Particle Swarm Estimation of Distribution Algorithm for Lymphoma Classification through Automatic Biopsies Analysis

Valentino Santucci, Alfredo Milani
Department of Mathematics and Computer Science
University of Perugia, Italy
Email: valentino.santucci@dmi.unipg.it, milani@unipg.it

***Abstract*—In this paper we propose a new system for lymphoma classification through automatic biopsy images analysis. The system is composed by two main modules: a computer vision module that extracts numerical features from the biopsy images acquired from a microscope, and a machine learning module that, basing on the numerical features of the images, builds an automated classifier that predicts the class label (i.e., the lymphoma type) of a new and unrecognized biopsy image coming in the system. Particle Swarm Estimation of Distribution Algorithm (PSEDA), a recently proposed meta-heuristic technique that hybridize Particle Swarm Optimization (PSO) and Estimation of Distribution Algorithms (EDAs), has been employed in order to perform the training of the classifier. Experiments were conducted on a standard and publicly available dataset of lymp-nodes tissue biopsy images, and they show that our approach results in a good classification accuracy with respect to other state-of-the-art and evolutionary classification schemes.**

I. Introduction

During the last years, automated image analysis tools have received a growing attention in the medical research field. Devices for automated image acquisition are able to provide a large amount of images of medical and biological interest. Think for example at biopsies and tissues images acquired by an electronic microscope, brain TAC or mammograms images. However, while medical image acquisition systems can provide a so large amount of images data, human analysis of these data is still very slow. This issue introduces the need for computer vision and machine learning techniques able to automatically analyze the acquired images and to provide some "human-understandable" results of their elaboration, for example in form of an automatic judgment or diagnosis. Although automated diagnoses are not mandatory for the doctor judgment, they can be employed as a filter for the large amount of data the doctor should analyze.

In this work we propose an automated system that, by analyzing a lymph-nodes biopsy image, is able to classify the type of lymphoma affecting the tissue contained in the biopsy image. Malignant lymphoma is a cancer affecting lymph-nodes. Three different types of malignant lymphoma are usually distinguished from the medical community, i.e., "chronic lymphocytic leukemia" (CLL), "follicular lymphoma" (FL), and "mantle cell lymphoma" (MCL). Basing on the recognized category of lymphoma, medical treatments on the affected patient can widely vary, thus this problem represents a relevant and practical issue for medical community.

The biopsies dataset adopted in this work is taken from [19], and it is composed by 374 images acquired through a brightfield microscope and representing biopsies sectioned and stained with hematoxylin/eosin. Moreover, for each biopsy image, the dataset provides also the information of the type of lymphoma to which it belongs.

The designed system is composed by two main modules: the computer vision module that elaborate the images in order to extract some numerical features, and the machine learning module that, basing on the numerical features of the images, builds an automated classifier that predicts, with an hopefully good accuracy, the class label (i.e., the lymphoma type) of a new and unrecognized biopsy image coming in the system.

For the computer vision module we have adopted some standard image features extraction techniques [10], [23], while the machine learning module has been designed by using the Particle Swarm Estimation of Distribution Algorithm (PSEDA) meta-heuristic [18].

The number of extracted features has been kept low in order to improve the execution times of the system as a whole. The numerical space induced by the features extraction step is probed by PSEDA which is a population-based optimization technique originally proposed in [18]. It combines some aspects of two widely adopted classes of meta-heuristic techniques, that is, Particle Swarm Optimization (PSO) [13] and Estimation of Distribution Algorithms (EDAs) [15]. Indeed, PSEDA combines the typical swarm-intelligence dynamics of PSO with the EDAs evolutionary search schemes based on an explicit probabilistic model of the population of candidate solutions. Although other PSO-EDAs hybridizations have already been proposed in literature (see for instance [5], [9], [24]), PSEDA exploits the EDAs probabilistic search in the PSO dynamics in a more neatly way.

In this work, PSEDA has been applied in order to build the prediction function of the biopsies classifier system. To do this, the training phase of the classification task has been cast to a numerical optimization problem by adopting some appropriate fitness functions that allow to judge the quality of a classifier on a given training set of labeled data (i.e., the images features and the class labels). A similar approach, although not employed for images classification, has been proposed in [3].

The rest of the paper is organized as follows. In Section II, the PSEDA numerical optimizer is described. Section III

provides a detailed description of the full classification system that is the aim of this work. Experimental results are provided in Section IV, while conclusions are drawn in Section V.

II. Particle Swarm Estimation of Distribution Algorithm

A. The philosophy

Particle Swarm Estimation of Distribution Algorithm (PSEDA), previously proposed in [18], is an iterative population-based meta-heuristic technique for numerical optimization problems that combines the typical swarm-intelligence dynamics of PSO [13] with the EDAs evolutionary search schemes [15] based on an explicit probabilistic model of the population of candidate solutions.

As in PSO, at each generation, a PSEDA individual tends to move in the search space toward a stochastic combination of its current position, the personal and social best positions found so far, and by also taking into account its previous search direction. The implementation of this stochastic movement in PSEDA simultaneously represents the difference with PSO and the analogy with the EDAs. Indeed, in PSEDA, the velocity concept of PSO is not employed and each individual moves to a new position in the search space by sampling a probability distribution built basing on the attraction positions adopted in PSO.

However, it is remarkably to note that, conversely from the more common EDAs in literature (see for instance [6], [8]), the PSEDA probabilistic model is not unique for the whole population. Indeed, each PSEDA individual has its own probabilistic model that iteratively updates and samples in order to move in the search space. Thus, instead of exploiting the statistical properties of the fittest part of the population as a whole (as done by typical EDAs), PSEDA relies on the implicit inter-dependencies modeling mechanism of swarm-intelligence based schemes [12]. Indeed, some statistical parameters are shared among the PSEDA probabilistic models, that is, among the PSEDA individuals. A similar approach has been also adopted in [11], [1].

B. The technique

A PSEDA population is composed by n individuals arranged in a neighborhood topology that defines the inter-individuals influences. Although in PSEDA, as in PSO, various neighborhood topologies can be adopted, in this work we have focused on the fully informed topology where the social attractor of each individual is represented by the population global best position visited so far.

Formally, given the objective/fitness function $f : \Theta \to \mathbb{R}$ (with $\Theta = [l_k, u_k]^d \subseteq \mathbb{R}^d$ and $1 \leq k \leq d$) to minimize, the genotype of a generic PSEDA individual i (with $1 \leq i \leq n$) at a generic generation t (with $t \in \mathbb{N}^+$) is represented by the following d-dimensional vectors:

- its current position at generation t, i.e., $x_{i,t} \in \Theta$,
- its personal best position found so far, i.e., $p_{i,t} = \arg\min_{s \leq t} f(x_{i,s})$,
- its social attractor, i.e., the population global best $g_t = \arg\min_{1 \leq i \leq n} f(p_{i,t})$ that is shared among the whole population,
- the position, personal best, and global best at the previous generation, i.e., respectively, $x_{i,t-1}, p_{i,t-1}, g_{t-1} \in \Theta$.

At each generation t, the genotype of a generic individual i is adopted to build the joint probability distribution $\mathcal{P}_{i,t}$ from which the next individual position is sampled, i.e., $x_{i,t+1} \sim \mathcal{P}_{i,t}$. It is assumed that the d-dimensional probability distribution $\mathcal{P}_{i,t}$ is factorized by a product of d unidimensional and independent densities, i.e., $\mathcal{P}_{i,t} = \prod_{1 \leq k \leq d} \mathcal{P}_{i,t,k}$. Therefore, each dimension k of $x_{i,t+1}$ is independently sampled from $\mathcal{P}_{i,t,k}$, i.e., $x_{i,t+1,k} \sim \mathcal{P}_{i,t,k}$.

Each $\mathcal{P}_{i,t,k}$ is modeled as a weighted finite mixture [20] of the following probability distributions:

- $\mathcal{TN}^{(x)}_{i,t,k}$, i.e., a normal probability distribution, with mean $x_{i,t,k}$ and standard deviation $\sigma^{(x)}_{i,t,k}$, truncated in the search space interval $[l_k, u_k]$,
- $\mathcal{TN}^{(p)}_{i,t,k}$, i.e., a normal probability distribution, with mean $p_{i,t,k}$ and standard deviation $\sigma^{(p)}_{i,t,k}$, truncated in the search space interval $[l_k, u_k]$,
- $\mathcal{TN}^{(g)}_{t,k}$, i.e., a normal probability distribution, with mean $g_{t,k}$ and standard deviation $\sigma^{(g)}_{t,k}$, truncated in the search space interval $[l_k, u_k]$,
- $\mathcal{U}_k$, i.e., a uniform probability distribution on the search space interval $[l_k, u_k]$,
- $\overline{\mathcal{P}}_{i,t-1,k}$, i.e., a relaxed variant of the previous-generation mixture $\mathcal{P}_{i,t-1,k}$ composed by only the three truncated normal distributions $\mathcal{TN}^{(x)}_{i,t-1,k}$, $\mathcal{TN}^{(p)}_{i,t-1,k}$, $\mathcal{TN}^{(g)}_{t-1,k}$ of the previous generation.

These component distributions are respectively weighted by the PSEDA parameters $w_x, w_p, w_g, w_u, w_m \in [0, 1]$ that sum up to 1. Therefore, the sampling procedure for the generic mixture $\mathcal{P}_{i,t,k}$ is divided in two phases [20]. First, the component distribution to sample is chosen by means of a roulette wheel tournament basing on the components weights. Then, the chosen component distribution is sampled. In the case that $\overline{\mathcal{P}}_{i,t-1,k}$ has been chosen in the first phase, the sampling process is repeated using as weights w_x, w_p, w_g normalized in order to sum up to 1. Moreover, for the sake of completeness, we report that a truncated normal distribution can be sampled either by using the accept-reject method, the inverse transform sampling method, or a combination of both [4].

Each one of the mixture's component distributions allows a PSEDA individual to mime the typical PSO swarm-intelligence dynamics by means of probabilistic search mechanisms. Indeed, due to the high probability density around the mean of a normal distribution, $\mathcal{TN}^{(x)}_{i,t,k}$, $\mathcal{TN}^{(p)}_{i,t,k}$, and $\mathcal{TN}^{(g)}_{t,k}$, respectively model: the PSO inertial property to stay in the last visited position, the PSO cognitive tendency to move toward the personal best visited so far, and the PSO social tendency

to move toward the population global best. The relaxed variant of the previous-generation mixture $\overline{\mathcal{P}}_{i,t-1,k}$ tries to reproduce in PSEDA the velocity concept of PSO, while the uniform distribution $\mathcal{U}_k$ has been introduced in order to regulate the exploration/exploitation balance of the evolutionary search.

Although the statistical models adopted are not explicitly multivariate, PSEDA, like PSO, implicitly and hiddenly copes with the inter-dependencies among the dimensions of the fitness landscape by means of its swarm-intelligence based behavior. Moreover, the use of truncated probability distributions allows to completely avoid the out-of-bounds issue that is typical of PSO and of many other evolutionary schemes.

Furthermore, note that the population diversity thus the step-size and the convergence speed of the algorithm highly depend on the standard deviation parameters of the three normal distributions that compose each mixture, i.e., $\sigma^{(x)}_{i,t,k}$, $\sigma^{(p)}_{i,t,k}$, and $\sigma^{(g)}_{t,k}$. In our original paper [18], a common value, that fades during the generations, has been adopted for each individual, dimension and component distribution. Here, we introduce a new technique that automatically adapts the standard deviation parameters depending on the particular component distribution, dimension and individual they belong to. Formally, just before the new position sampling at generation t, $\sigma^{(x)}_{i,t,k}$, $\sigma^{(p)}_{i,t,k}$, $\sigma^{(g)}_{t,k}$ are computed as follows:

$$\sigma^{(x)}_{i,t,k} \leftarrow \begin{cases} \min\{|x_{i,t,k} - p_{i,t,k}|, |x_{i,t,k} - g_{t,k}|\} & \text{if g.t. } 0 \\ \epsilon^{(x)}_{i,t,k} & \text{otherwise} \end{cases} \quad (1)$$

$$\sigma^{(p)}_{i,t,k} \leftarrow \begin{cases} \min\{|p_{i,t,k} - x_{i,t,k}|, |p_{i,t,k} - g_{t,k}|\} & \text{if g.t. } 0 \\ \epsilon^{(p)}_{i,t,k} & \text{otherwise} \end{cases} \quad (2)$$

$$\sigma^{(g)}_{t,k} \leftarrow \begin{cases} \min\{|g_{t,k} - x_{i,t,k}|, |g_{t,k} - p_{i,t,k}|\} & \text{if g.t. } 0 \\ \epsilon^{(g)}_{i,t,k} & \text{otherwise} \end{cases} \quad (3)$$

where $\epsilon^{(x)}_{i,t,k}$, $\epsilon^{(p)}_{i,t,k}$, $\epsilon^{(g)}_{i,t,k}$ are small random numbers uniformly distributed in $[0, 0.01]$. All the equations (1), (2), and (3) rely on the same basic idea. For each one of the three normal distributions of a given dimension and individual, the standard deviation parameter is set to the distance/difference between the mean/peak of the normal itself and the closer mean/peak among the other two normal distributions. In this way, in the case that a normal distribution has been chosen in the sampling procedure, and for the well known properties of normal distributions [21], at least (since the normal distributions are truncated) the 34.1% of chance to be sampled is given to the points located between the peaks/means of the normal distributions. Hopefully, this mechanism should be able to automatically switch the PSEDA search from an explorative to an exploitative behavior and vice versa during the evolution. Obviously, in the case that the three peaks are co-located, the small random movement allows anyway to progress the search in the surround of the last visited position.

Finally, for the sake of clarity the PSEDA pseudo-code is reported in Algorithm 1.

Algorithm 1 PSEDA Pseudo-Code

```
1: procedure PSEDA
2:     t ← 0
3:     Randomly initialize the n individuals
4:     while t < M do                    ▷ M is the allowed number of generations
5:         for all individuals i do
6:             Evaluate fitness f(x_{i,t})
7:             Save the previous personal best p_{i,t-1}
8:             Update personal best p_{i,t}
9:         end for
10:        Save the previous global best g_{t-1}
11:        Update global best g_t
12:        for all individuals i do
13:            for all dimensions k do
14:                Compute σ^(x)_{i,t,k}, σ^(p)_{i,t,k}, σ^(g)_{t,k} according to equations (1), (2), (3)
15:                Save the previous position x_{i,t-1}
16:                Sample x_{i,t+1,k} from P_{i,t,k}
17:            end for
18:        end for
19:        t ← t + 1
20:    end while
21: end procedure
```

III. Lymphoma Classification through Biopsy Image Analysis

A. The System Architecture

The aim of this work is to automatically classify a lymph-nodes biopsy image in one of three types of malignant lymphoma, i.e., "chronic lymphocytic leukemia" (CLL), "follicular lymphoma" (FL), and "mantle cell lymphoma" (MCL). Each biopsy image has been acquired through microscope and, before the acquisition, the lymph-nodes tissues have been sectioned and stained with Hematoxylin/Eosin (H+E) as described in [19]. Randomly selected images from each lymphoma class are reported in Figures 1, 2, 3.

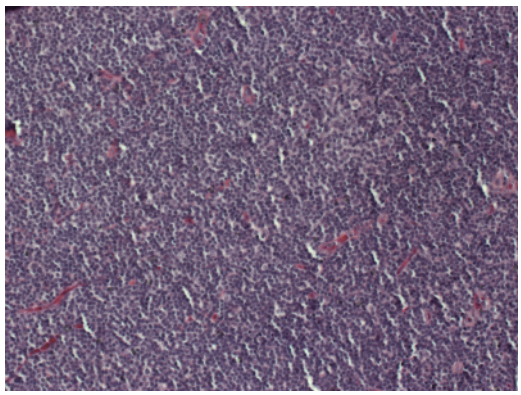

Fig. 1: Example of biopsy image from class CLL

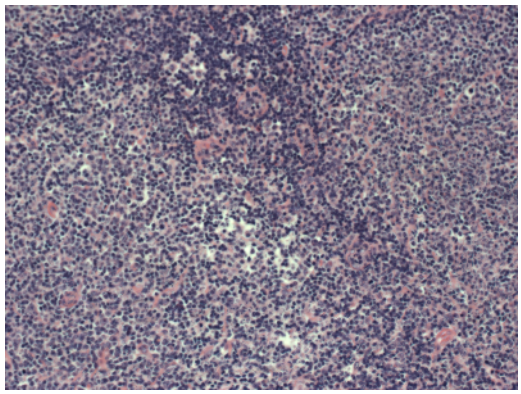

Fig. 2: Example of biopsy image from class FL

The automatic classification is performed following the macro-architecture reported in Figure 4. First, each biopsy

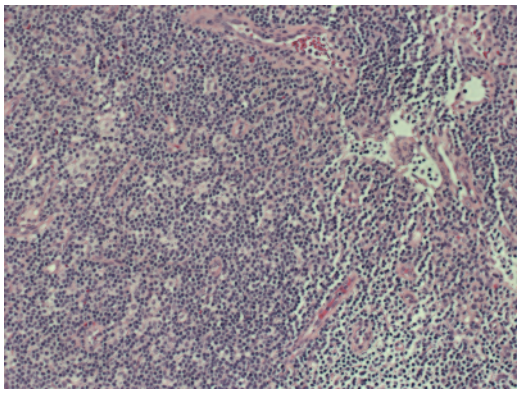

Fig. 3: Example of biopsy image from class MCL

image is preprocessed and then converted to a low-dimensional numerical vector through a process of features extraction. Then, a set of already classified images (the training set) allows the system to compute its internal parameters, thus to train a classification scheme that is employed in order to predict the class label of each incoming test image.

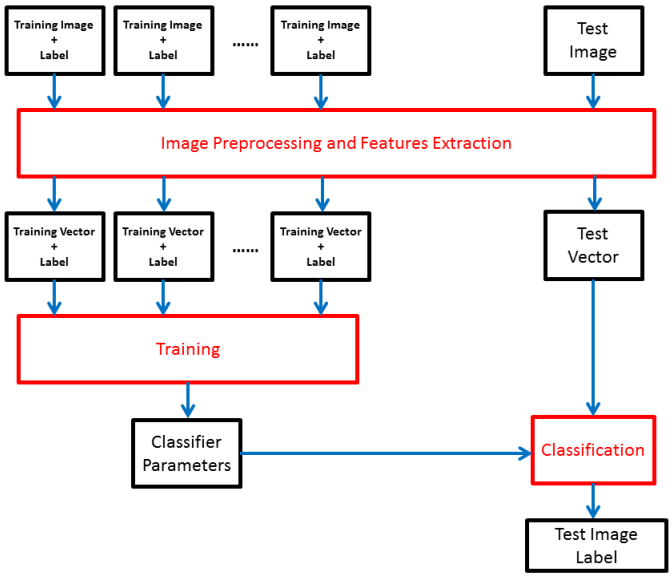

Fig. 4: System Architecture

B. Image Preprocessing and Features Extraction

The only image preprocessing step adopted is the conversion of the original image to a gray-scale format. In this way, the numerical features are extracted on an image composed by only one color channel.

Regarding the features extraction, we have employed two different kinds of features, that is, color and geometric features. The 3 color features considered are the intensity average, standard deviation, and skewness, that have been already employed in various other works (see for example [10]). Instead, the geometric features considered are the 7 Hu moments that have been extracted using the specific function included in the OpenCV library [17]. These kind of geometric features have been widely adopted in literature (see for example [23]), mainly for their invariance with respect to the image scale, rotation, and reflection.

Finally, note that, after the features extraction, a 10-dimensional numerical vector is made for each image. This low dimensionality allows to speed-up the classification step and to avoids the "curse of dimensionality" issue, observed in various classification technique [16], as well.

C. Classification through PSEDA

The classification of a *vectorized* image $a \in \mathbb{R}^{10}$ is performed through the following prediction function:

$$h_{x^{(CLL)},x^{(FL)},x^{(MCL)}}(a) = \underset{c\in\{CLL,FL,MCL\}}{\arg\min} d(x^{(c)}, a) \quad (4)$$

where a generic $x^{(c)} \in \mathbb{R}^{10}$ is a representative of the class $c \in \{CLL, FL, MCL\}$ in the 10-dimensional space induced by the features extraction process, while $d(\cdot,\cdot)$ is the usual Euclidean distance. Essentially, $h : \mathbb{R}^{10} \to \{CLL, FL, MCL\}$ assigns, to the vectorized image taken in input, the label of the closest class representative. Note that, due to the low dimensionality (10) and the low number of classes (3), the computational time of this function is almost negligible.

A training phase is needed in order to obtain a good setting for the parameters of the prediction function h, i.e., the classes representatives $x^{(CLL)}, x^{(FL)}, x^{(MCL)}$. However, in our approach, the training task can be easily cast to an optimization problem. Indeed, since each class representative is a point in $\mathbb{R}^{10}$, the learning of the prediction function parameters can be seen as the problem of finding the optimal position for each class representative in $\mathbb{R}^{10}$ given the training set T of already labeled images. Thus, the optimization search space Θ where to apply PSEDA (or other numerical optimization techniques) is a 30-dimensional space (3 classes $\times$ 10 numerical features), i.e., $\Theta \subseteq \mathbb{R}^{30}$, and a generic PSEDA candidate solution $x \in \Theta$ is composed by 3 10-dimensional blocks that encode the 3 class representatives, i.e., $x = \langle x^{(CLL)}, x^{(FL)}, x^{(MCL)} \rangle$ where $x^{(c)} \in \mathbb{R}^{10}$ for $c \in \{CLL, FL, MCL\}$.

In order to evaluate the quality of a generic solution x on the training set T, the two fitness functions adopted in [3] have been taken into account:

$$\psi_1(x) = \frac{1}{|T|} \sum_{(a_i,c_i)\in T} \delta(x, a_i, c_i) \quad (5)$$

$$\psi_2(x) = \frac{1}{|T|} \sum_{(a_i,c_i)\in T} \overline{d}(a_i, x^{(c_i)}) \quad (6)$$

where $c_i \in \{CLL, FL, MCL\}$ is the class label of the training image $a_i \in \mathbb{R}^{10}$, $\overline{d} : \mathbb{R}^{10} \times \mathbb{R}^{10} \to [0, 1]$ is the Euclidean distance normalized to range in $[0, 1]$[1], and

$$\delta(x, a_i, c_i) = \begin{cases} 1 & \text{if } h_{x^{(CLL)},x^{(FL)},x^{(MCL)}}(a_i) \neq c_i \\ 0 & \text{otherwise} \end{cases}$$

Both the fitness functions range in $[0, 1]$ and have to be minimized. ψ_1 represents the error rate on the training set,

[1]The euclidean distance is normalized as in [3], that is, each distance component is normalized with respect to the maximal range in the dimension and the sum of distance components is divided by the dimensionality m.

while ψ_2 computes the average distance of training instances from the class representative of the known classes. ψ_1 is a step function that can vary with "jumps" equal to $1/|T|$ only, while ψ_2 it is hoped to provide a greater continuity, although it can introduce large plateaus in the search space. Moreover, ψ_2 has a slightly lower complexity than ψ_1 since each training instance is compared with only one class representative instead of all the three.

Finally, a third fitness function ψ_3 is introduced in order to combine the characteristic of both ψ_1 and ψ_2:

$$\psi_3(x) = 0.75 \cdot \psi_1(x) + 0.25 \cdot \psi_2(x) \quad (7)$$

This function is a weighted convex combination of the previous, thus it still ranges in $[0, 1]$. The weights unbalanced toward ψ_1 are motivated by some preliminary experiments that we have conducted.

As a consequence of the choice of three different fitness functions, three PSEDA schemes, differing each other for the fitness function adopted, have been designed: PSEDA-ψ_1, PSEDA-ψ_2, and PSEDA-ψ_3.

IV. Experiments

The accuracy and the performances of the proposed automatic classification system have been evaluated using the biopsies images dataset proposed in [19]. The dataset is composed by 374 color images, each one with a resolution of 1388×1040 pixels and a depth of 32 bit. Each image represents a lymph-nodes tissues picture acquired using a brightfield microscope and following the method briefly described in Section III-B. 113 images are labeled as CLL, 139 as FL, and 122 as MCL. Every image has been preprocessed using the techniques described in Section III-B, therefore, classification schemes are applied on the 10-dimensional numerical vectors that describe the original images.

The PSEDA parameters settings has been chosen after some preliminary experiments and following the same methodology proposed in our original paper [18]. Therefore, we have chosen a population size of 50, and the following weights: $w_u = 0.01$, $w_x = w_m = 0.09705$, $w_p = w_g = 0.39795$. PSO, as well, has been executed using the setting adopted in [3] for other classification problems. Using the same notation of [3]: $n = 50$, $T_{max} = 1000$, $v_{max} = 0.05$, $v_{min} = -0.05$, $c_1 = 2.0$, $c_2 = 2.0$, $w_{max} = 0.9$, $w_{min} = 0.4$. Both PSEDA and PSO have been executed using the three fitness functions described in Section III-C. Furthermore, since we have adopted the software suite Weka (release 3.6) [7] in order to perform the classification (and the training) task, we have compared the two evolutionary techniques with two state-of-the-art classification schemes: a rule-based one, i.e., "Ripple Down Rule learner" (Ridor) [2], and a tree-based one, i.e., "Naive Bayes Tree" (NBTree) [14]. In these cases, the default parameters settings of Weka have been employed. Summarizing, we have executed a total of 8 classification schemes: PSEDA-ψ_1, PSEDA-ψ_2, PSEDA-ψ_3, PSO-ψ_1, PSO-ψ_2, PSO-ψ_3, Ridor, and NBTree.

These classification schemes have been evaluated using the stratified 10-folds cross-validation technique repeated for 10 times. This technique is commonly considered the standard method for classification performances evaluation [22]. It acts by performing 10 rounds of 10-folds cross-validation. For each round, the given dataset is randomly split in 10 equal-size and stratified folds (that is, each fold has approximately the same classes distribution of the entire dataset). Each classification scheme is trained using the instances belonging to 9 of these folds (i.e., the training set) and then evaluated on the remaining fold (i.e., the test set). In each round, this process is repeated for each of the 10 different combinations of training/test set. Finally, the classification performances are averaged among the 10 training/test set combinations and the 10 cross-validation rounds, resulting in a total of 100 executions for each classification scheme.

In Table I, for each classification scheme, the average, standard deviation, maximum and minimum percentual accuracies are reported.

TABLE I: Experimental Results

Scheme	Acc. Avg	Acc. StdDev	Acc. Max	Acc. Min
PSEDA-ψ_1	61.02	6.46	78.68	44.10
PSEDA-ψ_2	50.06	7.63	68.42	32.43
PSEDA-ψ_3	**63.92**	**6.36**	**80.32**	**48.32**
PSO-ψ_1	56.88	6.46	75.68	42.10
PSO-ψ_2	49.23	8.50	68.42	31.58
PSO-ψ_3	58.82	**6.36**	76.32	44.74
Ridor	59.24	8.29	76.32	37.84
NBTree	60.32	8.27	78.94	40.54

These results clearly show that PSEDA-ψ_3 outperforms the other schemes. Moreover, also a statistical significance test among PSEDA-ψ_3 and the other schemes has been conducted. The paired t-test with confidence 0.05 confirms the superiority of PSEDA-ψ_3 that is significantly better than all other schemes except PSEDA-ψ_1. By remembering that the fitness function ψ_3 is a weighted combination of ψ_1 and ψ_2 unbalanced toward ψ_1 (see equation (7)), it results clear why PSEDA-ψ_1 is not significantly worse than PSEDA-ψ_3. It is noticeable also the maximum accuracy achieved by PSEDA-ψ_3 that is the only one scheme able to reach the 80% of accuracy in at least a run. Furthermore, note that the evolutionary schemes (PSEDA and PSO) present a less accuracy standard deviation with respect to Ridor and NBTree. This can be clearly interpreted as a sign of robustness of the fitness functions employed.

Finally, in Figure 5, a comparison of the convergence graphs of PSEDA-ψ_3 and PSO-ψ_3 is provided. The graphs are averaged over the 100 executions performed by the 10-folds cross-validation repeated 10 times and show the improvement in fitness observed during the evolution. Figure 5 shows that also on the training set, PSEDA outperforms PSO. Moreover, although slowly in the second part of the evolution, both the meta-heuristic continue to improve their fitness value until the end of the evolution. This aspect can be considered as a further indicator of the goodness of the designed fitness function.

V. Conclusion and Future Work

In this paper, a system for lymphoma classification through automatic biopsy images analysis has been proposed. The system is composed by two main modules: a computer vision module that elaborates the tissue biopsy images in order to extract some numerical features, and a machine learning module that, basing on already labeled biopsy images, builds

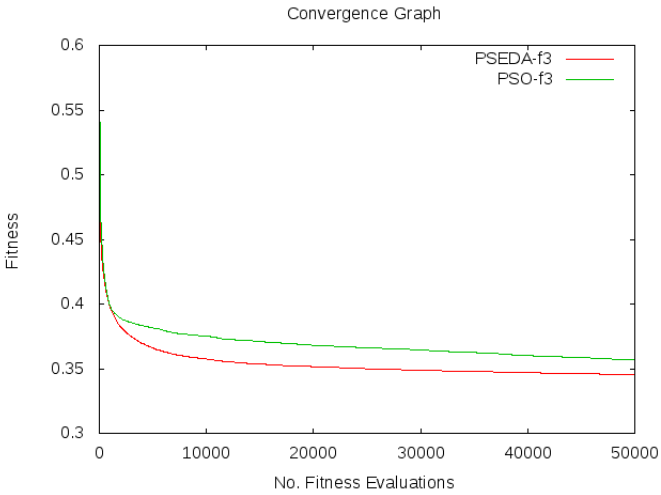

Fig. 5: Convergence Graphs for PSEDA-ψ_3 and PSO-ψ_3

a classifier for the automatic prediction of new and unlabeled biopsy images coming in the system.

A low number of numerical features has been considered in order to speed-up the computational time, thus to augment the responsiveness of the system, and to avoid the "curse of dimensionality" issue as well.

The classification and training tasks have been performed through the use of a recently introduced meta-heuristic technique, i.e., Particle Swarm Estimation of Distribution Algorithm (PSEDA). PSEDA consists in an hybridization of PSO and EDAs schemes and can be view as the probabilistic "alter-ego" of PSO. The numerical space induced by the features extraction task has been adopted as the search space for PSEDA. Each PSEDA candidate solution is mapped to a classifier/prediction function and the quality of the classification is evaluated by introducing three different fitness functions.

This results in the first application of PSEDA to classification problems, and in particular to images classification problems.

A standard and publicly available dataset has been adopted in order to test the system [19]. The experimental results provide two main conclusions.

Firstly, a low number of image features is enough to obtain a good accuracy level on the classification task. As a consequence, the computational time are drastically improved with respect to other systems that adopt a number of image features in the order of hundreds or thousands (see for example [19]). Moreover, the low dimensionality of the "vectorized" images allows to avoid the "curse of dimensionality" issue that is typical of the large majority of machine learning techniques.

Secondly, PSEDA, although originally introduced for numerical optimization, can be easily adopted for the training of a classification system. Experiments conducted show that it outperforms both some state-of-the-art classification schemes and, primarily, the swarm-intelligence technique from which it derives, i.e., PSO.

As future work, we are intentioned to study some other image features, like for example texture-based features that have not been considered in this work. Furthermore, PSEDA, thank to the use of different and independent probabilistic models for each dimension of the search space, can be adopted in hybrid discrete/continuous search spaces by simply considering discrete/continuous probabilistic models. We think that this characteristic could be important for those classification problems that consider also nominal and discrete attributes.

REFERENCES

[1] T. Blackwell. A study of collapse in bare bones particle swarm optimization. *Evolutionary Computation, IEEE Transactions on*, 16(3):354–372, 2012.

[2] P. Compton and R. Jansen. Knowledge in context: A strategy for expert system maintenance. In ChristopherJ. Barter and MichaelJ. Brooks, editors, *AI '88*, volume 406 of *Lecture Notes in Computer Science*, pages 292–306. Springer Berlin Heidelberg, 1990.

[3] I. De Falco, A. Della Cioppa, and E. Tarantino. Facing classification problems with particle swarm optimization. *Applied Soft Computing*, 7(3):652–658, 2007.

[4] L. Devroye. *Non-uniform random variate generation*, volume 4. Springer-Verlag New York, 1986.

[5] M. El-Abd and M. Kamel. Pso_bounds: A new hybridization technique of pso and edas. *Foundations of Computational Intelligence Volume 3*, pages 509–526, 2009.

[6] C. González, J.A. Lozano, and P. Larranaga. Mathematical modelling of umdac algorithm with tournament selection. behaviour on linear and quadratic functions. *International Journal of Approximate Reasoning*, 31(3):313–340, 2002.

[7] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.

[8] G.R. Harik, F.G. Lobo, and D.E. Goldberg. The compact genetic algorithm. *IEEE Transactions on Evolutionary Computation*, 3(4):287–297, 1999.

[9] M. Iqbal and M. de Oca. An estimation of distribution particle swarm optimization algorithm. *Ant Colony Optimization and Swarm Intelligence*, pages 72–83, 2006.

[10] HB Kekre, S.D. Thepade, T.K. Sarode, and S.P. Sanas. Image retrieval using texture features extracted using lbg, kpe, kfcg, kmcg, kevr with assorted color spaces. *International Journal of Advances in Engineering & Technology*, pages 2231–1963, 2012.

[11] J. Kennedy. Bare bones particle swarms. In *Swarm Intelligence Symposium, 2003. SIS'03. Proceedings of the 2003 IEEE*, pages 80–87. IEEE, 2003.

[12] J. Kennedy. Swarm intelligence. *Handbook of Nature-Inspired and Innovative Computing*, pages 187–219, 2006.

[13] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proc. of IEEE International Conference on Neural Networks, 1995.*, volume 4, pages 1942–1948. IEEE, 1995.

[14] R. Kohavi. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In *Proceedings of the second international conference on knowledge discovery and data mining*, volume 7, 1996.

[15] P. Larrañaga and J.A. Lozano. *Estimation of distribution algorithms: A new tool for evolutionary computation*, volume 2. Springer Netherlands, 2002.

[16] T.M. Mitchell. *Machine learning. 1997*. Burr Ridge, IL: McGraw Hill, 1997.

[17] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov. Real-time computer vision with opencv. *Communications of the ACM*, 55(6):61–69, 2012.

[18] V. Santucci and A. Milani. Particle swarm optimization in the edas framework. In *Soft Computing in Industrial Applications*, volume 96, pages 87–96. Springer, 2011.

[19] Lior Shamir, Nikita Orlov, David Mark Eckley, TomaszJ. Macura, and IlyaG. Goldberg. Iicbu 2008: a proposed benchmark suite for biological image analysis. *Medical & Biological Engineering & Computing*, 46:943–947, 2008.

[20] D.M. Titterington, A.F.M. Smith, U.E. Makov, et al. *Statistical analysis of finite mixture distributions*, volume 38. Wiley New York, 1985.

[21] H.M. Wadsworth. *Handbook of statistical methods for engineers and scientists*. McGraw-Hill Professional, 1998.

[22] I.H. Witten, E. Frank, and M.A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2011.

[23] J. Wu and S. Xiong. Research of web image retrieval technology based on hu invariant moments. *Advances in Swarm Intelligence*, pages 66–73, 2012.

[24] Y. Zhou and J. Jin. Eda-pso: a new hybrid intelligent optimization algorithm. In *Proc. of the Michigan University Graduate Student Symposium*, 2006.