% |
| OPF_{knn}  | 67.0% | 71.8% | 72.1% |
| SVM-RBF    | 67.3% | 71.4% | 71.9% |
| SVM-Linear | 55.2% | 56.8% | 57.3% |
| Bayes      | 59.8% | 60.0% | 60.7% |
| Random     |     |     |      |
| OPF_{cpl}  | 66.4% | 69.7% | 71.2% |
| OPF_{knn}  | 58.1% | 63.9% | 66.1% |
| SVM-RBF    | 62.1% | 65.6% | 63.7% |
| SVM-Linear | 53.2% | 54.5% | 52.7% |
| Bayes      | 56.2% | 53.0% | 54.4% |
| OPF clustering |     |     |      |
| OPF_{cpl}  | 71.2% | 77.7% | 78.9% |
| OPF_{knn}  | 63.9% | 71.3% | 75.7% |
| SVM-RBF    | 68.0% | 70.2% | 69.7% |
| SVM-Linear | 61.3% | 64.7% | 64.4% |
| Bayes      | 50.2% | 53.0% | 54.4% |

Last but not least, the best results among all three feature extractors (i.e., the ones marked with ‘⋆’) were obtained using OPF clustering for dictionary generation and OPF_{cpl} for classification with 500 and 1,000 visual words considering SURF and SIFT, and the same pair (i.e., OPF clustering and OPF_{cpl}) regarding A-KAZE with 1,000 visual words, and finally k-means and OPF_{cpl} with 500 words. Notice the best absolute result was obtained using OPF clustering for visual words generation and OPF_{cpl} for classification purposes with SIFT-based features (i.e., 78.9%).

5.2 Augsburg Dataset

Table 4 presents the results concerning the SURF feature extractor. Once again, OPF-based classifiers obtained the best results in most of the scenarios, being OPF clustering and k-means the best approaches for visual dictionary generation. The best absolute classification results were obtained by OPF_{cpl} and Bayes with accuracies nearly to 68%. The average number of Pols used for training and test sets concerning SURF feature extractor were 14, 251 and 6, 108, respectively, taking an average computational load of 4.18 minutes.

Table 4 Mean accuracy results using A-KAZE Features and 100, 500, and 1,000 visual words.

| Dictionary | 100 | 500 | 1000 |
|------------|-----|-----|------|
| k-means    |     |     |      |
| OPF_{cpl}  | 60.7% | 69.9% | 65.6% |
| OPF_{knn}  | 61.0% | 66.1% | 70.1% |
| SVM-RBF    | 60.4% | 63.5% | 63.1% |
| SVM-Linear | 55.3% | 60.4% | 62.1% |
| Bayes      | 56.9% | 60.1% | 63.3% |
| Random     |     |     |      |
| OPF_{cpl}  | 59.4% | 68.4% | 69.9% |
| OPF_{knn}  | 59.9% | 61.4% | 62.4% |
| SVM-RBF    | 57.9% | 62.2% | 63.2% |
| SVM-Linear | 53.5% | 58.8% | 58.9% |
| Bayes      | 56.5% | 57.1% | 61.0% |
| OPF clustering |     |     |      |
| OPF_{cpl}  | 68.4% | 68.7% | 72.0% |
| OPF_{knn}  | 59.8% | 60.3% | 70.3% |
| SVM-RBF    | 59.4% | 63.0% | 69.8% |
| SVM-Linear | 57.7% | 57.3% | 62.7% |
| Bayes      | 62.4% | 60.7% | 63.1% |

Table 5 presents the results concerning the SIFT feature extractor. In this case, OPF-based classifiers and SVM-RBF figured as the most accurate techniques and OPF clustering as the best one for visual dictionary generation (absolute results). A comparison against A-
KAZE and SURF showed these to be quite more accurate than SIFT, an opposite situation that occurred over MICCAI 2015 dataset, where SIFT achieved the best recognition rates. Additionally, the average number of PoIs used for training and test sets concerning SIFT feature extractor were 89, 514 and 38, 363, respectively, taking an average computational load of 8.71 minutes.

Table 6 Mean accuracy results using SIFT Features and 100, 500, and 1,000 words.

| Dictionary       | 100   | 500   | 1000  |
|------------------|-------|-------|-------|
| k-means          |       |       |       |
| OPF\textsubscript{cpl} | 60.3% | 60.5% | 59.3% |
| OPF\textsubscript{knn}  | 58.9% | 60.6% | 62.0% |
| SVM-RBF          | 60.8% | 61.8% | 59.8% |
| SVM-Linear       | 55.5% | 57.1% | 59.9% |
| Bayes            | 53.1% | 54.8% | 58.7% |
| Random           |       |       |       |
| OPF\textsubscript{cpl} | 59.2% | 60.5% | 61.6% |
| OPF\textsubscript{knn}  | 57.0% | 58.4% | 60.5% |
| SVM-RBF          | 57.8% | 62.6% | 62.1% |
| SVM-Linear       | 54.4% | 55.6% | 61.5% |
| Bayes            | 51.9% | 57.0% | 59.0% |
| OPF clustering   |       |       |       |
| OPF\textsubscript{cpl} | 60.4% | 62.8% | 62.1% |
| OPF\textsubscript{knn}  | 58.1% | 61.6% | 63.9% |
| SVM-RBF          | 57.0% | 60.5% | 62.1% |
| SVM-Linear       | 58.8% | 58.9% | 58.7% |
| Bayes            | 61.1% | 62.2% | 61.8% |

5.3 Discussion

In this section, we aim at providing a more in-depth discussion about the experiments, as well as insightful conclusions regarding the usage of bag-of-visual words in the context of computer-aided differentiation between Barrett’s esophagus and adenocarcinoma. Table 7 presents a summary with the best results obtained in the previous two sections concerning the number of visual words and feature extractor. Concerning both datasets, OPF\textsubscript{cpl} figured as the more accurate classification technique, meanwhile OPF clustering appears as the best dictionary generation approach.

Table 7 Summarization of the results.

| Dataset       | Accuracy | Feature Extractor | #visual words |
|---------------|----------|-------------------|---------------|
| MICCAI 2015   | 78.9%    | SIFT              | 1,000         |
| Augsburg      | 72.6%    | A-KAZE            | 1,000         |

The results support the primary contributions stated previously, which are related to the robustness of OPF-based classifiers for both supervised and unsupervised learning in the context of automatic adenocarcinoma identification. Additionally, the number of visual words strongly affects the results, but we believe a trade-off between the size of the dictionary and the information it carries on shall be established beforehand.

Table 8 summarizes the mean sensitivity and specificity results of both datasets with the best configuration of the number of visual words, dictionary generation approach, feature extractor, and classification technique mentioned above. Sensitivity stands for the classification rate considering adenocarcinoma identification, i.e., those positive to Barrett’s esophagus and to adenocarcinoma, and specificity denotes the accuracy regarding those negative to adenocarcinoma, i.e., positive only to BE. Considering such sensitivity and specificity results, some conclusions can be drawn: (i) for the MICCAI 2015 dataset, the sensitivity results presented higher values than the specificity ones, suggesting a very good generalization in the positive adenocarcinoma identification. Even with lower results, the specificity still showed a convincing value, and the misclassification can be justified by two factors: the fuzzy region (region in which the experts disagree in the annotation) and lack of enough key points in the noncancerous regions during the feature vector calculation. For the Augsburg results of sensitivity and specificity, a better trade-off between the correct classification of positive and non-positive adenocarcinoma samples could be found, but still with lower results when compared to the MICCAI 2015 dataset ones. The Augsburg dataset presents images with different behavior and acquisition technology when compared to the MICCAI 2015 ones, thus justifying the worse results.

To provide more insightful comments and to better understand the working mechanism of visual words in the context of computer-assisted BE identification, we performed some additional experiments with cancerous images that were classified either as cancer or as Barrett’s esophagus since we have their delineated regions.

Table 5 Mean accuracy results using SURF Features and 100, 500, and 1,000 words.

| Dictionary       | 100   | 500   | 1000  |
|------------------|-------|-------|-------|
| k-means          |       |       |       |
| OPF\textsubscript{cpl} | 66.3% | 67.9% | 61.5% |
| OPF\textsubscript{knn}  | 62.8% | 63.2% | 65.4% |
| SVM-RBF          | 57.1% | 61.1% | 62.9% |
| SVM-Linear       | 56.7% | 57.1% | 59.4% |
| Bayes            | 60.8% | 59.9% | 61.1% |
| Random           |       |       |       |
| OPF\textsubscript{cpl} | 60.0% | 62.2% | 63.5% |
| OPF\textsubscript{knn}  | 54.2% | 58.1% | 60.8% |
| SVM-RBF          | 61.3% | 61.9% | 62.0% |
| SVM-Linear       | 57.1% | 55.4% | 56.4% |
| Bayes            | 51.9% | 59.0% | 59.1% |
| OPF clustering   |       |       |       |
| OPF\textsubscript{cpl} | 59.2% | 62.1% | 66.1% |
| OPF\textsubscript{knn}  | 61.1% | 63.9% | 64.5% |
| SVM-RBF          | 58.5% | 62.0% | 65.4% |
| SVM-Linear       | 53.5% | 60.4% | 64.6% |
| Bayes            | 59.8% | 67.0% | 67.9% |
In a nutshell, the main idea is to compute the percentage of PoIs located inside those regions with respect to the remaining ones (i.e., those located outside cancerous areas). This information allows us to compare whether the number of PoIs placed inside the delineated regions are enough or not to provide accurate classifications.

Table 9 presents the mean percentage of PoIs located inside the cancerous area for the whole dataset, as well as the average percentage of PoI inside the cancerous area concerning the misclassified images (i.e., cancerous images that were classified as BE). Since we conducted a cross-validation approach with 20 runs, the average percentages concerning the misclassified images (i.e., Cancer→BE) were computed to each run, for the further computation of the average value of all. Additionally, since MICCAI 2015 dataset comprises delineations from five experts, we took the intersection of them all as the final delineated area to compute the percentage of PoIs into account.

One can observe that the percentage of PoIs inside the cancerous images were more significant than the values obtained from the misclassified images. Such assumption is pretty interesting since we can conclude that the number of PoIs inside the delineated regions are essential to achieve accurate results and to avoid misclassifications. The only exception stands for the Augsburg dataset with A-KAZE features, where the number of PoIs were slightly higher for the misclassified images. Note that the percentage of PoIs inside the cancer region is in general higher for the Augsburg dataset than for the MICCAI 2015 dataset. This can be explained because the Augsburg images use the near-focal imaging technique, in which the suspicious region is displayed larger.

For visualization purposes, Figures 8 to 9 depict some cancer patients that were misclassified as BE from both datasets. The PoIs showed in Figure 8 were calculated using SIFT and belong to the MICCAI 2015 dataset, and their percentage of incidence is 21.72%, which is slightly lower considering the average percentage presented in Table 9 (23.04%).

One can observe a considerable amount of PoIs located at the left-middle portion of Figure 8, mainly due to some air bubbles and foam. Problems with light (upper part of the image) also contribute to placing PoIs outside the delineated area.

The PoIs showed in Figure 9 were calculated using A-KAZE on an image from the Augsburg dataset. Their percentage of incidence is 7.5%, which is quite low considering the average percentage presented in Table 9 (53.77%). In this case, the main reason for placing PoIs outside the delineated area concerns illumination problems (brighter areas).

6 Conclusions and Future Works

In this paper, we dealt with the problem of computer-assisted Barrett’s esophagus identification by means of bag-of-visual-words calculated using the OPF clustering technique. Such technique showed promising results, outperforming the previous handcrafted feature results in the same context. This suggests the generalization relevance of such technique, which can improve previous results in the same field not only for the BE context but for other in which the image representation configures the context to be evaluated. BE stands for an illness that is likely to be confused with adenocarcinoma, and its early detection and prevention is of great concern.

We observed that only a very few works attempted at coping with the problem of automatic BE identification using computer vision and machine learning techniques to date. In this work, we fostered the research towards such area by introducing a supervised variant of the Optimum-Path Forest classifier for automatic BE recognition, as well as we showed how to build proper visual dictionaries using unsupervised OPF learning, outperforming the results obtained in some recent works in which the same database and protocol were applied [35,36]. Considering some previous works [35,37,38], the use of handcrafted features were based on the SURF and SIFT PoIs, but without the use of the OPF clustering as a way of dimensional reduction of the problem. Moreover, considering the improvements of the results, the use of the OPF clustering provides a new and promising way of BE and adenocarcinoma problem evaluation based on extracted key
points. The presented results showed the relevance of such technique addressed to the BE and adenocarcinoma evaluation and description, contributing to the context literature and influencing the evaluation and description of other tissue diseases. Comparing the proposed method with others already published, we can ensure that with the use of the OPF for the BoVW step, improvements could be achieved considering the higher results obtained. Also, such technique provides advantages in the dimension reduction of the feature vector calculation, once even with a different number of key points per image, a standard method of feature calculation is established. Again, the OPF clustering may provide flexibility and time saving for such task.

The experimental results were considered over two datasets: (i) MICCAI 2015, and (ii) Augsburg. For both scenarios, we evaluated five classification techniques and three unsupervised learning approaches to build the visual dictionaries. Also, we considered dictionaries with three distinct sizes and even three different feature extractors.

The experiments pointed out that bag-of-visual-words techniques are suitable to handle BE automatic identification, and there must be a trade-off between the number of visual words and the amount of information they can encode (i.e., size of the clusters). Additionally, both supervised and unsupervised OPF-based classifiers achieved the most accurate results, thus supporting the main contributions of this paper.

In the following, a bullet list of trends based on the achieved results is presented:

- the OPF classifier presented the highest results of accuracy in all experiments, and may be highly recommended considering the high generalization that provided for such a context, even for different description scenarios;
- the representation of BE and adenocarcinoma by means of image description techniques may provide encouraging results, and with less computation processing cost as needed in more sophisticated techniques;
- the A-KAZE features showed the very best results for BE and adenocarcinoma description in the Augsburg dataset evaluation, suggesting to be a very important technique for the description of such diseases;
- the use of handcrafted features still has potential to evaluated for BE and adenocarcinoma problem, considering the several number of techniques, such as fisher vectors and sparse coding;
the use of OPF clustering improved the current results and can be applied to a large number of cases of image description of the BE and adenocarcinoma regions;

- the way of improving the selection of key points in each region (cancerous and non-cancerous) still shows potential considering the influence of the number of key points in each region for the correct classification result.

Regarding future works, we aim at considering deep learning and post-processing techniques after the construction of the bags, such as feature selection (i.e., visual word selection). Additionally, this post-processing can be performed using the large number of machine learning techniques, such as SVM, OPF or even Convolutional Neural Networks, providing intermediate learning for the dictionaries calculation. More techniques for image description are also considered to be evaluated using the bag-of-visual-words provided by the OPF clustering.

**Acknowledgements**

The authors are grateful to DFG grant PA 1595/3-1, Capes/Alexander von Humboldt Foundation grant number BEX 0581-16-0, CNPq grants 306166/2014-3 and 307066/2017-7, as well as FAPESP grants 2013/07375-0, 2014/12236-1, and 2016/19403-6. This material is based upon work supported in part by funds provided by Intel AI Academy program under Fundesp Grant No.2597.2017.

**References**

1. Afonso, L.C.S., Papa, J.P., Papa, L.P., Marana, A.N., Rocha, A.R.: Automatic visual dictionary generation through optimum-path forest clustering. In: 2012 19th IEEE International Conference on Image Processing, pp. 1897–1900 (2012)

2. Alcantarilla, P.F., Nuevo, J., Bartoli, A.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: Proceedings of the British Machine Vision Conference, BMVC, pp. 1–11 (2013)

3. Amorim, W.P., Falcão, A.X., Papa, J.P., Carvalho, M.H.: Improving semi-supervised learning through optimum connectivity. Pattern Recognition 60, 72–85 (2016)

4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer Vision and Image Understanding 110(3), 346–359 (2008)
5. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2(3), 1–27 (2011)

6. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Proceedings of the Workshop on Statistical Learning in Computer Vision, pp. 1–22 (2004)

7. Dent, J.: Barrett’s esophagus: a historical perspective, an update on core practicalities and predictions on future evolutions of management. Journal of Gastroenterology and Hepatology 26, 11–30 (2011)

8. Falcão, A.X., Stolfi, J., Lotufo, R.A.: The image foresting transform: theory, algorithms, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(1), 19–29 (2004)

9. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, vol. 2, pp. 524–531 (2005)

10. González-Castro, V., Valdés-Hernández, M.C., Armitage, J.: Barrett’s esophagus. Endoscopy 48(5), 555–562 (2016)

11. Hassan, A.R., Haque, M.A.: Computer-aided gastrointestinal hemorrhage detection in wireless capsule endoscopy videos. Computers in Biology and Medicine 122, 341–353 (2015)

12. Itseez: Open source computer vision library. https://github.com/itseez/opencv (2015)

13. Klomp, S., van der Sommen, F., Swager, A.F., Zinger, S., Schoon, E.J., Curvers, W.L., Bergman, J.J., de With, P.H.N.: Evaluation of image features and classification methods for barrett’s cancer detection using vle imaging. In: Proceedings of the SPIE Medical Imaging, vol. 10134, p. 101340D (2017)

14. Koh, J.E.W., Ng, E.Y.K., Bhandary, S.V., Hagiwara, Y., Lowe, D.G.: Distinctive image features from scale-invariant spaces in Brain MRI Using Bag of Visual Words, ICIAR, vol. 9730, pp. 642–649. Springer International Publishing (2016)

15. Lagergren, J., Lagergren, P.: Oesophageal cancer. BMJ 313, 117–126 (2006)

16. Papuc, V., Nassar, M., Vlachos, D.I.: Effective area-based binary local descriptors. In: Proceedings of the 18th IEEE International Conference on Image Processing, pp. 3525–3528 (2011)

17. Peng, X., Wang, L., Wang, X., Qiao, Y.: Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. Computer Vision and Image Understanding 150, 109–125 (2016)

18. Mandel, R., Ebigbo, A., Probst, A., Messmann, H., Palm, C.: Multimodal and high-performance alternative technology for the screening of esophageal cancer and barrett’s esophagus. IEEE Transactions on Biomedical Engineering 55(3), 1032–1042 (2008)

19. McMahon, L., Zhang, K., Poon, R.E., Bisschops, R., Pech, O., Raghunath, K., Weusten, B.L.A.M., et al.: Multimodality endoscopic eradication for neoplastic barrett oesophagus: results of an European multicentre study (euro-il). Gut 65(4), 555–562 (2016)

20. Papa, J.P., Fernandes, S.E.N., Falcão, A.X.: Optimum-path forest based on k-connectivity: Theory and applications. Pattern Recognition Letters 87, 117–126 (2017)

21. Papa, J.P., Rocha, A.R.: Image categorization through optimum path forest and visual words. In: Proceedings of the 18th IEEE International Conference on Image Processing, pp. 3525–3528 (2011)

22. Probst, A., Messmann, H., Palm, C.: Barrett’s esophagus analysis using convolutional neural networks. Pattern Recognition 45(1), 512–520 (2012)

23. Probst, A., Messmann, H., Palm, C.: Nature-inspired framework for hyperspectral band selection. IEEE Transactions on Geoscience and Remote Sensing 52(4), 2126–2137 (2014). DOI 10.1109/TGRS.2013.2258351

24. R., Probst, A., Messmann, H., Palm, C.: Barrett’s esophagus analysis using convolutional neural networks. Pattern Recognition 45(1), 512–520 (2012)

25. Seibel, E., Carroll, R.E., Dominitz, J.A., Johnston, R.S., Melville, C.D., Lee, C.M., Seitz, S.M., Kimmy, M.B.: Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and barrett’s esophagus. IEEE Transactions on Biomedical Engineering 55(3), 1032–1042 (2008)

26. Sharma, P., Bergman, J.J.G.H.M., Goda, K., Kato, M., et al.: Development and validation of a classification system to identify high-grade dysplasia and esophageal adenocarcinoma in barrett’s esophagus using narrow-band imaging, Gastroenterology 150(3), 591 – 598 (2016)

27. Sharma, P., Brill, J., Canto, M., DeMarco, D., Ferran, M.B., Gupta, N., L.: White paper aga: Advanced imaging. Gastroenterology 150(3), 591 – 598 (2016)

28. Segui, S., Drozdalz, M., Pascual, G., Radeva, P., Malage, C., Azpiroz, F., Vitià, J.: Generic feature learning for wireless capsule endoscopy analysis. Computers in Biology and Medicine 79, 163–172 (2016)

29. Papa, J.P., Fernandes, S.E.N., Falcão, A.X.: Optimum-path forest based on k-connectivity: Theory and applications. Pattern Recognition Letters 87, 117–126 (2017)

30. Seibel, E., Carroll, R.E., Dominitz, J.A., Johnston, R.S., Melville, C.D., Lee, C.M., Seitz, S.M., Kimmy, M.B.: Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and barrett’s esophagus. IEEE Transactions on Biomedical Engineering 55(3), 1032–1042 (2008)

31. Seibel, E., Carroll, R.E., Dominitz, J.A., Johnston, R.S., Melville, C.D., Lee, C.M., Seitz, S.M., Kimmy, M.B.: Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and barrett’s esophagus. IEEE Transactions on Biomedical Engineering 55(3), 1032–1042 (2008)

32. Seibel, E., Carroll, R.E., Dominitz, J.A., Johnston, R.S., Melville, C.D., Lee, C.M., Seitz, S.M., Kimmy, M.B.: Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and barrett’s esophagus. IEEE Transactions on Biomedical Engineering 55(3), 1032–1042 (2008)

33. Seibel, E., Carroll, R.E., Dominitz, J.A., Johnston, R.S., Melville, C.D., Lee, C.M., Seitz, S.M., Kimmy, M.B.: Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and barrett’s esophagus. IEEE Transactions on Biomedical Engineering 55(3), 1032–1042 (2008)
37. Souza Jr, L.A., Palm, C., Mendel, R., Hook, C., Ebigbo, A., Probst, A., Messmann, H., W., S., Papa, J.P.: A survey on Barrett’s esophagus analysis using machine learning. Computers in Biology and Medicine 96, 203–213 (2018). DOI 10.1016/j.compbiomed.2018.03.014

38. Suzuki, C.T.N., Gomes, J.F., Falcão, A.X., Papa, J.P., Hoshino-Shimizu, S.: Automatic segmentation and classification of human intestinal parasites from microscopy images. IEEE Transactions on Biomedical Engineering 60(3), 803–812 (2013)

39. Swager, A.F., van der Sommen, F., Zinger, S., Meijer, S.L., Schoon, E.J., Bergman, J., de With, P.H., Curvers, W.L.: Feasibility of a computer algorithm for detection of early Barrett’s neoplasia using volumetric laser endomicroscopy. Gastroenterology 150(4, Supplement 1), S56 (2016)

40. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bulletin 1(6), 80–83 (1945)