SA-SVM based automated diagnostic System for Skin Cancer

Ammara Masood, Adel Al-Jumaily
University of Technology Sydney, Australia

ABSTRACT

Early diagnosis of skin cancer is one of the greatest challenges due to lack of experience of general practitioners (GPs). This paper presents a clinical decision support system aimed to save time and resources in the diagnostic process. Segmentation, feature extraction, pattern recognition, and lesion classification are the important steps in the proposed decision support system. The system analyses the images to extract the affected area using a novel proposed segmentation method H-FCM-LS. The underlying features which indicate the difference between melanoma and benign lesions are obtained through intensity, spatial/frequency and texture based methods. For classification purpose, self-advising SVM is adapted which showed improved classification rate as compared to standard SVM. The presented work also considers analyzed performance of linear and kernel based SVM on the specific skin lesion diagnostic problem and discussed corresponding findings. The best diagnostic rates obtained through the proposed method are around 90.5 %.

Keywords: Skin Cancer, diagnosis, feature extraction, classification, support vector machine.

1. INTRODUCTION

Malignant melanoma is one of the deadliest forms of skin cancer. A rapid increase in melanoma cases is observed in Europe, North America, and Australia over the last decade. Over 76,250 new cases of invasive melanoma were diagnosed in the US in 2012 [1].An estimated 1,890 Australians die from skin cancer each year [2]. From treatment point of view, skin cancer is one of the most expensive forms of cancer. But early diagnosis can make the situation better as melanoma has near 95% cure rate if diagnosed and treated in early stages [3]. In visual diagnosing, it is hard for medical professionals to differentiate normal from the abnormal mole. Dermatologist's data indicate that even in specialized centers diagnostic accuracy is only about 60%, and at the same time they are overloaded by referrals [4].

This study proposes an automated diagnostic system for skin cancer. The benefits of computerized diagnostic tools are twofold: a) it can be used as a standalone warning tools for helping the General practitioners during early diagnosis; (2) it can provide quantitative information about lesion to experts that can be considered during biopsy decision making.

A Computer Aided Diagnostic (CAD) system for diagnosis of skin cancer is aimed to find the exact boundaries of a lesion automatically and also to provide an estimate of the probability of a disease. Differentiation of malignant melanoma images from non-melanoma demands efficient image processing, segmentation, feature extraction and classification algorithms. There are various diagnostic systems proposed in literature [5-7] but as we discussed in [8] detailed research is required to make the best choice and for setting the benchmarks for diagnostic system development and validation. This paper presents a part of our research being carried out to come up with the best combination of segmentation, feature extraction and classification algorithms which can consequently form the basis of a more generalized and efficient skin cancer diagnostic system. The diagnostic model proposed in this paper is shown in Fig. 1.

The decision support system proposed in this paper uses adaptive median filter for pre-processing of image to reduce the ill effects and various artifacts like hair that may be present in the dermoscopic images. It is followed by the detection of the lesion by our Histogram based fuzzy C means thresholding algorithm presented in [9]. This algorithm provided efficient segmentation results as compared to other popular segmentation methods; the comparative analysis is presented in [10]. Once the lesion is localized, different intensity, spatial/frequency and texture based features are quantified. Finally, Self-advisable Support Vector machine (SA-SVM) is used for classification of cancerous and non-cancerous skin lesions. SA-SVM is an improved version of SVM. In order to

justify the performance of the proposed method, classification results of the proposed method are also compared with the diagnosis results of standard SVM.

This paper is organized as follows: Section 2 describes the computer-aided diagnosing (CAD) which consists of the pre-processing, segmentation, features extraction and classification stages. Section3 presents the experimental results, comparative analysis and Discussion, and finally section for conclusions and intended future work.
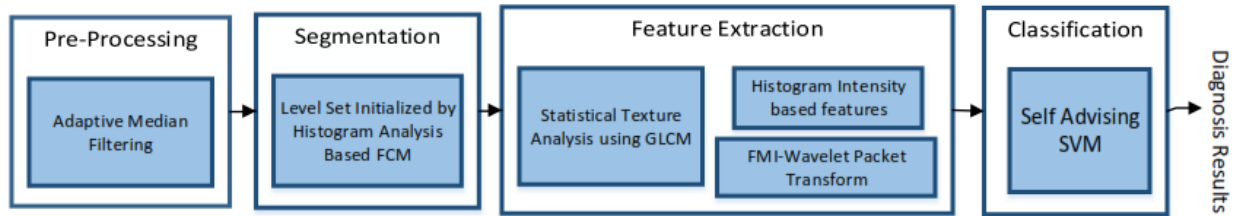

Figure. 1 Computer Aided Diagnostic Support System

2. MATERIAL AND METHODS

2.1 Experimental Dataset

A clinical database of dermoscopic and clinical view lesion images were obtained from different sources but most of the images came from the Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital, Vienna General Hospital and Hospital Charles Nicolle CHU Rouen, France. The images have been stored in the RGB color format having different dimensions which were rescaled to a resolution of 720x 472 with bit depth 24 and size around 526KB. A total of 168 images (56 benign and 112 melanoma) were included in the experimental data set.

2.2 Pre-Processing

Skin images have certain extraneous artifacts such as skin texture, dermoscopic gel and hair that make border detection a bit difficult. In order to reduce the effect of these artifacts on segmentation results, it is necessary to pre-process the images with a smoothing filter like adaptive median filter. The median filter also performs well as long as the spatial density of the impulse noise is not too large. However the adaptive median filtering has a better capability to handle impulse noise with even larger probabilities. An additional benefit of the adaptive median filter is that it seeks to preserve details while smoothing the non-impulse noise [11]. Considering the high level of noise that may be present in skin lesion images and the need of preserving structural details, the adaptive median algorithm performed quite well.

2.3 Segmentation

We proposed a segmentation algorithm, histogram analysis based fuzzy C mean algorithm for Level Set initialization (H-FCM-LS) as presented in figure 2. The mathematical details of our proposed segmentation algorithm can be found in our research publication [12]. We found that the practical significance of fuzzy clustering for multidimensional feature space is of great importance when it comes to skin cancer diagnosis. In this case deterministic misclassification can be very costly and a sophisticated thresholding procedure is required for increasing the accuracy of segmentation results. In the proposed method, histogram analysis of image was done to see average intensity distribution in the images and then the hard threshold was selected between classes with dominant intensity values. Thus, this method solves the difficulty that may arise in finding effective threshold automatically for different images regardless of intensity variety. This method was further used as an initializing step for complex segmentation method like Level set having spatial information. Segmentation results for some of the skin lesion images are shown in Fig. 2.
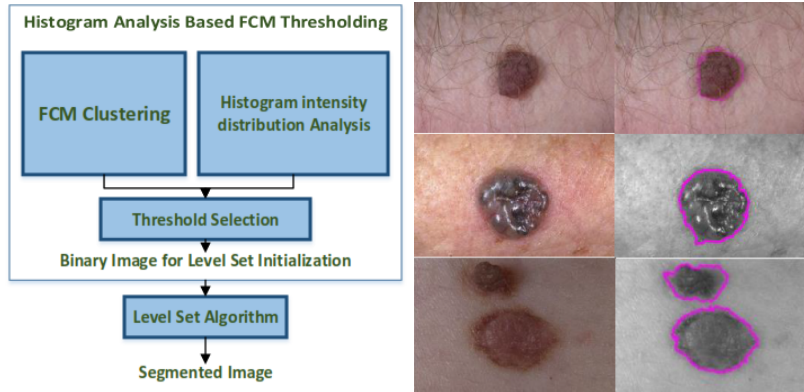
Figure. 2. Block Diagram & Results of Segmentation Algorithm (H-FCM-LS)

2.4 Feature Extraction

The features analysis of skin lesions suggest that surface characteristics of the skin lesion changes as it progresses towards melanoma. In the early stages of melanoma the features used in visual examination are hardly visible and thus it may lead to a false diagnosis. However, if images of skin lesions can be collected that has intensity, spatial/frequency and texture information, a non-invasive method of lesion classification based on the surface characteristics can be developed. Thus, in this paper, we used intensity histogram based features, texture-based features calculated using Grey-Level Co-occurrence Matrix and spatial/frequency based features calculated using Fuzzy mutual-information based wavelet packet transform (FMIWPT).

2.4.1 Intensity Histogram Features

The characteristic of the histogram is closely related with the characteristic of image such as brightness and contrast. Histogram based features include:

$$\text{Mean } \mu = \sum_{i=1}^{N} p_i x_i \tag{1}$$

$$\text{Variance } \sigma^2 = \sum_{i=1}^{N} p_i \ (x_i - \mu)^2 \tag{2}$$

$$\text{Standard deviation } \sigma = \sqrt{\sigma^2} \tag{3}$$

$$\text{Skewness} = \sum_{i=1}^{N} \frac{(x_i - \mu)^3}{(N-1)(\sigma)^3} \tag{4}$$

$$\text{Kurtosis} = \sum_{i=1}^{N} \frac{(x_i - \mu)^4}{(N-1)(\sigma)^4} \tag{5}$$

$$\text{Energy} = \sum_{i=1}^{N} (p_i \cdot p_i) \tag{6}$$

$$\text{Entropy} = -\sum_{i=1}^{N} (p_i \cdot \log(p_i)) \tag{7}$$

Mean reveals the general brightness of an image. Standard deviation or variance tells the contrast of an image. Image with good contrast have high variance. Standard Deviations (SD) also characterize the cluster. Skew measures how asymmetrical (unbalance) is the distribution of the gray level. Image with object in contrast background have high variance but low skew distribution i.e. one peak at each side of mean. Energy measurement is also closely related to skewness. Highly skew distribution usually gives high-energy measurement. Entropy provides the average number of bits to code each gray level. It has inverse relationship with skew and energy measurement. Highly skew distribution tends to yield low entropy. Within ROI (region of interest) in the segmented skin image a histogram distribution of the image is computed. Then six features are calculated for classification.

2.4.2 Fuzzy-Wavelet packet based features

The wavelet packet method is a generalization of wavelet decomposition and offers a richer signal analysis. Wavelet packet are waveforms indexed by three naturally interpreted parameters i.e. position, scale and frequency. The wavelet decomposition procedure splits approximation coefficients into two parts; a vector of approximation coefficients and a vector of detail coefficients both at a coarser scale. The information lost between two successive approximations is captured in detail coefficients and the new approximation coefficient vector is split again. In the wavelet packet approach, each detail coefficient vector is also decomposed into two parts as the approximation vector. Different extensions of wavelet packet transform are present in literature for different applications [13, 14]. It

is observed that features extracted using wavelet transforms provide significant increase in the classification accuracy. After converting skin images to corresponding vectors, Fuzzy mutual-information based wavelet packet transform (FMI-WPT) is used here to extract features from skin images. The detailed steps of FMIWPT algorithm are as follows:

1) For each original image vector, perform a full WPT decomposition to the maximum level J (taken as 3 here). For all j = 0, 1,. ., J and k = 0, 1, . . , 2j − 1, construct features according to relation $E_{\Omega j.k} = \log\left(\frac{\sum_n (w^T_{j,k,n} x)^2}{N/2^j}\right)$. Where $\Omega j,k$ is the decomposition subspace with j denoting scale and k denoting sub-band index within the scale [14].

2) Construct associated fuzzy sets and compute fuzzy entropies and mutual information. Then evaluate classification ability of n No. of features using fuzzy-set based criterion Fi where Fi = I(C; fi)/H (fi) for i = 1, 2, . . . , n.

Note: I (f;C) = H(f) − H(f |C) where H(f) is marginal entropy of f and H(f |C) is conditional entropy of f and C [15].

3) Determine the optimal WPT decomposition X, being the one that corresponds to the maximum value of F.

4) The set X is the final FMIWPT-based decomposition.

The aforementioned algorithm is applied to optimize the WPT tree of the skin image vector to form a feature vector that will be used for classification.

2.4.3 Gray Level Co-occurrence Matrix (GLCM)
The GLCM features characterize the texture of an image by calculating how often pairs of pixel occur in an image with specific values and in a specified spatial relationship [16]. This calculation is used to create GLCM, and then statistical measures are extracted from this matrix. In GLCM the number of rows and columns is equal to the number of gray levels, G, in the image. The matrix element P (i, j | Δx, Δy) is the relative frequency with which two pixels, separated by a pixel distance (Δx, Δy), occur within a given neighborhood, one with intensity 'i' and the other with intensity 'j'. The matrix element P (i, j | d, ɵ) contains the second order statistical probability values for changes between gray levels 'i' and 'j' at a particular displacement distance d and at a particular angle (ɵ). The class mark of ith intensity level is denoted by xi, frequency of ith level is denoted by fi and the relative frequency of the ith level is denoted by pi=fi/N. The mathematical of GLCM based features is provided in table 1.

Table 1. GLCM based Features

Feature	Mathematical Equation
Autocorrelation	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(ixj)p(i,j)$
Contrast	$\sum_{n=0}^{N-1} n^2 \left\{\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)\right\}, |i-j| = n$
Correlation	$= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{(ixj)p(i,j)-\{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$
Cluster Prominence	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i+j-\mu_x-\mu_y)^4 \times p(i,j)$
Cluster shade	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i+j-\mu_x-\mu_y)^3 \times p(i,j)$
Dissimilarity	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}|i-j|.p(i,j)$
Energy	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}p(i,j)^2$
Entropy	$-\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}p(i,j) \times \log(p(i,j))$
Homogeneity	$\frac{\sum_{i=0}^{N-1}\sum_{i=0}^{N-1}(ij)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$
Maximum probability	$MAX_{i,j}\, p(i,j)$

Difference variance	variance of p_{x-y}
Difference entropy	$-\sum_{i=0}^{N-1} p_{x-y}(i)\log(p_{x-y}(i))$
Inverse difference normalized	$=\sum_{i,j=1}^{N} \frac{C(i,j)}{1+|i-j|^2/N^2}$ where $C(i,j) = \frac{p(i,j)}{\sum_{i,j=1}^{N} p(i,j)}$
Inverse difference moment normalized	$\sum_{i,j=1}^{N} \frac{C(i,j)}{1+(i-j)^2/N^2}$
Information measure of correlation	$\dfrac{-\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)\times log(p(i,j)) + \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)\times log\{p_x(i)\,p_x(j)\}}{\max\{entropy\ of\ p_x, entropy\ of\ p_y\}}$

2.5 Classification

Classification of the lesion as cancer or non-cancer is the final step. Classification is a significant area of research in a variety of fields including pattern recognition, artificial intelligence medicine and vision analysis. For classification of skin lesions, an improved version of support vector machine (SVM), named Self Advising SVM is adapted here. SVM is a well-known machine learning method proposed by Vapnik [17]. The idea of SVM is to construct a maximized separating hyperplane that can separate the data in the feature space.

The classic SVM ignores the training data that has not been separated linearly by the kernels during the training phase. This occurs through the introduction of the tolerance parameters in the objective function and constraints. Thus, if data that is similar or identical to this misclassified data appears in the test set, it will be classified wrongly. This is because the data which is close to the misclassified data is uncertain. This misclassification is not reasonable and it can be handled if the available data and information in the training phase has not been ignored by the SVM algorithm.

In this study we adapted a non-iterative self-advising approach [18] for SVM that extracts subsequent knowledge from training phase. The misclassified data can come from two potential sources 1) outliers 2) data that have not been linearly separated by using any of the types of kernels. SA-SVM deals with the ignorance of SVM from the knowledge that can be acquired from the misclassified data by generating advice weights based on the use the misclassified training data, and through the use of these weights together with decision values of the SVM in the test phase. These weights also help the algorithm to eliminate the outlier data. The details of SA-SVM algorithm are as follows:

Training phase

1. Finding the hyperplane by using decision function $f(x) = \text{sign}(\sum_{\alpha_{i>0}} y_i \alpha_i\, k(x, x_i)\ +b)$, i.e. the normal SVM training. Note: the kernel function used here is radial Basis Function so $K(x_i x_j) = e^{-\gamma|x_i-x_j|^2}$.

2. To benefit from the misclassified data of the training phase, the misclassified data sets (MD) in the training phase is determined as

$$MD = \bigcup_{i=1}^{N} x_i\ |y_i \neq \text{sign}(\sum_{\alpha_j>0} y_i\, \alpha_j\, k(x_i, x_j) + b) \tag{8}$$

The MD set can be null, but experimental results have revealed that the occurrence of misclassified data in training phase is a common occurrence. Note that xi is the input vector corresponding to the ith sample and labeled by yi depending on its class and αi is the nonnegative Lagrange multiplier as used in standard SVM [17].

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

$$NL(x_i) = \text{minimum}_{x_j} (\|x_i - x_j\| \mid y_i \neq y_j) \tag{9}$$

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

Here the training data is mapped to a higher dimension,, the distance between xi and xj is be computed according to the following equation with reference to the related kernel k (RBF),

$$\|\theta(x_i) - \theta(x_j)\| = (k(x_i, x_i) + k(x_j, x_j) - 2k((x_i, x_j))^{0.5} \tag{10}$$

Testing phase

1. Compute the advised weight AW(xk) for each xk, from the test set. Where AW is computed as follows

$$\begin{cases} 0 & \forall x_i \in MD, \|x_k - x_i\| > NL(x_i) \text{ or } MD = NUL, \\ \sum 1 - \dfrac{\sum_{x_i} \|x_k - x_i\|}{\sum_{x_i} NL(x_i)} & x_i \in MD, \|x_k - x_i\| \leq NL(x_i) \end{cases} \tag{11}$$

These AWs represent how close the test data are to the misclassified data.

2. Compute the absolute value of the SVM decision values for each xk from the test set and scale it to [0, 1].
3. For each xk from the test set, If $(AW(x_k) < \text{decision value}(x_k)$ then $y_k = \text{sign}\left(\sum_{\alpha_j>0} y_j \alpha_j k(x_k, x_j) + b\right)$ which is normal SVM labeling, otherwise $y_k = y_i \mid (\|x_k - x_i\| \leq NL(x_i) \text{ and } x_i \in MD)$

3. EXPERIMENTAL RESULTS

Algorithms discussed in the previous sections have been implemented using MATLAB software R2013 and simulated by a system with corei5 3.10 GHz processor and 4 GB memory under Windows7 operating system. For training the SA-SVM, 84 images are used, 84 images were used for the testing. The classification results of SA-SVM are compared with standard SVM (both linear and kernel based i.e. polynomial kernel based SVM, Quadratic kernel based SVM, Radial basis Function based SVM and multilayer perceptron based SVM). The best results obtained for each classifier during the experiments are shown in the Table 2.

Table 2. Diagnosis Results of SA-SVM vs. Standard SVM

SVM type	True Positive	False Positive	True Negative	False Negative
Linear	40	7	21	16
Polynomial (Kernel)	46	6	22	10
Quadratic (Kernel)	47	6	22	9
Radial Basis Function (Kernel)	49	5	23	7
Multi-Layered Perceptron (Kernel)	36	15	13	20
SA-SVM	50	2	26	6

For cross validating the results, hold-out validation a specific type of k-fold cross validation is used. For each fold, we randomly divided images into two sets S1 and S2, so that both sets are of equal size. In order to implement it we shuffled the experimental data set and split it into two parts. Then we trained the SA-SVM using S1 and tested on S2. This is followed by training using S2 and testing using S1. This validation method has the advantage that our training and test sets are both large, and each data point is used for both training and validation on each fold. In order to ensure better validation of classifier performance the hold-out validation was repeated 5 times. The performance of the diagnostic system is analyzed using statistical parameters such as sensitivity, specificity and accuracy, see table 3. Higher value of both sensitivity and specificity shows better performance of the system.

Table 3. Measurement Parameters

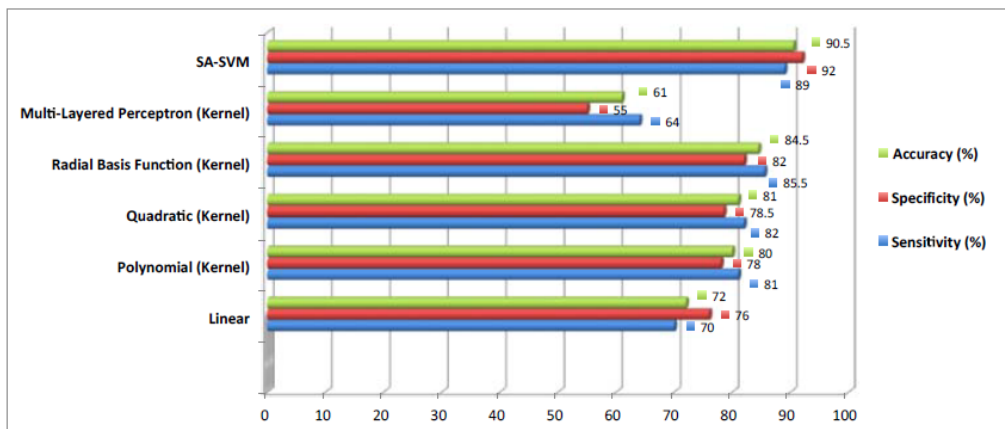| Measure	Formula
Sensitivity	True positive/ (True positive + False negative)
Specificity	True negative / (True negative + False positive)
Accuracy	(True Positive + True negative) /(True positive + False negative + True negative + False positive)

Figure.3. Relative Performance Measure of SA-SVM Vs. Standard SVM

The experimental analysis in Fig. 3 shows clearly that the average classification results obtained by the self-advising SVM are significantly better than the results of traditional SVM.

4. CONCLUSION

In this paper, an automated skin cancer diagnostic system is proposed based on self-advising SVM. Self-advising SVM uses information generated from misclassified data in the training phase and thus, improves the SVM performance by transferring more information from the training phase to the test phase. It is also taken into account that diverse range of features can be extracted from skin images using state of the art feature extraction methods to enhance the classification performance of the classifier. The diagnostic results obtained are quite satisfactory with sensitivity of 89 % and specificity of 92%.

Despite the high accuracy that can be achieved by the proposed computer aided diagnostic, for developing a more reliable diagnostic system we intend to test multiple classifier based systems as well to undo the chances of any misclassification due to classifier limitations. For this purpose we intend to do experiments combining different classification algorithms like neural networks, support vector machine, deep learning and extreme learning machine. The technical achievements of recent years in the areas of image acquisition and processing allow the improvement and lower cost of image analysis systems. Such tools may serve as diagnostic adjuncts for medical professionals for the confirmation of a diagnosis, and can give the opportunity of implementing more accurate, faster and reliable classification system.

REFERENCES

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

[2] Causes of death 2010, C.W.o. Australia, Editor, Australian Bureau of Statistics: Canberra, Australia.

[3] Thursfield V, F.H., Cancer in Victoria: Statistics and trends 2010, C.C. Victoria, Editor December 2011: Melbourne, Australia.

[4] Perrinaud, A., et al., Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems. British Journal of Dermatology, 2007. 157(5): p. 926-933.

[5] Ganster, H., et al., Automated melanoma recognition. IEEE Transactions on Medical Imaging,, 2001. 20(3): p. 233-239.

[6] Ignazio Stanganelli, A.B., Luigi Calori, Roberto Gori,, et al., Computer aided diagnosis of melanocytic tumors. ANTICANCER RESEARCH, 2005. 25: p. 4577-4582.

[7] Ruiz, D., et al., A decision support system for the diagnosis of melanoma: A comparative approach. Expert Systems with Applications, 2011. 38(12): p. 15217-15223.

[8] Masood, A., Ali Al-Jumaily, Adel, Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms. International Journal of Biomedical Imaging, 2013. 2013: p. 22.

[9] Ammara Masood, A.A.A.-J., Fuzzy C mean Thresholding based Level Set for Automated Segmentation of Skin Lesions. Journal of Signal and Information Processing, 2013. 4(3B): p. 66-71.

[10] Masood, A.J., Adel Ali Al Hoshyar, Azadeh Noori Masood, Omama. Automated segmentation of skin lesions: Modified Fuzzy C mean thresholding based level set method. in 2013 16th International Multi Topic Conference (INMIC), . 2013.

[11] Sun, T. and Y. Neuvo, Detail-preserving median based filters in image processing. Pattern Recognition Letters, 1994. 15(4): p. 341-347.

[12] Masood, A.A.-J., AdelAli Maali, Yashar, Level Set Initialization Based on Modified Fuzzy C Means Thresholding for Automated Segmentation of Skin Lesions, in Neural Information Processing, M.H. Lee, Akira Hou, Zeng-Guang Kil, RheeMan, Editor 2013, Springer Berlin Heidelberg. p. 341-351.

[13] Khushaba, R.N.A.-J., A. Al-Ani, A. Novel feature extraction method based on fuzzy entropy and wavelet packet transform for myoelectric Control. in Communications and Information Technologies, 2007. ISCIT '07. International Symposium on. 2007.

[14] 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,, 2011. 58(1): p. 121-131.

[15] Proakis J.G., M.D.G., Digital Signal Processing Principles, Algorithms, and Applications1996, New Jersey: Prentice-Hall.

[16] Haralick, R.M., Shanmugan, K., and Dinstein, I., Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, 1973. 3(6): p. 610–621.

[17] V., V., The nature of statistical learning theory. 2nd ed2000, New York: Springer.

[18] Maali, Y.A.-J., Adel, Self-advising support vector machine. Knowledge-Based Systems, 2013. 52(0): p. 214-222.