Differential Evolution based Advised SVM for Histopathalogical Image Analysis for Skin Cancer Detection

Ammara Masood, Adel Al- Jumaily, Member, IEEE

Abstract— **Automated detection of cancerous tissue in histopathological images is a big challenge. This work proposed a new pattern recognition method for histopathological image analysis for identification of cancerous tissues. It comprised of feature extraction using a combination of wavelet and intensity based statistical features and autoregressive parameters. Moreover, differential evolution based feature selection is used for dimensionality reduction and an efficient self-advised version of support vector machine is used for evaluation of selected features and for the classification of images. The proposed system is trained and tested using a dataset of 150 histopathological images and showed promising comparative results with an average diagnostic accuracy of 89.1%.**

I. INTRODUCTION

Malignant melanoma is one of the deadliest forms of skin cancer. In US, one person dies of melanoma every 57 minutes [1]. In 2014, an estimated 128,000 new cases of cancer were diagnosed in Australia, with that number expected to rise to 150,000 by 2020 [2]. Cancer treatment costs constitute more than $3.8 billion (7.2%) of health system costs. Pathologists attribute the success of skin cancer diagnosis and treatment to early detection.

Traditionally, in the skin cancer diagnosis process, pathologists use histopathological images of biopsy samples removed from patients and examine them under a microscope. A pathologist typically examines the image to observe the deviations in the cell structures and/or the change in the distribution of the cells across the tissue under examination. However, these judgments depend on their personal experience and expertise and often lead to considerable variability [3]. To overcome this problem and improve the reliability of diagnosis process, it is important to develop computational tools for automated diagnosis that operate on quantitative measures. Such tools can facilitate objective mathematical judgment complementary to that of a pathologist, and help them in identifying the affected areas efficiently. One of the biggest challenges in developing such tools for histopathological images is the feature selection to represent a cell/tissue in the task of cellular or tissue level property [4]. The features should provide distinguishing quantitative measures to detect cancerous regions. In addition

Ammara Masood is a PhD student at University of Technology Sydney, P.O. Box 123 Broadway, NSW 2007 Australia (Phone: (+61)466504940; e-mail: ammara.masood@student.uts.edu.au).

Adel Ali Al-Jumaily is Associate Professor at School of Electrical, Mechanical and Mechatronic Systems, University of Technology, Sydney (e-mail: Adel.Al-Jumaily@uts.edu.au).

to this an efficient algorithm is also required for the evaluation and classification stage.

Due to the complex nature of images during histopathological image analysis [5], it is not suitable to rely on just one type of feature extraction method. Thus, we proposed use of set of features based on intensity distribution analysis using both first and second order statistics, fuzzy mutual-information based wavelet packet transform and autoregressive modeling.

Apart from getting a good variety of differentiating features to start with, a good feature selection method is also important for removing irrelevant and redundant features for reducing amount of data for classifier learning, improving algorithms' predictive accuracy and increasing the comprehensibility of the constructed models. The most important factors defining a feature selection method are the search procedure and the evaluation measure.

Searching for the optimal subset, which can result in best training and testing performance is a quite challenging task. The exhaustive search of all feature space, can guarantee the optimal solution, but it is impractical even with moderate size feature sets. A number of other search strategies varying in optimality have been proposed in the literature. Some of the state of the art methods include stochastic methods such as genetic algorithm (GA), swarm intelligence like Particle swarm optimization (PSO) and differential evolution and extensions of these methods. A review of evolutionary optimization can be found in [6]. When considering the feature selection problem, many methods perform well on certain datasets, while such methods may fail to escape local minima when applied to other datasets. This paper proposes a new differential evolution algorithm based method for feature selection that showed promising results for histopathological image analysis for skin cancer diagnosis.

On the other hand, evaluation measure is usually categorized as 1) filter based methods which depend on some kind of estimation of importance of feature subsets, and 2) wrapper based methods where the importance of features is measured using a classification algorithm. Wrapper based methods are computationally more expensive than filter but they are more accurate and accuracy is one of our main concerns while developing a diagnostic model.

In this paper, we have used wrapper based approach using our proposed advised support vector machine for evaluating the selected feature subsets. The proposed classification algorithm has a tendency to deal with misclassified data of the training phase and reducing the effect of outliers that may

affect classification process. The final trained model on the basis of the selected subset of features is then analyzed with test dataset to validate the general accuracy of the proposed model. The experimental analysis is based on a dataset of 150 histopathological images and results are compared with some other methods used in literature for histopathological images.

The paper is organized as follows: Section II provides the details of the proposed methodology. It includes brief introduction to the suggested initial feature set; detail steps of the feature selection process and mathematical details of the evaluation/classification algorithm. Section III presents our experimental results. It includes analysis of the effect of different sizes of selected feature subsets on the classification accuracy of the overall model and comparisons of the proposed method with some other state of the art methods in this area. Finally, conclusion is provided in Section IV.

II. METHODOLOGY

A. Feature Extraction

A total of 45 features were extracted using three types of methods 1) autoregressive modeling [7] , 2) statistical intensity analysis using Gray Level Co-occurrence matrix and Grey-Tone difference matrix and 3) transform-based approach using fuzzy mutual-information based wavelet packet transform. Autoregressive parameters provide a powerful tool for distinguishing images with cancerous tissue from the regular healthy ones. Statistical analysis [8] based Grey-Level Co-occurrence Matrix and Grey-Tone Difference Matrix provided the features using pure numerical analysis of pixels intensity distribution that is different for healthy and cancerous tissues. Transform approaches provided an equivalent transformation to the image that is then analyzed as a representative proxy for original image. An improved extension of wavelet packet transform, Fuzzy mutual-information based wavelet packet transform [9] is used here for extracting features that can help in differentiating the cancerous and no cancerous tissues. For mathematical details of suggested features extraction methods refer to our previous work [10].

B. Differential Evolution based Feature Selection

Differential evolution (DE) is a population based optimization method, which has attracted an increased attention [11]. It is capable of handling nonlinear objective functions. It has parallel and direct search approach and good convergence. Like genetic algorithm, it uses crossover and mutation as selection mechanisms. However, mutation is the central procedure and is based on differences of randomly sampled pairs of solutions within the population. The proposed feature selection method is an extension of DE-based feature selection technique proposed in [12] . 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. Input the original feature set that mentioned in the (feature extraction) subsection A.

2. Generate a population of NP members each of D-dimensional real valued parameters, where NP is the population size and D is the number of parameters to be optimized.

3. Initialize the advised Support vector machine for evaluating the accuracy of each feature subset.

4. **while** the termination condition (maximum number of iterations) is not met do.

5. for all population members – vector z_i do

6. Create a mutant vector $v_{i,g}$ by merging three different randomly selected vectors.

$$v_{i,g} = z_{r0,g} + F \times (z_{r1,g} - z_{r2,g}) \qquad (1)$$

7. Employ uniform cross over for z_i and $v_{i,g}$ for building trial vector $u_{i,g}$ as follows.

$$u_{j,i,g} = \begin{cases} v_{j,i,g} & if \ rand(0,1) \leq C_r \\ z_{j,i,g} & otherwise \end{cases} \qquad (2)$$

Here Fϵ (0,1) is a scale factor controlling the rate at which population evolve. C_r is the cross over probability controlling the fraction of parameter values that are copied form the mutant. The index g represents the generation to which the corresponding vector belongs. $'i'$ is the population index ranging from 0 to NP-1 and parameters inside vectors are indexed with 'j' which operates from 0 to D-1.

8. For each member in the group

i. Specify the corresponding feature subset and advised SVM parameters according to the member.

ii. Use Advised SVM refer to figure.1 to calculate the classification output and accuracy as the fitness (F) of the member. The accuracy is based on the correctness of y(k) value (0 cancer , 1 non cancer) when compared to the actual diagnostic done by the pathologist for the test image.

9. In or-der to overcome the problem of duplicate features roulette wheel weighing scheme is utilized as mentioned in the original DEFS proposed in [12]. Thus a cost weighting is implemented where the probabilities of individual features are calculated from the distribution factor that is associated with each feature. The distribution factor of feature f_i within the current generation g is calculated as follows:

$$FD_{j,g} = a_1 \times \left(\frac{PD_j}{PD_j + ND_j}\right) + \frac{NF - DNF}{NF} \times (1 - \frac{(PD_j + ND_j)}{\max(PD_j + ND_j)}) \qquad (3)$$

where NF is the total number of features and DNF number of desired features. PD_j and ND_j is the number of times feature f_i has been used in the good subsets and less competitive subsets respectively. Whereas, a_1 is the constant that reflects the importance of features in PD.

10. Calculate the relative difference using following relationship [13].

$$T = (FD_{g+1} - FD_g) \times FD_{g+1} + FD_g) \qquad (4)$$

Thus the distribution factor provided to roulette wheel is calculated by the difference between the relative frequencies of distribution estimated form FD_g and FD_{g+1}. This helps in supressing the domination of certain features on the distribution factor.

12. if $F(u_i) \geq F(z_i)$ (fitness function i.e. accuracy of classifier) $z_i \leftarrow u_i$

13. Update the vectors end if

14. end for

15. end while

C. Advised Support Vector Machine based classification

In this work, a non-iterative self-advising approach for SVM is adapted that extracts subsequent knowledge from the misclassified data in training phase that can be a result of outliers or the data that have not been separated correctly. This is done by generating advice weights [14] based on the distance of misclassified training data from the correctly classified training data, and through use of these weights together with decision values of SVM in the test phase. These weights also help the algorithm to eliminate the outlier data.

The details of Advised SVM algorithm is as follows:

1. The classifying hyperplane is found by using decision function $f(x) = sign(\sum_{\alpha_l > 0} y_l \alpha_l \, k(x, x_l) + b)$, here x_l is the input vector corresponding to the lth sample and labelled by y_l depending on its class and α_l is the nonnegative Lagrange multiplier that is inconsistent with standard SVM training.

Note that in order to use SVM to produce non-linear decision functions as the data is comprised of nonlinearly separable cases, radial basis function kernel $K(x_l x_m) = e^{-\gamma |x_l - x_m|^2}$ is used for necessary operations in input space.

2. The data samples that are misclassified in the initial training phase are identified. The misclassified data sets (MD) in the training phase is determined as

$$MD = \cup_{l=1}^{N} x_l \mid y_l \neq sign(\sum_{\alpha_m > 0} y_l \, \alpha_m \, k(x_l, x_m) + b) \quad (5)$$

The MD set can be null, but experimental results revealed that the occurrence of misclassified data in training phase is a common occurrence.

3. If the MD is null, go to the testing phase, else compute neighbourhood length (NL) for each member of MD . NL is given as

$$NL(x_l) = minimum_{x_m} (\|x_l - x_m\| \mid y_l \neq y_m) \quad (6)$$

Where x_m , m=1, …., N are the training data that do not belong to the MD set.

4. For each sample x_n from the test set advised weight $AW(x_n)$ is computed. Where AW is computed as in equation (7), These AWs represent how close the test data is to the misclassified data

$$\begin{cases} AW = 0 & \forall x_l \in MD \,, \|x_n - x_l\| > NL(x_l) \, or \, MD = NUL, \\ AW = \sum 1 - \dfrac{\sum_{x_l} \|x_n - x_l\|}{\sum_{x_i} NL(x_i)} & x_l \in MD \,, \|x_m - x_l\| \leq NL(x_l) \end{cases} \quad (7)$$

5. The absolute value of the SVM decision values for each x_n from the test set are calculated and scaled to [0, 1].

6. For each x_k from the test set, If (AW (x_k) < decision value (x_k) then $y_k = sign (\sum_{\alpha_m > 0} y_m \alpha_m \, k(x_k, x_m) + b)$

which is in consistence with normal SVM labelling, otherwise $y_k = y_l \mid (\|x_k - x_l\| \leq NL(x_l) \, and \, x_l \in MD)$
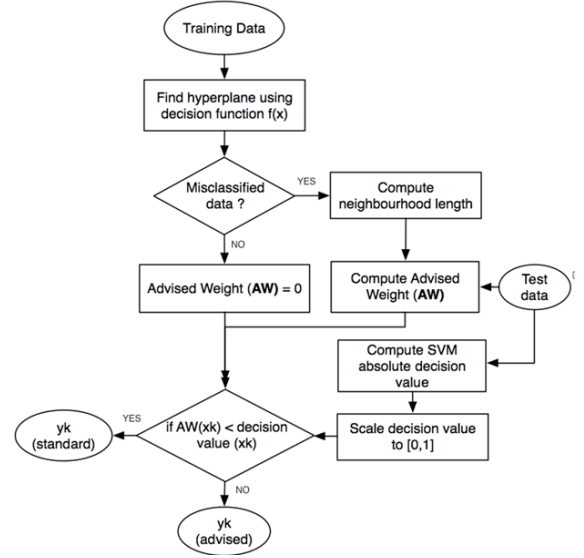

Figure 1. Advised support vector machine

III. EXPERIMENTAL ANALYSIS AND DISCUSSION

The database used for analysis of proposed method includes 150 histopathological images taken from biopsies of skin cancer patients, few samples shown in Figure 2. Most images were obtained from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. A total of 45 features (15 GLCM, 5 GTDM, 15 FMI_WPT, and 10 Autoregressive) were extracted for each image. 100 images were used for training and 50 images were used for testing. The whole process is implemented using MATLAB software R2013.
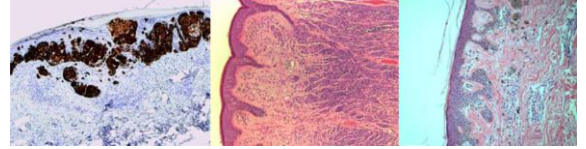

Figure 2. Sample of few Histopathological images used

The performance of the proposed model will be compared with the ones based on well-established binary Genetic Algorithm BGA [15], Binary PSO (BPSO) [16], improved BPSO [17] and hybrid GA [18].

For GA probability of mutation = 0.02 and probability of crossover was chosen as 0.5 after running several tests. This is used to make sure to have the number of '1's in the strings matching a predefined number of desired features. For BPSO the inertia weight was made to decrease linearly from 0.9 to 0.4 while the maximum velocity was set to be clipped within 20% of the corresponding variable; and acceleration constants were set to 2.0. Both of BGA and BPSO utilize binary strings representing a feature subset with ones and zeros to indicate the selection and neglecting of features respectively. Improved binary particles warm (IBPSO) was implemented according to the algorithm described in [12].Hybrid genetic search algorithm (HGA) was implemented as proposed in [13] to search for subsets of fixed sizes. It should be noted HGA is computationally very

expensive for larger datasets, as the number of subsets to be formed and evaluated increases with the number of features in the dataset.

All methods were made to start from the same initial population with the population size set to 50 and same number of iterations. The chosen fitness function was set to the classification accuracy, which is utilized to check the performance of the proposed method. Due to the relatively small number of samples in the available data, a 10 fold cross validation technique was used. Since the appropriate size of the most predictive feature subset is unknown, experimental analysis is done for various feature set sizes ranging from 5 to 20 with a step of 1. The methods were employed for 10 runs when searching for each specific feature subset size and the average is reported as classification accuracy here (with advised SVM as classifier). It was observed, as shown in figure 2, that at average the best performance was with feature subset size of 19.

In addition to this, table 1 shows the performance metrics of proposed model when compared with those based on KNN and standard SVM classifiers. The results prove the effectiveness of the proposed analysis method based on differential evolution and advised SVM. It provided best results in searching dataset with different number of features for the subset that best interact together to develop a model that works efficiently for identification of cancer in histopathological images.
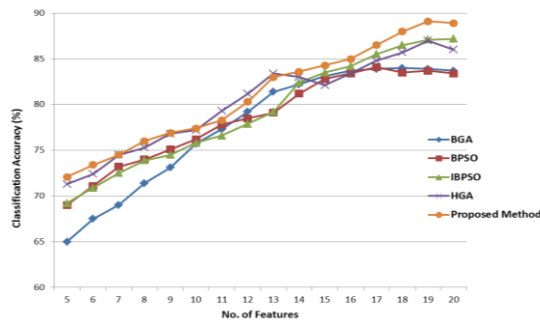

Figure 3. Average classification accuracies vs. feature subset sizes

TABLE I.   PERFORMANCE EVALUATION BASED ON ACCURACY

Classifier	Feature Selection Method				
	BGA	BPSO	HGA	IBPSO	DEFS$_A$
KNN	80.5	81	83.5	82	84
SVM	82.3	81.2	84.1	84.2	85.5
A-SVM	83.9	83.7	87	87.1	89.1

IV. CONCLUSION

This paper deals with the development of pattern recognition model for histopathological images for skin cancer detection. It presented a novel feature subset selection method based on the combination of differential evolution and advised support vector machine. Experimental analysis shows that the proposed model works well and provides an optimal number of feature set with higher classification rate when compared with some other popular methods used in literature. Although for this application we had access to limited amount of data but this method can be applied to larger datasets and can help in reducing the computational cost of the system along with achieving good performance.

REFERENCES

[1] American Cancer Society. Cancer Facts & Figures 2014. Accessed January 5, 2015]; Available from: http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc-042151.pdf.

[2] Australian Institute of Health and Welfare 2014. , in ACIM (Australian Cancer Incidence and Mortality) Books. 2014: Canberra: AIHW.

[3] Preti, M., et al., "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*, vol. 107, pp. 594-599, 2000.

[4] Yener, C.D.a.B., "Automated cancer diagnosis based on histopathological images: a systematic survey", Rensselaer Polytechnic Institute, Department of Computer Science, 2009.

[5] Verhaegen, P.D.H.M.V.Z., Paul P. M. Pennings, Noor M. Van Marle, Jan Niessen, Frank B. Van Der Horst, Chantal M. A. M. Middelkoop, Esther, "Differences in collagen architecture between keloid, hypertrophic scar, normotrophic scar, and normal skin: An objective histopathological analysis", *Wound Repair and Regenera*tion, vol. 17, pp. 649-656, 2009.

[6] Nguyen, T.T.Y., Shengxiang Branke, Juergen, "Evolutionary dynamic optimization: A survey of the state of the art", *Swarm and Evolutionary Computation*, vol. 6, pp. 1-24. 2012.

[7] Mao, J. and A.K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models", *Pattern Recognition*, vol. 25, pp. 173-188, 1992.

[8] P. Mohanaiah, P.S., L. GuruKumar, "Image Texture Feature Extraction Using GLCM Approach", *International Journal of Scientific and Research Publications (IJSRP),* vol. 3(5): pp. 1-5, 2013.

[9] 9. Khushaba, R.N.K., S. Lal, S. Dissanayake, G., "Driver Drowsiness Classification Using Fuzzy Wavelet-Packet-Based Feature-Extraction Algorithm", *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 121-131, 2011.

[10] Masood, A Al-Jumaily, Adel, Anam, Khairul, "Texture Analysis Based Automated Decision Support System for Classification of Skin Cancer Using SA-SVM", in Neural Information Processing, vol. 8835, pp. 101-109, 2014.

[11] Price, K.S., Rainer M Lampinen, Jouni A, Differential evolution: a practical approach to global optimization, Springer Science & Business Media, 2006.

[12] Khushaba, R.A.-A., Ahmed, Al-Jumaily, Adel, "Feature Subset Selection Using Differential Evolution", Advances in Neuro-Information Processing, vol. 5506, pp. 103-110, 2009.

[13] Bharathi, P.S., P, "Optimal Feature Subset Selection Using Differential Evolution and Extreme Learning Machine", International Journal of Science and Research, vol. 3, pp. 1898-1905, 2012.

[14] Maali, Y.A.-J., Adel, "Self-advising support vector machine", Knowledge-Based Systems, 2013, vol 52, pp. 214-222.

[15] Haupt, R.L.H., Sue Ellen, Practical genetic algorithms, John Wiley & Sons, 2004.

[16] Firpi, H.A.G., Erik D. "Swarmed Feature Selection," in Proceedings of the 33rd Applied Imagery Pattern Recognition Workshop, pp. 112-118, 2004.

[17] Khushaba, R.N.A., Akram Al-Ani, Ahmed Al-Jumaily, Adel Zomaya, Albert Y, "A novel swarm based feature selection algorithm in multifunction myoelectric control," Journal of Intelligent and Fuzzy Systems, vol. 20, pp. 175-185, 2009.

[18] Oh, I.-S.L., Jin-Seon Moon, Byung-Ro, "Hybrid genetic algorithms for feature selection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26(11), pp. 1424-1437, 2004.