Journal of Computer and Communications, 2015, 3, 184-190
Published Online November 2015 in SciRes. http://www.scirp.org/journal/jcc
http://dx.doi.org/10.4236/jcc.2015.311029

Semi Advised SVM with Adaptive Differential Evolution Based Feature Selection for Skin Cancer Diagnosis

Ammara Masood, Adel Al-Jumaily

School of Electrical, Mechanical, Mechatronic Systems, University of Technology, Sydney, Australia
Email: ammara.masood@student.uts.edu.au, Adel.Al-Jumaily@uts.edu.au

Received October 2015

Abstract

Automated diagnosis of skin cancer is an important area of research that had different automated learning methods proposed so far. However, models based on insufficient labeled training data can badly influence the diagnosis results if there is no advising and semi supervising capability in the model to add unlabeled data in the training set to get sufficient information. This paper proposes a semi-advised support vector machine based classification algorithm that can be trained using labeled data together with abundant unlabeled data. Adaptive differential evolution based algorithm is used for feature selection. For experimental analysis two type of skin cancer datasets are used, one is based on digital dermoscopic images and other is based on histopathological images. The proposed model provided quite convincing results on both the datasets, when compared with respective state-of-the art methods used for feature selection and classification phase.

Keywords

Classification, Feature Selection, Skin Cancer, Support Vector Machine, Differential Evolution

1. Introduction

Malignant melanoma is one of the most dangerous forms of skin cancer. Melanoma cases are recorded in big numbers over the last few decades [1]. In 2014, in Australia an estimated 128,000 new cases of cancer were diagnosed and the number may rise to 150,000 by 2020 [2].

Traditionally, in the skin cancer diagnosis process, dermoscopic images are used by dermatologists while pathologists use histopathological images of biopsy samples taken from patients and examine them using microscope. However, all the analysis and judgments depend on personal experience and expertise and often lead to considerable variability [3]. The biggest challenges in developing automated diagnostic tools is selection of distinguishing quantitative features and development of an efficient classification algorithm that can be trained using both labeled and unlabeled data due to the limited availability of related skin cancer datasets.

Due to the complex nature of skin cancer images specially the histopathological images [4], it is suggested to have a good variety of differentiating features to begin with, and then use an efficient feature selection method for removing redundant features for reducing amount of data for classifier learning.

How to cite this paper: Masood, A. and Al-Jumaily, A. (2015) Semi Advised SVM with Adaptive Differential Evolution Based Feature Selection for Skin Cancer Diagnosis. *Journal of Computer and Communications*, **3**, 184-190.
http://dx.doi.org/10.4236/jcc.2015.311029

Searching for the optimal feature subset, which can result in best training as well as testing performance, is a challenging task. Various studies show that DE has outperformed many other optimization algorithms in terms of robustness over common benchmark problems and real world applications [5] [6]. This paper proposes a new adaptive differential evolution algorithm based search strategy for feature selection that showed promising results for skin cancer diagnosis both for dermatological as well as histopathological image datasets. It should be noted that the control parameters and learning strategies involved in DE are highly dependent on the problem under consideration and need to be adjusted more adaptively [7]. Thus, one of the research objectives was to adaptively adjust the related control parameter and select the feature subset simultaneously for the problem under consideration, without degrading the classification accuracy.

On the other hand, for the evaluation of selected feature sets, there are various classification/learning methods proposed in literature [8] [9]. However, as we discussed in [10] more research is required that should take into consideration the experts' advice using the labeled data along with the capability of using unlabeled data due to the high costs and time involved in getting labeled datasets. Unfortunately, this is also a case in developing skin cancer diagnosis models as obtaining the labeled datasets from experts is not a trivial task and it cost a lot of money and is quite time consuming. To solve this issue, developing semi supervised learning methods that can use unlabeled data together with labeled data to build better learners, is an important area that needs consideration. Typical semi-supervised methods used for different applications include Expectation-maximization (EM) algorithm with generative mixture models [11], transductive support vector machines [12] and graph-based methods [13] etc. In addition to using unlabeled data for training it is also of utmost importance that learning process should have self-advising and self-correcting capability to deal with the misclassified data to avoid its effect on the diagnostic performance of the learning model. The proposed semi advised support vector machine is meant to deals with both of the issues mentioned above.

The paper is organized as follows: Section 2 provides the details of the adaptive differential evolution algorithm proposed for feature selection and the semi-advised support vector machine algorithm proposed for classification. Section 3 provides the overview of the experimental model based on the proposed algorithms and presents the experimental results and finally conclusion is given in Section 4.

2. Proposed Algorithms

2.1. Adaptive Differential Evolution Based Feature Selection

Differential evolution (DE) is a population based optimization method, which has attracted an increased attention in the past few years. Although it showed quite promising results in various applications but in complex applications the search performance get highly depended on the mutation strategy, crossover operation and control factors including scale factor (F), Cross over rate (Cr) and population size (NP) [7] [14].

The paper proposes a DE-based feature selection technique with an adaptive approach to make the feature selection process more dynamic to be applied for complex pattern recognition applications like the histopathological image analysis. It will use advised support vector machine explained in following section for evaluation of selected feature subset. The steps of the feature selection procedure are as follows.

1) Initialize the population of *NP* individuals $\text{Pop}_G = \{ \vec{X}\_1_G , \cdots, \vec{X}\_NP_G \}$ where $\vec{X}\_i_G = [\text{x1\_i}_G, \text{x2\_i}_G, \text{x3\_i}_G, \cdots, \text{xD\_i}_G]$, with i = [1, 2, ⋯, NP] where D is the number of parameter to be optimized.

2) Set the mutation parameter (F) and cross over control parameter (Cr) using the following equations

$F_i = \text{Cauchy}(F_m, 0.1)$ with $F_m = (w_F.F_m) + ((1 - w_F)F_m\_\text{best})$ and

$w_F = 0.8 + 0.2 \times \text{rand}(0,1) Cr_i = \text{Gaussian}(Cr_m, 0.1)$ with $Cr_m = (w_{Cr}.Cr_m)) + ((1 - w_{Cr})Cr_m\_\text{best})$ and

$w_{Cr} = 0.9 + 0.1 \times \text{rand}(0,1)$. Note F_m is initialized with value of 0.5 while Cauchy distribution prevent premature convergence due to its wider tail property. While F_m_best is the most successful scale factor in the current generation. While Cr_m is initialized with vlaue of 0.6 and Gaussian distribution is used as opposite to Cauchy distribution its short tail property help in keeping the value of Cr within unity [7] which is also required here. Cr_m_best is the successful crossover probability in the current generation .

3) While the termination criterion (maximum number of iterations) is not satisfied

Do **for** i = 1 to NP //do for each individual

a. Perform Mutation: A mutant vector $\vec{V}\_i_G = \{\text{v1\_i}_G, \cdots, \text{vD\_i}_G\}$ is created corresponding to the ith target

vector $\vec{X}\_i_G$ by merging three different randomly selected vectors *i.e.* using the DE/rand/1 Mutation strategy.

$$\vec{V}\_i_G = \overrightarrow{X_1}\_i_G + F_i \cdot (\overrightarrow{X_2}\_i_G - \overrightarrow{X_3}\_i_G) \tag{1}$$

b. Crossover operation : Employ binomial crossover on each of the D variable as follows for building trial vector

$$uj\_i_G = \begin{cases} vj\_i_G & if\ (randi, j[0,1] \le Cr_i\ or\ j = j_{rand} \\ xj\_i_G & otherwise \end{cases} \tag{2}$$

Here $j_{rand} \in [1, 2, \cdots, D]$ is a randomly selected index to ensure that $\vec{U}\_i_G$ gets at least some component from $\vec{V}\_i_G$.

c. Evaluate the population with the objective function.

d. Perform Selection: Evaluate the trial vector $\vec{U}\_i_G$ with the fitness function f = accuracy of classifier

If $f(\vec{U}\_i_G) \ge f(\vec{X}\_i_G)$, then $\vec{X}\_i_{G+1} = \vec{U}\_i_G$

Else $\vec{X}_{i_{G+1}} = \vec{X}_{i_G}$ end if end **for**

4) Repeat from step 2 - 3 until G_{max}

2.2. Semi Advised SVM

For using automated learning techniques in any area in order to improve performance requires a proper choice of the learning algorithm and of their statistical validation. Classifier training with insufficient number of labeled data is a well-known hard problem [15] [16]. Development of computer aided diagnostic models for skin cancer is a difficult problem given the relative paucity of labeled lesion data and consequently the training data available is not of high quality [17]. The proposed algorithm addresses the skin lesion classification problem with training the classifier using unlabeled data by making efficient use of limited labeled data.

In this paper, a semi-advising algorithm for SVM is proposed that extracts subsequent knowledge during the training phase using both labeled and sets of unlabeled data that is added in a batch-mode. The effect of misclassified data during the training phase is controlled by generating advice weights [18] based on using misclassified training data. The details of semi-advised SVM algorithm are as follows:

Given the dataset is divided into two subsets: a labeled data set $D_L = \{(x_i, y_i)\}_{i=1}^{L}$ and an unlabeled data set

$$D_{UL} = \{(x_i)\}_{i=L+1}^{L+UL}$$

Step 1: The unlabeled data set D_{UL} is equally divided into n subsets $D_{UL1}, D_{UL2}, \cdots, D_{ULn}$, then D_L is taken as the initial training set T_s and initialize i = 1, where i denotes the i^{th} loop of the algorithm.

Step 2: SVM classifier is trained using labeled data & classifying hyperplane is found using decision function

$$f(x) = sign(\sum_{\alpha_{i>0}} y_i \alpha_i k(x, x_i) + b) \tag{3}$$

where x_i is input vector corresponding to the i^{th} sample and is labeled by y_i depending on its class, b is constant and α_i is the nonnegative Lagrange multiplier that is inconsistence with standard SVM training.

As the data is comprised of nonlinearly separable cases so kernel based SVM is used to produce non-linear decision functions and radial basis function kernel (RBF)

$$K(x_i x_j) = e^{-\gamma |x_i - x_j|^2} \tag{4}$$

is used to make all necessary operations in the input space.

Step 3: The misclassified data sets (MD) in the training phase is determined using following relationship.

$$MD = \bigcup_{i=1}^{N} x_i \mid y_i \ne sign(\sum_{\alpha_j>0} y_i \alpha_j k(x_i, x_j) + b) \tag{5}$$

The MD set can be null, but most experiments showed that the occurrence of misclassified data in training

phase is a common occurrence. It must also be noted that any method that tries to get benefit from misclassified data, must also have some control on the impact of outlier data. We observed that when the misclassified data is comprised of resembling samples, the use of misclassified data actually improved the classification accuracy more as it can lead to the variations required in the final separating hyperplane.

If the MD is null, go to the next step, else compute neighbourhood length (NL) for each member of MD using the following mathematical relation, which is then used during advised weight calculation.

$$NL(x_i) = minimum_{x_j}(\|x_i - x_j\| \mid y_i \neq y_j) \tag{6}$$

where x_j, j = 1, ⋯, N are the training data that do not belong to the MD set. Here as the training data is mapped to a higher dimension, the distance between x_i and x_j is computed according to the following equation with reference to the related RBF kernel.

$$\|\theta(x_i) - \theta(x_j)\| = (k(x_i, x_i) + k(x_j, x_j) - 2k((x_i, x_j))^{0.5} \tag{7}$$

Step 4: The labels for data samples in D_{UL1} are estimated using current classifier, and then the most confidently classified elements are determined according to the distance between the element and the separating boundary. The criteria is formulated as $|x \cdot w - b\} \geq Th$, where constant $Th > 0$ is the distance threshold. If distance between element and separating boundary is larger than Th, we take it as confident element. The most confidently classified elements with their predicted labels are represented as set R and are added with their predicted labels, to training set T_s, *i.e.*, $T_s = T_s \cup R$. Remaining elements of D_{UL1} are denoted as unlabeled query UL_Q_i.

For each sample x_k from the unlabelled Query set UL_Q_i advised weight $AW(x_k)$ is computed using following mathematical relationship. These AWs represent how close data is to the misclassified data from the labelled set.

$$\begin{cases} 0 \quad \forall x_i \in MD, \|x_k - x_i\| > NL(x_i) \text{ or } MD = NUL, \\ \sum 1 - \dfrac{\sum_{x_i} \|x_k - x_i\|}{\sum_{x_i} NL(x_i)} x_i \in MD, \ \|x_k - x_i\| \leq NL(x_i) \end{cases} \tag{8}$$

The absolute value of the SVM decision values for each x_k from the unlabeled Query set set are calculated and scaled to [0, 1]. For each x_k from unlabelled Query set, if (AW (x_k) < decision value (x_k) then $y_k = \text{sign}(\sum_{\alpha_j > 0} y_j \alpha_j k(x_k, x_j) + b)$ which is in consistence with normal SVM labelling, otherwise $y_k = y_i \mid (\|x_k - x_i\| \leq NL(x_i) \text{ and } x_i \in MD)$. After getting labels of UL_Q_i add UL_Q_i with predicted labels to T, that is, $T = T \cup Qi$.

Step 5: i = i + 1

Step 6: If i equals n, terminate; otherwise, go back to Step 2.

3. Experimental Analysis

3.1. Overall Learning Model

The proposed model is presented in **Figure 1**. In order avoid the domination of features in greater numeric ranges on the ones with smaller numeric ranges, both the datasets under consideration are linearly scaled to the range [−1, +1] or [0, 1]. The optimization of feature subsets and control parameters is done based on the adaptive differential evolution algorithm explained in the previous section. Using the selected feature sets, the training sets are fed into classification stage where the semi advised SVM classifier explained in the previous section is used. Once the trained model is obtained it is then used for classification of the test data.

3.2. Experimental Results

Two datasets were used in the experiments, dataset 1 is based on dermoscopic images and dataset 2 is based on histopathological images obtained from the biopsy samples of skin cancer patients. Most of the images in the datasets came from Sydney Melanoma Diagnostic Centre. Dataset one comprise of 300 labeled and 500 unlabeled images. While the Dataset 2 consists of 160 images including 60 labeled and 100 unlabeled samples.

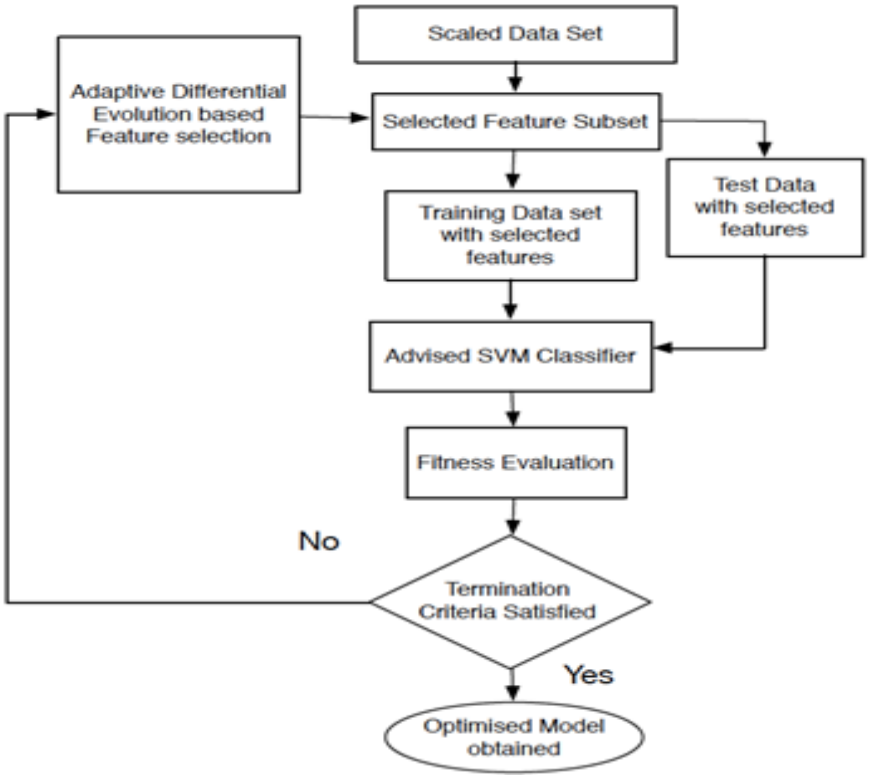

Figure 1. Proposed learning model.

For testing the effect of feature selection method and the number of selected features on the overall performance of the model, the performance of the proposed feature selection method is also compared with the ones based on well-established binary Genetic Algorithm BGA [19], Binary PSO (BPSO) [20] and Standard Differential evolution based feature selection [21] (see **Figure 2**). All methods were made to start from the same initial population with the population size set to 50 and the evolution process is terminated at the same number of iterations set to 100. The fitness function used for evaluation was the classification accuracy. It can be seen that the proposed method attained comparable or better classification accuracies (using the proposed semi-advised SVM classifier) as compared to other methods for comparatively lesser number of selected features.

This shows that if parameter tuning and feature selection is done simultaneously and effect of misclassified data/outliers is minimized, it can improve the classification performance of the learning models. In addition to that, it can also help in minimizing the use of redundant/irrelevant features in the final optimized model, which will make system computationally less complex and will also decrease the chances of having over fitted models.

10 fold cross validation rule is used to validate the performance of the overall model for both datasets. We also compared classification performance of proposed classification algorithm with SVM and T-SVM. **Figure 3** shows average classification error rate of different classifiers with respect to the change in the ratio of labeled and unlabeled data samples used for training phase. In consistent with a lot of finding in different other applications [11] it was observed that by increasing the number of labeled data in the training phase helps in reducing the classification error. The classification error reduced to around 16.5% for Histopathological images and 6% for Dermoscopic images when the learning model used 50% of labeled and 50% of unlabeled data.

4. Conclusion

This paper presents a novel learning model with adaptive differential evolution based feature selection and semi advised support vector machine based classification. The proposed feature selection method is meant to adaptively adjust the tuning parameter for the differential evolution process and do the feature selection for the corresponding dataset simultaneously. On the other hand, the proposed semi advised SVM is trained using labeled data along with adding sets of unlabeled data to deal with the misclassified data elements and improve the gene-

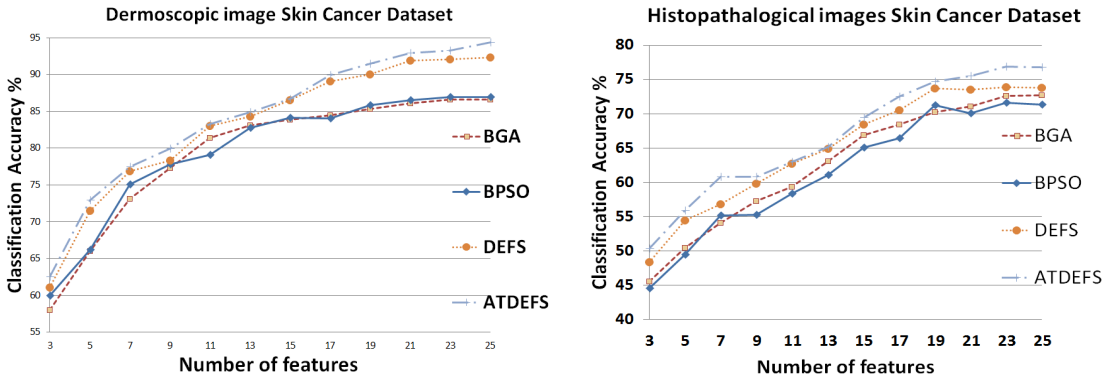

Figure 2. Average classification accuracies vs. feature subset sizes.

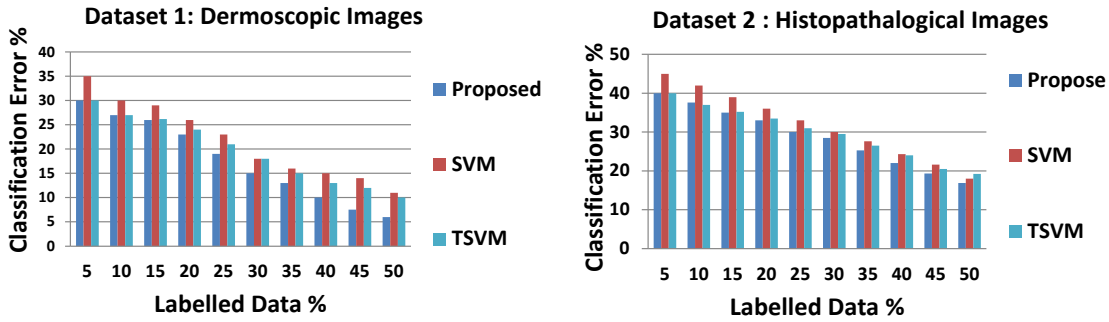

Figure 3. Classification error rate of different classifiers with increasing percentage of labeled training data.

ralization performance of the classifier by using increased amount of training data. Experimental analysis shows that the proposed learning model works well and provides an optimal feature set with higher classification rate when compared with some other popular methods used in literature. The efficient use of unlabeled data with the aid of limited labeled dataset helped in obtaining better generalization of the model over the test data and obtained accuracy of around 94% for dermoscopic images and 86.5% for histopathological images.

References

[1] Siegel, R., Naishadham, D. and Jemal, A. (2012) CA: A Cancer Journal for Clinicians. *Cancer Statistics*, **62**, 10-29.

[2] Australian Institute of Health and Welfare (2014) In ACIM (Australian Cancer Incidence and Mortality) Books.

[3] Preti, M., *et al.* (2000) Inter-Observer Variation in Histopathological Diagnosis and Grading of Vulvar Intraepithelial Neoplasia: Results of an European Collaborative Study. *BJOG*: *An International Journal of Obstetrics & Gynaecology*, **107**, 594-599. http://dx.doi.org/10.1111/j.1471-0528.2000.tb13298.x

[4] Verhaegen, P.D., van Zuijlen, P.P., Pennings, N.M., van Marle, J., Niessen, F.B., van der Horst, C.M. and Middelkoop, E. (2009) Differences in Collagen Architecture between Keloid, Hypertrophic Scar, Normotrophic Scar, and Normal Skin: An Objective Histopathological Analysis. *Wound Repair and Regeneration*, **17**, 649-656. http://dx.doi.org/10.1111/j.1524-475X.2009.00533.x

[5] Khushaba, R.N., Ahmed, A.-A. and Al-Jumaily, A. (2011) Feature Subset Selection Using Differential Evolution and a Statistical Repair Mechanism. *Expert Systems with Applications*, **38**, 11515-11526. http://dx.doi.org/10.1016/j.eswa.2011.03.028

[6] Islam, S.M.D., Ghosh, S., Roy, S., Suganthan, S. and Nagaratnam, P. (2012) An Adaptive Differential Evolution Algorithm with Novel Mutation and Crossover Strategies for Global Numerical Optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*: *Cybernetics*, **42**, 482-500. http://dx.doi.org/10.1109/TSMCB.2011.2167966

[7] Skrovseth, S.O., *et al.* (2010) A Computer Aided Diagnostic System for Malignant Melanomas. 3*rd International Symposium on Applied Sciences in Biomedical and Communication Technologies* (*ISABEL*). http://dx.doi.org/10.1109/isabel.2010.5702825

[8] Ruiz, D., *et al.* (2011) A Decision Support System for the Diagnosis of Melanoma: A Comparative Approach. *Expert Systems with Applications*, **38**, 15217-15223. http://dx.doi.org/10.1016/j.eswa.2011.05.079

[9] Rahman, M.M., Bhattacharya, P. and Desai, B.C. (2008) A Multiple Expert-Based Melanoma Recognition System for Dermoscopic Images of Pigmented Skin Lesions. 8*th IEEE International Conference on BioInformatics and BioEngineering, BIBE*.

[10] Masood, A. and Ali Al-Jumaily, A. (2013) Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms. *International Journal of Biomedical Imaging*, **22**. http://dx.doi.org/10.1155/2013/323268

[11] Mitchell, K.N. (2000) Text Classification from Labeled and Unlabeled Documents Using EM. *Machine Learning*, **39**, 103-134. http://dx.doi.org/10.1023/A:1007692713085

[12] Collobert, R., Sinz, F., Weston, J. and Bottou, L. (2006) Large Scale Transductive SVMs. *Journal of Machine Learning Research*, **7**, 1687-1712.

[13] Blum, A. and Chawla, S. (2001) Learning from Labeled and Unlabeled Data Using Graph Mincuts. *Intl. Conference on Machine Learning*.

[14] Kumar, P.P. and Millie, A (2010) Self Adaptive Differential Evolution Algorithm for Global Optimization. *Swarm, Evolutionary, and Memetic Computing*, 103-110. http://dx.doi.org/10.1007/978-3-642-17563-3_13

[15] Zhou, Z.-H., Zhan, D.-C. and Yang, Q. (1999) Semi-Supervised Learning with Very Few Labeled Training Examples. *Proceedings of the National Conference on Artificial Intelligence*, Menlo Park, Cambridge, London.

[16] Li, K., Luo, X. and Jin, M. (2010) Semi-Supervised Learning for SVM-KNN. *Journal of Computers*, **5**, 671-678. http://dx.doi.org/10.4304/jcp.5.5.671-678

[17] Masood, A., Al-Jumaily, A. and Anam, K. (2015) Self-Supervised Learning Model for Skin Cancer Diagnosis. 7*th International IEEE/EMBS Conference on Neural Engineering* (*NER*), 1012-1015. http://dx.doi.org/10.1109/ner.2015.7146798

[18] Masood, A., Al-Jumaily, A. and Anam, K. (2014) Texture Analysis Based Automated Decision Support System for Classification of Skin Cancer Using SA-SVM. *Neural Information Processing*, Springer International Publishing, 101-109. http://dx.doi.org/10.1007/978-3-319-12640-1_13

[19] Haupt, R.L. and Haupt, S.E. (2004) Practical Genetic Algorithms. John Wiley & Sons.

[20] Firpi, H.A.G. and Erik, D. (2004) Swarmed Feature Selection. 33*rd Applied Imagery Pattern Recognition Workshop*. http://dx.doi.org/10.1109/AIPR.2004.41

[21] Khushaba, R.N., Al-Ani, A. and Al-Jumaily, A. (2008) Differential Evolution Based Feature Subset Selection. 19*th International Conference on Pattern Recognition*. http://dx.doi.org/10.1109/icpr.2008.4761255