A Comparative Study of the Classification of Skin Burn Depth in Human

P.N.Kuan[1], S.Chua[1], E.B.Safawi[2], H.H.Wang[1] and W.Tiong[2]
[1]*Faculty of Computer Science and Information Technology Universiti Malaysia Sarawak.*
[2]*Faculty of Medicine and Health Science Universiti Malaysia Sarawak,*
94300 Kota Samarahan, Sarawak, Malaysia.
16020107@siswa.unimas.my

***Abstract*—A correct first evaluation of skin burn injury is essential as it is an important step in providing the first treatment to the patient by determining the burn depths. The objective of this paper is to conduct a comparative study of different types of classification algorithms on the classification of different burn depths by using an image mining approach. 20 classification algorithms were compared on a skin burn dataset comprising skin burn images categorized into three classes by medical experts. The dataset was evaluated using both a supplied test set and 10-fold cross validation methods. Empirical results showed that the best classification algorithms that were able to classify most of the burn depths using a supplied test set were Logistic, Simple Logistic, MultiClassClassifier, OneR, and LMT, with an average accuracy of 68.9% whereas for 10-fold cross validation evaluation, the best result was obtained through the Simple Logistic algorithm with an average accuracy of 73.2%. It can be concluded that Simple Logistic has the potential to provide the best classification for the degree of skin burn depth.**

***Index Terms*—Skin Burn; Classification; Segmentation; Image Mining Approach.**

I. INTRODUCTION

Human skin is the largest organ that covers the outer part of the body. Generally, human skin is made up of three layers as shown in Figure 1: (i) the epidermis, which is the outermost layer of the skin, (ii) the dermis, lay underneath the epidermis layer and is divided into two sub-layers, which are papillary layer (superficial) and reticular layer (deep) and (iii) the hypodermis, which is the inner layer of the skin, constitutes of fat and connective tissue [1].

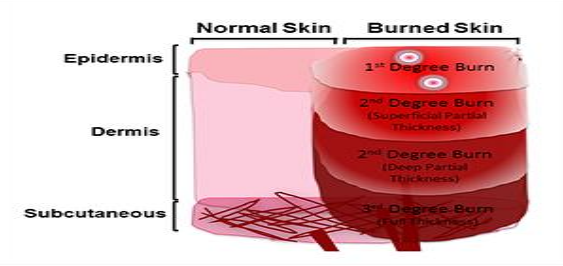

Figure 1: Human skin structure [2]

There are three degrees of skin burns: (i) First degree burn, which include only the epidermis, (ii) Second degree burn, classified into (a) superficial partial thickness burn, which involve the entire epidermis and the upper layer of the dermis (papillary layer) and (b) deep partial thickness burn, which involve the entire epidermis and most of the dermis, and (iii) Third degree burn, also known as full thickness burn, in which all the layers of the skin are destroyed, and some may extend into muscle and bone [3]. The severity of the burn injury is usually determined by the depth of the burn.

Patients with burn injuries usually consult doctors for treatment, where assessment is based on visual findings on examination. Sometimes the depth of the burn is not easily defined, as there could be mixed depth appearance. Medical practitioners with limited experience may at times be confused with the depth and severity of the burns, especially in non-clear-cut cases. In rural areas, patients may only have access to other healthcare staff at nurse-led clinics. A wrong assessment of burn depth results in inappropriate and inaccurate initial management of the burn injuries. Such mistakes translates into poor healing process, infections, undesirable scars and reduced body functions post burns.

The current state-of-the-art in burn depth classification is focused on identifying features that are capable of differentiating between healthy skin and the burn wound as well as being dependent on the feature selection performed by intelligent classifiers, such as deep learned convolutional neural network. However, the images used were manually registered with infrared markings [4].

In this work, an image mining approach is used to evaluate the image of a skin burn injury and to classify the burn injury into one of the burn depths. Based on the burn depth classification, suitable treatment can then be recommended. Many previous work in the literature used colour as the main characteristic to differentiate between different burn depths. However, this research work is focused on extracting both the colour and texture features. The feature extraction is carried out using discrete wavelet transform (DWT) and followed by applying principle component analysis (PCA) to reduce the feature dimensions [5, 6]. Gray Level Co-occurrence Matrix (GLCM) is then used to extract texture features from the decomposed images [6]. The classification was conducted using the binary classification approach, by taking one class versus all other classes. The evaluation measures used are accuracy, precision, recall, and F-measure. The main contribution of this work is the comparative study of classification of skin burn depths based on the features extracted. Another contribution includes a hybrid segmentation method, which used RGB threshold values to separate the body part from the background and Otsu's method of thresholding [7] to separate the burn wound from the body part in the image. In addition, in feature extraction, both colour and texture features were adopted and used in combination.

II. Literature Review

Previous works focused on evaluating the skin burn depth in order to reduce the specialist's high experience requirement during visual examination. The works in the literature either used a segmentation-based approach, which meant segment the burn wound from the skin region in the image before feature extraction or a segmentation-free approach, which extracted features directly from the skin burn image. In this section, some works related to image acquisition methods, image segmentation algorithms, features extraction and image classification methods are presented.

A. Image Acquisition Methods

Acha et al. [8] proposed an acquisition protocol apart from the influence of illumination and the camera calibration issues that they identified when using a digital photographic camera. For illumination influence issue, they concluded that the xenon flash illumination was able to dominate the illumination after their experiments under three different situations, which were in a darkroom with built in flash, in a darkroom with florescent light, and in a room under diffused sunlight [8]. For the issue of camera calibration, in order to convert RGB coordinates to a device-independent colour representation system, they proposed a matrix transformation between the RGB and CIE XYZ based on the Macbeth ColourChecker DC chart. This calibration is specific for each camera, thus different cameras used will need to perform calibration once [8]. The acquisition protocol for burn wounds are: (i) distance between camera and patient should be about 40-50 cm, (ii) healthy skin should appear in the image when possible, (iii) the background should be a green/blue sheet, (iv) the flash must be on, and (v) the camera should be placed parallel to the burn [8-11].

Wantanajittikul et al. [12] used 5 burn images provided by the medical center in their work. These original images contain background information. They divided the 5 burn images into a total of 34 sub-images and classify them [12].

The database used by Deepak et al. [13] consists of 120 images which are collected from the internet, self-captured from hospitals and scanned from books. All the images used are in jpeg format. Images are poor in quality due to poor illumination.

Suvarna et al. [14, 15] collected their skin burn images with three different burn depths from the internet, self-captured from hospital and scanned from biomedical books. Each of these burn depths comprised 40 images. These images were pre-labeled by a plastic surgeon and used the standard jpeg format.

Tran et al. [16] used real-life burn images provided by the hospital. These images were pre-labeled and a total of 396 burn images were used in this work, with 180 images of degree II burn, 192 images of degree III burn and 24 images of degree IV burn.

B. Image Segmentation Algorithms

Many different segmentation algorithms had been applied in the literature to segment the skin burn wounds from the healthy skin in images. A manual segmentation based on CIE L*u*v* colour coordinate space was proposed in the works of [8, 10, 11, 17]. According to them, L*u*v* and L*a*b* colour representation systems were known as uniform system. The reason was because the Euclidean distance between the measured colour in these spaces were almost similar to human perception of colour differences [8]. These two spaces were slightly different from each other as a* and b* in L*a*b* are independent from luminance, with the colour perception being strongly influenced by the luminance. However, both were equally good in providing a very good estimation of colour differences between the two colour vectors. Thus, the authors had chosen the L*u*v* colour space to be used in their work [8]. The segmentation algorithm steps proposed were: (i) select a small region in the burn wound, and then preprocess the image, (ii) convert to a single channel image, and finally (iii) threshold and post process [8, 10, 11, 17]. The drawback for this algorithm was the requirement of the user to manually select the colour to be segmented, thus, making it a semi-automated and user-guided approach. This may also be prone to bias in human colour perception. This segmentation of the burn wound based on the colour selected by a user may affect the burn degree classification results. For example, if the burn wound consists of mixed burns (i.e. Superficial thickness burn and deep partial thickness burn), which consist of two colours of different degree of burn classes in one burn wound, the user may select the colour to be segmented which belonged to the superficial thickness burn, despite the burn wound being actually deep partial thickness burn.

Wantanajittikul et al. [12] proposed a new segmentation algorithm to separate the skin region from the background and then in turn, separate the wound region from the healthy skin. The algorithm converted the entire RGB image to the Cr-space. Fuzzy C-means (FCM) clustering was then used to separate the skin region from the background. After that, in order to emphasize the burn wound region, the skin region from the RGB-space was converted to the L*u*v* space. FCM clustering was used again to separate the wound region from the healthy skin. Finally, the segmented wound region was preprocessed to eliminate noise by using mathematical morphology.

Tran et al. [16] used a normalization approach by normalizing the skin burn image into a standard size with the rate of 4:3 in order to remove the non-burn region of the images. The burn wound was then segmented according to the colour information used in the work of Acha et al. [8], which was done by user selection.

With regards to skin burn images, a related work on other medical images in the literature was also reviewed. In a study on the assessment of diabetic foot ulcers, the segmentation algorithm used was as follows: (i) The foot outline was determined within the image by finding the largest connected component whereby the colour of the component was similar to the preset standard skin colour. (ii) The wound boundary was then determined from within the foot image based on three assumptions. First, there should be a little irrelevant background information from the foot image. Second, a nearly uniform colour feature of the healthy skin was assumed on the sole of the foot. Third, the edge of the foot outline assumed with no ulcer was located. (iii) After the wound boundary had been segmented successfully, colour segmentation was performed on it using K-means to classify the wound into granulation, slough and necrosis classes [18].

From the previous works reviewed, a segmentation algorithm that is able to detect and segment skin burn

wound without the involvement of the user is still preferable to ensure an automatic approach and to prevent user bias. Therefore, a hybrid segmentation method addressing this issue is proposed and will be discussed in Section III (A).

C. Features Extraction

Many different sets of features, either colour, or colour and texture, had been proposed in previous works. Colour and texture are the two main characteristics proven by burn specialist in determining the depth of a burn wound [8]. The descriptors chosen in the work of Acha et al. (2005) were: mean of lightness (L*), mean of hue (h), mean of chroma (c), standard deviation of lightness, standard deviation of hue, standard deviation of chroma, mean of u*, mean of v*, standard deviation of u*, standard deviation of v*, skewness of lightness, kurtosis of lightness, skewness of u*, kurtosis of u*, skewness of v* and kurtosis of v* [8], [10], [11], [17]. The optimum set of features after applying the descriptor selection method were: lightness, hue, standard deviation of the hue component, u* chrominance component, standard deviation of the v* component, and skewness of lightness.

Wantanajittikul et al. [12] focused on extracting colour and texture features. The features selected were: mean of hue (h), the standard deviation of hue (h), the contrast and the homogeneity.

Deepak et al. and Suvarna et al. [13, 14, 15] focused on extracting colour features such as the mean and $(2,1)^{th}$ coefficient of Discrete Cosine Transform (DCT) function of V1 chrominance plane of the L*a*b* colour space.

The features extracted by Tran et al. [16] were the multi-colour channels Red, Green, Blue, and Gray, which were considered as fast feature extraction with regards to real time processing speed. The multi-colour channels were converted to binary value to improve the performance of machine learning.

In the classification of MRI brain images, wavelet transformation method was used for feature extraction. Wavelet transformation has the property of multi-resolution analytic, thus it is capable of analyzing images at various levels of resolution. The problems with using this were the requirement of a large storage space and its expensive computational cost. Thus, principal component analysis (PCA) was used instead to reduce the dimensions of the feature vector to increase the discriminative power [5].

Sawakare et al. [6] used Daubechies Wavelet Transform (DWT) for feature extraction in MRI brain tumor images. They chose DWT is because DWT provide a good contrast to an image. They also make a comparison between Discrete Cosine Transform (DCT) and DWT by stating that DWT provide a better image quality than DCT at the higher threshold value while at the lower threshold value, both DCT and DWT have the same performance. The DWT transform the image into four sub band, which are LL, LH, HH, and HL images. DWT calculation only performed on LL sub band image. Gray Level Co-occurrence Matrix (GLCM) was used for texture feature extraction. In their work, statistical texture features such as contrast, correlation, energy, homogeneity and entropy were obtained at the first five levels of wavelet decomposition of LH and HL sub bands. PCA was used to reduce the dimensions and the computational complexity as well as to extract the best feature. The feature extraction approach proposed were: (i) Feature extraction by using DWT, (ii) Texture feature extraction by using GLCM and (iii) Feature selection by using PCA.

D. Image Classification Methods

The classifier used by Acha et al. [8] for classification of burn depth was a Fuzzy-ARTMAP neural network. This network was developed by Grossberg and Carpenter [19] that was based on Adaptive Resonance Theory. Fuzzy-ARTMAP is used because it has understandable theoretical properties, is efficient to implement, and has clustering properties that are similar to human perception. It is also successfully used in industrial and medical applications. Apart from that, the small number of design parameters of this network, as well as the initial and architecture values are always consistent [8]. This network was tested on 62 burn images, with an average success percentage of 82.26%. The percentage of misclassifications was 55%, in which the images were classified as superficial dermal types when they were actually deep dermal burn, and vice versa [8], [10]. Serrano et al. [11] tested on 35 burn images with the same descriptor and Fuzzy-ARTMAP neural network. Their work yielded an average success percentage of 88.6%. An average success percentage of 88.89% was obtained when they tested 18 times on 16 images and two of them presented two different depths in another work of Acha et al. [17].

The classifier used in the work of Wantanajittikul et al. [12] was support vector machine (SVM). The results were compared with that from Bayes and kNN classifiers. The best result was obtained by SVM on the validation sets of 4-fold cross validation with 89.29% whereas for the blind test experiment, a correct classification of 75.33% was obtained.

The classifiers used by Deepak et al. [13] and Suvarna et al. [15] for the classification of skin burn grades were Template Matching (TM), nearest neighbor classifier (kNN) and Support Vector Machine (SVM). SVM produced the best results of 90% for both works among the three classifiers, with template matching method yielding an efficiency of 66% while kNN classifier yield an efficiency of 75%. In another work of Suvarna et al. [14], they used Template Matching (TM), k Nearest Neighbor Classifier (kNN) and Artificial Neural Network (ANN) to compare the performance of these classifiers on skin burn images. ANN produced the best results by yielding 95% for Grade 1 (Superficial) burn, 97.5% for Grade 2 (Partial Thickness) burns and 95% for Grade 3 (Full Thickness) burns as compared to TM, which yielded 55%, 72.5% and 70% for Grade 1, Grade 2, and Grade 3 respectively and for kNN which produced 67.5%, 82.5%, and 75% respectively.

The classifier used by Tran et al. [16] for burn image classification was one-class SVM instead of the traditional SVM due to the imbalance degrees of burn data available. The best classification results obtained using one-class SVM with polynomial kernel was an accuracy of 77.78% compared to using SVM with polynomial kernel which had an accuracy of 73.73%.

Zhang et al. [5] used kernel support vector machine (KSVM) with K-fold stratified cross validation for the classification of MRI brain tumor images. They tested with four different kernels which are LIN, HPOL, IPOL, and GRB. They obtained the highest classification accuracy of 99.38% with GRB kernel as compared to 95%, 96.88%, and 98.12% for LIN, HPOL, and IPOL respectively.

Sawakare et al. [6] used Probabilistic Neural Network (PNN) for the classification of MRI brain tumor images.

They yield a maximum recognition rate of 100% for the classification method.

III. IMAGE MINING APPROACH

This work proposes to use an image mining approach to evaluate the image of a skin burn injury and classify the burn injury into one of the burn depths. Image mining is not just an extension of data mining to image domain. It is an interdisciplinary field with a combination of techniques such as computer vision, image processing, image retrieval, data mining, machine learning, database and artificial intelligence [20]. Figure 2 shows the image mining approach that is used in this work. The image mining approach consists of several processes.

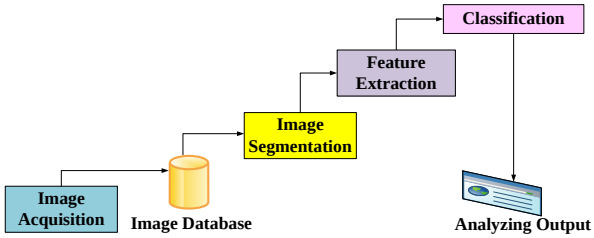

Figure 2: Image mining approach

The first process is image acquisition whereby a sample of images are collected and stored in the image database. For every image, the segmentation algorithm will detect and segment the skin burn wound by removing the background information and healthy skin region. The segmented image undergoes various transformation and image features are extracted to represent the corresponding image content. The extracted features then act as an input for various classifiers to classify. The final step is to analyze and compare the different classification results obtained and identifying the best algorithms for all the three burn depths. The following sub-sections will discuss in depth on each of the processes in the image mining approach.

A. *Image Acquisition*

The burn images used in this work were collected by burn specialist. The skin burn depths considered in this work are second degree burn and third degree burn. The burn images are categorized into superficial partial thickness (SPT) burn, deep partial thickness (DPT) burn and full thickness (FT) burn. The total images collected are shown in Table 1.

Table 1
Total Images Collected

Burn Depth	Total Images Collected
Superficial Partial Thickness Burn	82
Deep Partial Thickness Burn	48
Full Thickness Burn	34
Total	164

B. *Image Segmentation*

A hybrid segmentation method was used, which hybrided the method used by Saranya et al. [18] for separating the healthy skin region from the background information in an image and Otsu's method of thresholding for segmentation of tumor in brain images used by Otsu [7] and Manu [21].

The segmentation process consists of two parts: the body part outline detection and the burn wound detection. The body part outline detection finds the largest connected component in which the colour of the component is closest to the standard skin colour. The burn wound detection is based on the body part outline detection results. If the body part outline detection results is correct, the body part area is then marked as 'white' and the rest of the background image is marked as 'black' in a binary image. This way, it is easier to identify the burn wound located within the body part region.

The body part outline detection uses RGB colour space to discover the colour of connected components that is close to the skin colour. Many other experiments had been carried out in finding the suitable colour space, for example, HSV and YCbCr colour spaces, which are able to detect the body part outline that is closest to the standard skin colour. However, the best results were obtained using the RGB colour space.

The threshold value that was used to segment the skin regions based on RGB colour space is shown in Equation 1 below [22]:

$$\begin{cases} R > 95 \;\&\; G > 40 \;\&\; B > 20 \\ \max(R, G, B) - \min(R, G, B) > 15 \\ |R - G| > 15 \;\&\; R > G \;\&\; R > B \end{cases} \quad (1)$$

where: R= Red channel
G= Green channel
B= Blue channel

The segmented body part region was converted from RGB to CIE L*a*b colour space. After that, the a* component is chosen to be used for burn wound determination due to its high intensity compared to the b* component as shown in Figure 3, 4 and 5.

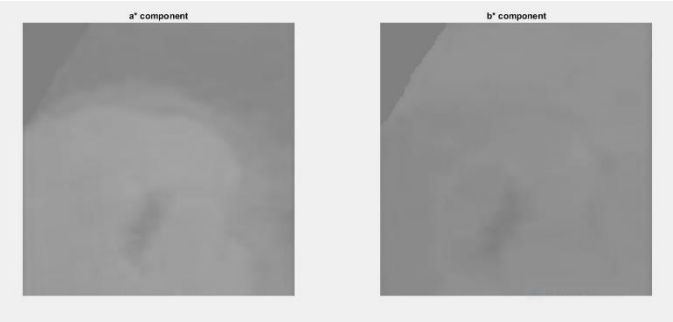

Figure 3: Comparison between a* and b* components in superficial partial thickness burn image

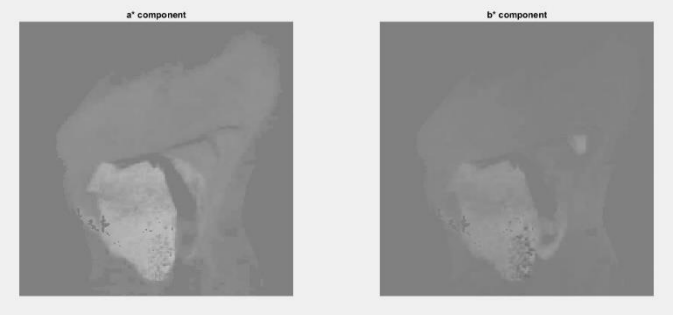

Figure 4: Comparison between a* and b* components in deep partial thickness burn image

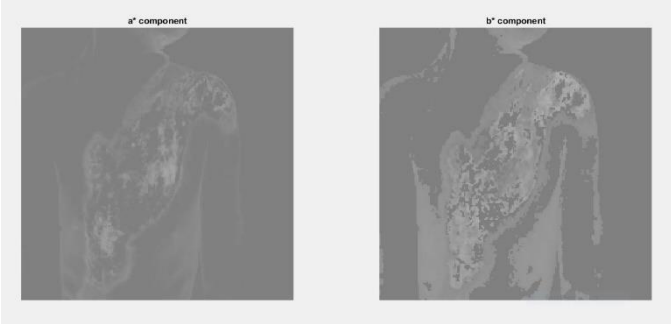

Figure 5: Comparison between a* and b* components in full thickness burn image

The Otsu's method of thresholding is then used to segment the burn wound from the body region [7]. The post-processing is performed to smooth the segmented burn wound regions by filling the small holes, which are the small gaps found in between the segmented burn wound. The algorithm for the hybrid segmentation method used is presented as follows:

1. BEGIN
2. INPUT: colour skin burn image
3. Resize the image to 200 x 200 pixels
4. IF size (*image*, 3) > 1
   final_image = zeros of size (*image*, 1) and size (*image*, 2);
   FOR *i* = 1: size (*image*, 1)
   FOR *j* = 1: size (*image*, 2)
   SET *R* variable = *image* (*i*, *j*, 1);
   SET *G* variable = *image* (*i*, *j*, 2);
   SET *B* variable = *image* (*i*, *j*, 3);
   IF *R* > 95 & *G* > 40 & *B* > 20
   IF max(*R*, *G*, *B*) – min(*R*, *G*, *B*) > 15
   IF |*R* - *G*| > 15 & *R* > *G* & *R* > *B*
   SET *final_image* variable to '1'
   ENDIF
   ENDIF
   ENDIF
   ENDFOR
   ENDFOR
5. ENDIF
6. Convert segmented body part region from RGB to CIE L*a*b colour space
7. Extract a*, in which a* = lab (:, :, 2);
8. Convert a* matrix to grayscale
9. Apply graythresh(), an Otsu's method of thresholding on the grayscale
10. The burn wound is segmented
11. Apply imdilate() to dilate the burn wound
12. Apply imfill() to fill the small holes in the burn wound
13. Apply imerode() to erode the burn wound
14. END

The burn wound detection can work well with the following assumptions. First, the burn image contains little irrelevant background information and it is even better if the background colour is different from the skin colour. Second, the healthy skin of any body part should present a nearly uniform colour feature. Third, the burn wound is within the skin region.

C. Feature Extraction

After the burn wound had been successfully segmented, feature extraction was performed. Both colour and texture were used in combination as features to evaluate the different burn depths. The colour feature extraction was done by finding the statistical colour moments, for example, mean, standard deviation, skewness and so on for each coordinate of the L*a*b* colour spaces as well as the derived hue and chroma image plane [8].

For texture feature extraction, discrete wavelet transform (DWT) was used, followed by principle component analysis (PCA) which reduced the feature dimensions [5]. The skin burn images in the dataset comprised many different qualities and resolution levels. Therefore, to extract all the significant features from the skin burn images with various qualities, both DWT and PCA were used in this work. After the feature reduction by PCA, Gray Level Co-occurrence Matrix (GLCM) method was applied to extract the statistical texture features [6], [21]. The extracted features are mean of lightness, mean of hue, standard deviation of hue, standard deviation of A* component, standard deviation of B* component, and skewness of lightness for colour [8] and contrast, correlation, energy, homogeneity, mean, entropy, smoothness, kurtosis, skewness and inverse difference moment (IDM) [21] for texture.

Feature vectors for the skin burn images in three different burn depths were formed consecutively to be used as input to train the classifier. The feature vectors for test images were also formed in the same way as in feature vectors for train images.

D. Burn Depth Classification

The performance of different classification algorithms was compared using a machine learning workbench, the Waikato Environment for Knowledge Analysis (WEKA) [23]. 20 classification algorithms were used on the skin burn dataset for this comparative study. The different classification algorithms used are briefly described in Table 2.

Classification of the skin burn images was carried out using the binary classification approach, by taking one class versus all the other classes. The test methods used are supplied test set and 10-fold cross validation. The supplied test set method was performed by splitting the dataset into two sets, 70% and 30% for training and testing respectively.

The metrics used to evaluate the performance of the classifiers were accuracy, precision, recall and F-measure. The values for precision, recall and F-measure were recorded by taking the weighted average.

IV. RESULTS AND DISCUSSION

A. Image Segmentation

Table 3 shows the total images that were correctly segmented which were then used for feature extraction and classification. Figures 6, 7 and 8 show a sample segmentation result for each of the three different burn depths. The results showed that segmentation of the burn wound were correctly performed.

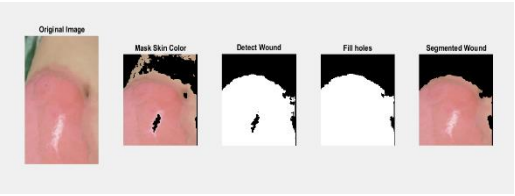

Figure 6: Segmentation result for superficial partial thickness burn

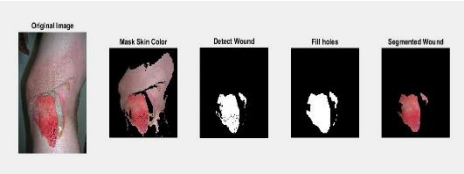

Figure 7: Segmentation result for deep partial thickness burn

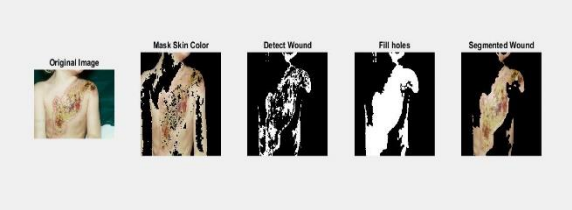

Figure 8: Segmentation result for full thickness burn

Table 2
Different Classification Algorithms used in Weka [23]

Classification Algorithms	Description
BayesianLogisticRegression	Bayesian Logistic Regression for both Gaussian and Laplace Priors is implemented.
NaiveBayes	Estimator classes is used and the training data are analyzed to choose the estimator precision values.
Logistic	A multinomial logistic regression model with a ridge estimator is built and used by this classifier.
Simple Logistic	Logistic regression models is built using this classifier. LogitBoost is used as base learners to fit the logistic models.
SMO	This classifier was implemented by John Platt's sequential minimal optimization algorithm for training a support vector classifier.
VotedPerceptron	This classifier was implemented by Freund and Schapire, which globally replaces all the missing values, and transforms nominal attributes into binary ones.
KStar	An instance-based classifier, K* which is the class of a test instance is based on the class of those training instances similar to it, as determined by some similarity function. An entropy-based distance function is used which make it differs from other instance-based learners.
LWL (Locally Weighted Learning)	An instance-based algorithm is use to assign instance weights which are then used by a specified WeightedInstancesHandler.
Bagging	A class that reduces variance, and is able to perform classification and regression depending on the base learner.
ClassificationViaClustering	A clusterer is use for classification.
ClassificationViaRegression	Regression methods is used to do classification.
MultiClassClassifier	A meta classifier that handles multi-class datasets with 2-class classifiers.
RandomSubSpace	A decision tree based classifier is constructed that maintains highest accuracy on training data and improves on generalization accuracy as it grows in complexity.
VFI (Voting Feature Intervals)	The intervals are constructed around each class for each attribute, and class counts are recorded for each interval on each attribute.
DecisionTable	A simple decision table majority classifier is build and use by this classifier.
JRip	A classifier that proposed by William W. was implemented with a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER).
OneR	A 1R classifier is built and used by this classifier, which means minimum-error attribute is used for prediction and discretizing numeric attributes.
J48	A pruned or unpruned C4.5 decision tree is generated.
LMT (Logistic Model Trees)	A 'logistic model trees' is built, by which the classification trees with logistic regression functions at the leaves.
RandomForest	A forest of random trees is constructed.

Table 3
Total Images that are Correctly Segmented

Burn Depth	Total Images Collected	Total Images that are Correctly Segmented
Superficial Partial Thickness Burn	82	65
Deep Partial Thickness Burn	48	41
Full Thickness Burn	34	17
Total	164	123

B. Feature Extraction

Table 4 shows the types of features extracted to be used as input to the classifiers. These features were extracted for each class of burn depths for both training and testing images.

Table 4
Features Extracted for Colour and Texture

Type of features	Features extracted
Colour	Mean of lightness, mean of hue, standard deviation of hue, standard deviation of A* component, standard deviation of B* component, and skewness of lightness
Texture	Contrast, correlation, energy, homogeneity, mean, entropy, smoothness, kurtosis, skewness and inverse difference moment (IDM)

C. Classification

Table 5 shows the dataset specification for evaluation using both the supplied test set and 10-fold cross validation methods.

Table 5
Dataset Specification for Supplied Test Set and 10-Fold Cross Validation Methods

Burn Depth	Supplied Test Set Method		10-Fold Cross Validation Method
	Train Set (70%)	Test Set (30%)	
Superficial Partial Thickness Burn	43	22	65
Deep Partial Thickness Burn	24	17	41
Full Thickness Burn	11	6	17
Total	78	45	123

Table 6, 7, and 8 show the classification results of the supplied test set method whereas Table 9, 10 and 11 show the classification results of the 10-fold cross validation method for all the three burn depths.

Based on Table 6, Logistic and MultiClassClassifier showed the best performance using the supplied test set method on superficial partial thickness burn images. Based on Table 7 and 8, for deep partial thickness burn and full thickness burn images respectively, OneR produced the best performance using the supplied test set method.

From Table 9, VFI has the best performance using the 10-fold cross validation method on superficial partial thickness burn images. For Table 10 and 11 respectively, the best performance for deep partial thickness burn images was ClassificationViaClustering and for full thickness burn images, the ClassificationViaRegression produced the best performance.

Table 6
Classification Results of Supplied Test Set for Superficial Partial Thickness Burn

Classification Algorithms	Supplied Test Set (Superficial Partial Thickness Burn)			
	Accuracy	Precision	Recall	F-measure
BayesianLogisticRegression	53.3%	0.5	0.5	0.5
NaiveBayes	53.3%	0.6	0.5	0.5
Logistic	**62.2%**	**0.7**	**0.6**	**0.6**
Simple Logistic	57.8%	0.6	0.6	0.5
SMO	48.9%	0.5	0.5	0.5
VotedPerceptron	46.4%	0.5	0.5	0.4
KStar	51.1%	0.5	0.5	0.5
LWL	46.7%	0.5	0.5	0.5
Bagging	44.4%	0.4	0.4	0.4
ClassificationViaClustering	51.1%	0.5	0.5	0.4
ClassificationViaRegression	46.7%	0.5	0.5	0.5
MultiClassClassifier	**62.2%**	**0.7**	**0.6**	**0.6**
RandomSubSpace	48.9%	0.5	0.5	0.5
VFI	51.1%	0.5	0.5	0.5
DecisionTable	48.9%	0.5	0.5	0.5
JRip	42.2%	0.4	0.4	0.4
OneR	44.4%	0.4	0.4	0.4
J48	42.2%	0.4	0.4	0.3
LMT	57.8%	0.6	0.578	0.5
RandomForest	53.3%	0.5	0.533	0.5

Table 7
Classification Results of Supplied Test Set for Deep Partial Thickness Burn

Classification Algorithms	Supplied Test Set (Deep Partial Thickness Burn)			
	Accuracy	Precision	Recall	F-measure
BayesianLogisticRegression	62.2%	0.6	0.6	0.6
NaiveBayes	60.0%	0.6	0.6	0.6
Logistic	68.9%	0.7	0.7	0.7
Simple Logistic	62.2%	0.6	0.6	0.6
SMO	60.0%	0.5	0.6	0.5
VotedPerceptron	62.2%	0.6	0.6	0.6
KStar	60.0%	0.6	0.6	0.6
LWL	62.2%	0.6	0.6	0.6
Bagging	60.0%	0.5	0.6	0.5
ClassificationViaClustering	64.4%	0.6	0.6	0.6
ClassificationViaRegression	62.2%	0.6	0.6	0.6
MultiClassClassifier	68.9%	0.7	0.7	0.7
RandomSubSpace	60.0%	0.5	0.6	0.5
VFI	62.2%	0.6	0.6	0.6
DecisionTable	60.0%	0.5	0.6	0.5
JRip	60.0%	0.6	0.6	0.6
OneR	**73.3%**	**0.7**	**0.7**	**0.7**
J48	57.8%	0.5	0.6	0.5
LMT	62.2%	0.6	0.6	0.6
RandomForest	57.8%	0.6	0.6	0.6

Table 8
Classification Results of Supplied Test Set for Full Thickness Burn

Classification Algorithms	Supplied Test Set (Full Thickness Burn)			
	Accuracy	Precision	Recall	F-measure
BayesianLogisticRegression	86.7%	0.8	0.9	0.8
NaiveBayes	73.3%	0.8	0.7	0.8
Logistic	75.6%	0.7	0.8	0.7
Simple Logistic	86.7%	0.8	0.9	0.8
SMO	86.7%	0.8	0.9	0.8
VotedPerceptron	86.7%	0.8	0.9	0.8
KStar	73.3%	0.7	0.7	0.7
LWL	86.7%	0.8	0.9	0.8
Bagging	86.7%	0.8	0.9	0.8
ClassificationViaClustering	75.6%	0.7	0.9	0.8
ClassificationViaRegression	84.4%	0.7	0.8	0.8
MultiClassClassifier	75.6%	0.7	0.8	0.7
RandomSubSpace	86.7%	0.8	0.9	0.8
VFI	84.4%	0.9	0.8	0.9
DecisionTable	86.7%	0.8	0.9	0.8
JRip	86.7%	0.8	0.9	0.8
OneR	**88.9%**	**0.9**	**0.9**	**0.9**
J48	86.7%	0.8	0.9	0.8
LMT	86.7%	0.8	0.9	0.8
RandomForest	84.4%	0.7	0.8	0.8

The 10-fold cross validation method takes the average of the different test partitions in the dataset while the supplied test set method uses a fixed test set. These two evaluation methods were experimented to see if using different test sets affect the performance of the classifiers. This was certainly the case as observed in the results presented for each of the three burn depths. Since the 10-fold cross validation method takes the average of all partitions in the dataset, its results would be void of bias and more consistent.

On closer inspection of the misclassifications, there were some superficial partial thickness burn which were misclassified as deep partial thickness burn, and vice versa. The reason of this is because in some burn wounds, two depths of wound are present, which is known as "mix partial thickness burn". When this happens, the burn wound is usually classified into the more serious type of burn depth. However, in this work, for example, some images of deep partial thickness burn was misclassified as superficial partial thickness burn. This was due to the reason that the deep partial thickness burn, usually cream or almost white in colour, was surrounded by superficial partial thickness burn. Therefore, the classifier would recognize it as a superficial thickness burn. It is also possible that the classifier misclassified a burn depth by its surface area. For instance, in a mix partial thickness burn of superficial partial thickness burn and full thickness burn, in which the image should actually be a full thickness burn, the classifier misclassified it as superficial partial thickness due to the fact that the superficial partial thickness burn had a larger surface area than the full thickness burn.

The overall performances for each classifier for all three burn depths are shown in Table 12 and 13 for the supplied test set and 10-fold cross validation methods respectively, taking the average accuracy of the three burn depths. From Table 12, it can be seen that Logistic, Simple Logistic, MultiClassClassifier, OneR and LMT produced the same accuracies and are best in classifying the images in the dataset using the supplied test set method, with an average accuracy of 68.9%. In Table 13, an average accuracy of 73.2% was achieved with Simple Logistic using the 10-fold cross validation method for classifying the images in the dataset. The results indicated that Simple Logistic using the 10-fold cross validation method produced the best classification performance for the skin burn dataset. While it was noted that SVM performed very well in literature, this was not observed in the experiments with Sequential Minimal Optimization (SMO), an algorithm to train a support vector classifier in this work.

Table 9
Classification Results of 10-Fold Cross Validation for Superficial Partial Thickness Burn

Classification Algorithms	10-fold Cross Validation (Superficial Partial Thickness Burn)			
	Accuracy	Precision	Recall	F-measure
BayesianLogisticRegression	58.5%	0.6	0.6	0.6
NaiveBayes	57.7%	0.6	0.6	0.6
Logistic	61.0%	0.6	0.6	0.6
Simple Logistic	64.2%	0.6	0.6	0.6
SMO	57.7%	0.6	0.6	0.6
VotedPerceptron	56.9%	0.6	0.6	0.6
KStar	61.8%	0.6	0.6	0.6
LWL	48.0%	0.5	0.5	0.5
Bagging	50.4%	0.5	0.5	0.5
ClassificationViaClustering	53.7%	0.5	0.5	0.5
ClassificationViaRegression	55.3%	0.6	0.6	0.6
MultiClassClassifier	61.0%	0.6	0.6	0.6
RandomSubSpace	57.7%	0.6	0.6	0.6
VFI	**65.9%**	**0.7**	**0.7**	**0.6**
DecisionTable	49.6%	0.4	0.5	0.4
JRip	53.7%	0.5	0.5	0.5
OneR	56.1%	0.6	0.6	0.6

J48	60.2%	0.6	0.6	0.6
LMT	63.4%	0.6	0.6	0.6
RandomForest	54.5%	0.5	0.5	0.5

Table 10
Classification Results of 10-Fold Cross Validation for Deep Partial Thickness Burn

Classification Algorithms	10-fold Cross Validation (Deep Partial Thickness Burn)			
	Accuracy	Precision	Recall	F-measure
BayesianLogisticRegression	69.1%	0.7	0.7	0.6
NaiveBayes	65.9%	0.6	0.7	0.6
Logistic	65.0%	0.6	0.7	0.6
Simple Logistic	69.9%	0.7	0.7	0.7
SMO	67.5%	0.6	0.7	0.6
VotedPerceptron	69.1%	0.7	0.7	0.6
KStar	60.2%	0.6	0.6	0.6
LWL	65.0%	0.6	0.7	0.6
Bagging	65.9%	0.6	0.7	0.6
ClassificationViaClustering	**72.4%**	**0.7**	**0.7**	**0.7**
ClassificationViaRegression	66.7%	0.7	0.7	0.7
MultiClassClassifier	65.0%	0.6	0.7	0.6
RandomSubSpace	67.5%	0.7	0.7	0.6
VFI	67.5%	0.7	0.7	0.7
DecisionTable	66.7%	0.6	0.7	0.5
JRip	62.6%	0.6	0.6	0.6
OneR	64.2%	0.6	0.6	0.6
J48	65.0%	0.6	0.7	0.6
LMT	69.9%	0.7	0.7	0.7
RandomForest	68.3%	0.7	0.7	0.7

Table 11
Classification Results of 10-Fold Cross Validation for Full Thickness Burn

Classification Algorithms	10-fold Cross Validation (Full Thickness Burn)			
	Accuracy	Precision	Recall	F-measure
BayesianLogisticRegression	86.2%	0.7	0.9	0.8
NaiveBayes	53.7%	0.8	0.5	0.6
Logistic	82.9%	0.8	0.8	0.8
Simple Logistic	85.4%	0.8	0.9	0.8
SMO	86.2%	0.7	0.9	0.8
VotedPerceptron	86.2%	0.7	0.9	0.8
KStar	75.6%	0.8	0.8	0.8
LWL	79.7%	0.8	0.8	0.8
Bagging	83.7%	0.7	0.8	0.8
ClassificationViaClustering	70.7%	0.7	0.8	0.8
ClassificationViaRegression	**87.0%**	**0.8**	**0.9**	**0.8**
MultiClassClassifier	82.9%	0.8	0.8	0.8
RandomSubSpace	86.2%	0.7	0.9	0.8
VFI	64.2%	0.8	0.6	0.7
DecisionTable	82.1%	0.7	0.8	0.8
JRip	78.9%	0.8	0.8	0.8
OneR	83.7%	0.8	0.8	0.8
J48	80.5%	0.8	0.8	0.8
LMT	85.4%	0.8	0.9	0.8
RandomForest	84.6%	0.8	0.8	0.8

V. CONCLUSION

A comparative study of the classification of skin burn depth in human was conducted using an image mining approach. A hybrid segmentation method was implemented, in which the method would first separate the body part or skin region from the background of the image. After successfully segmenting the body part, the method would detect the burn wound within the skin region. Both colour and texture features were then extracted and used in combination. Discrete wavelet transform (DWT) and principle component analysis (PCA) were used for feature extraction. Gray Level Co-occurrence Matrix (GLCM) method was used to extract texture features. There were a total of 20 classification algorithms used to classify the skin burn dataset comprising of three different burn depths, and the performance of these classifiers were measured by accuracy, precision, recall, and F-measure. Two test methods were used in the experiments, which were the supplied test set and 10-fold cross validation methods. The classification was carried out according to the binary classification approach, by taking one class versus all other classes. The best classification algorithm that was able to classify most of the three burn depths was Simple Logistic using the 10-fold cross validation method, producing an average accuracy of 73.2%. Some observations on misclassifications provided insights for future improvement of the image mining approach. In future, multi-class classification will be conducted to see how the compared classifiers perform.

Table 12
Average Accuracies for All Three Burn Depths Using the Supplied Test Set Method

Classification Algorithms	Supplied Test Set			
	Superficial Partial Thickness Burn	Deep Partial Thickness Burn	Full Thickness Burn	Average Accuracy
BayesianLogisticRegression	53.3%	62.2%	86.7%	67.4%
NaiveBayes	53.3%	60.0%	73.3%	62.2%
Logistic	**62.2%**	**68.9%**	**75.6%**	**68.9%**
Simple Logistic	**57.8%**	**62.2%**	**86.7%**	**68.9%**
SMO	48.9%	60.0%	86.7%	65.2%
VotedPerceptron	46.4%	62.2%	86.7%	65.1%
KStar	51.1%	60.0%	73.3%	61.5%
LWL	46.7%	62.2%	86.7%	65.2%
Bagging	44.4%	60.0%	86.7%	63.7%
ClassificationViaClustering	51.1%	64.4%	75.6%	63.7%
ClassificationViaRegression	46.7%	62.2%	84.4%	64.4%
MultiClassClassifier	**62.2%**	**68.9%**	**75.6%**	**68.9%**
RandomSubSpace	48.9%	60.0%	86.7%	65.2%
VFI	51.1%	62.2%	84.4%	65.9%
DecisionTable	48.9%	60.0%	86.7%	65.2%
JRip	42.2%	60.0%	86.7%	63.0%
OneR	**44.4%**	**73.3%**	**88.9%**	**68.9%**
J48	42.2%	57.8%	86.7%	62.2%
LMT	**57.8%**	**62.2%**	**86.7%**	**68.9%**
RandomForest	53.3%	57.8%	84.4%	65.2%
Average	50.7%	62.3%	83.5%	65.5%

Table 13
Average Accuracies for All Three Burn Depths Using the 10-Fold Cross Validation Method

Classification Algorithms	10-fold Cross Validation			
	Superficial Partial Thickness Burn	Deep Partial Thickness Burn	Full Thickness Burn	Average Accuracy
BayesianLogisticRegression	58.5%	69.1%	86.2%	71.3%
NaiveBayes	57.7%	65.9%	53.7%	59.1%
Logistic	61.0%	65.0%	82.9%	69.6%
Simple Logistic	**64.2%**	**69.9%**	**85.4%**	**73.2%**
SMO	57.7%	67.5%	86.2%	70.5%
VotedPerceptron	56.9%	69.1%	86.2%	70.7%
KStar	61.8%	60.2%	75.6%	65.9%
LWL	48.0%	65.0%	79.7%	64.2%
Bagging	50.4%	65.9%	83.7%	66.7%
ClassificationViaClustering	53.7%	72.4%	70.7%	65.6%
ClassificationViaRegression	55.3%	66.7%	87.0%	69.7%
MultiClassClassifier	61.0%	65.0%	82.9%	69.6%
RandomSubSpace	57.7%	67.5%	86.2%	70.5%
VFI	65.9%	67.5%	64.2%	65.9%
DecisionTable	49.6%	66.7%	82.1%	66.1%
JRip	53.7%	62.6%	78.9%	65.1%
OneR	56.1%	64.2%	83.7%	68.0%
J48	60.2%	65.0%	80.5%	68.6%
LMT	63.4%	69.9%	85.4%	72.9%
RandomForest	54.5%	68.3%	84.6%	69.1%
Average	57.4%	66.7%	80.3%	68.1%

REFERENCES

[1] *Boundless Anatomy and Physiology* in *Structure of the Skin: Dermis*. 2016. [Online]. Available: https://www.boundless.com/physiology

/textbooks/boundless-anatomy-and-physiology-textbook/integumentary-system-5/the-skin-64/structure-of-the-skin-dermis-395-7489/.

[2] M. Rani and M. G. Schwacha, "Aging and the pathogenic response to burn," *Aging and Disease*, vol. 3, no. 2, pp. 171–180, Apr. 2012.

[3] "Burn Classification," in *UNM hospitals*. [Online]. Available: http://hospitals.unm.edu/burn/classification.shtml. Accessed: Jan. 7, 2017.

[4] M. S. Badea, C. Vertan, C. Florea, L. Florea, and S. Badoiu, "Automatic burn area identification in color images," *International Conference on Communications (COMM)*, Institute of Electrical and Electronics Engineers (IEEE), 2016, pp. 65–68.

[5] Y. Zhang and L. Wu, "An mr brain images classifier via principal component analysis and kernel support vector machine," *Progress In Electromagnetics Research*, vol. 130, pp. 369–388, 2012.

[6] S. Sawakare and D. Chaudhari, "Classification of brain tumor using discrete wavelet transform, principal component analysis and probabilistic neural network," *International Journal for Research in Emerging Science and Technology*, vol. 1, no. 6, pp. 13–19, Nov. 2014.

[7] N. Otsu, "A threshold selection method from gray-level histograms," in *IEEE Transactions on Systems, Man, and Cybernetics*, Institute of Electrical and Electronics Engineers (IEEE), 1979, vol. 9, no. 1, pp. 62–66.

[8] B. Acha, C. Serrano, J. I. Acha, and L. M. Roa, "Segmentation and classification of burn images by color and texture information," *Journal of Biomedical Optics*, vol. 10, no. 3, pp. 034014–03401411, 2005.

[9] C. Serrano, L. Roa, and B. Acha, "Evaluation of a telemedicine platform in a burn unit," *Information Technology Applications in Biomedicine, 1998. ITAB 98. Proceedings. 1998 IEEE International Conference*, pp. 121–126, May 1998.

[10] B. Acha, C. Serrano, J. I. Acha, and L. M. Roa, "CAD tool for burn diagnosis," *Biennial International Conference on Information Processing in Medical Imaging*, pp. 294–305, Jul. 2003.

[11] C. Serrano, B. Acha, T. Gómez-Cía, J. I. Acha, and L. M. Roa, "A computer assisted diagnosis tool for the classification of burns by depth of injury," *Burns*, vol. 31, no. 3, pp. 275–281, May 2005.

[12] K. Wantanajittikul, S. Auephanwiriyakul, N. Theera-Umpon, and T. Koanantakool, "Automatic segmentation and degree identification in burn color images," in *The 4th 2011 Biomedical Engineering International Conference (BMEiCON)*, Institute of Electrical and Electronics Engineers (IEEE), 2012, pp. 169–173.

[13] L. Deepak, J. Antony, and C. Niranjan U, "Hardware Co-Simulation of skin burn image analysis," in *19th IEEE International Conference in High Performance Computing (HiPC-2012): Student Research Symposium. Pune, India*.

[14] M. Suvarna, K. Kumar, Sivakumar, and N. U. C, "Diagnosis of burn images using template matching, k-nearest neighbor and artificial neural network," *International Journal of Image Processing (IJIP)*, vol. 7, no. 2, 2013.

[15] M. Suvarna, S. kumar, and N. U. C, "Classification methods of skin burn images," *International Journal of Computer Science and Information Technology*, vol. 5, no. 1, pp. 109–118, 2013.

[16] H. Tran, T. Le, and T. Nguyen, "Burn image classification using One-Class support vector machine," in *International Conference on Context-Aware Systems and Applications*, Springer International Publishing, 2015, pp. 233–242.

[17] B. Acha, C. Serrano, and J. I. Acha, "Segmentation of burn images using the L*u*v* space and classification of their depths by color and texture imformation," *Medical Imaging 2002: Image Processing*, pp. 1508–1515, May 2002.

[18] G. Saranya, S. S. Thasny, and K. Sobhia, "Parallel implementation of wound image analysis system for diabetic patient," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 5, no. 11, pp. 112–118, May 2016.

[19] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 698–713, 1992.

[20] R. Sudhir, "A survey on image mining techniques: Theory and applications," *Computer Engineering and Intelligent Systems*, vol. 2, no. 6, pp. 44–52, Oct. 2011.

[21] B. N. Manu, "Brain MRI Tumor Detection and Classification," in *MathWorks*, 2016. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/55107-brain-mri-tumor-detection-and-classification.

[22] J. Kovac, P. Peer, and F. Solina, "Human skin color clustering for face detection," in *The IEEE Region 8 EUROCON 2003. Computer as a Tool*, Institute of Electrical and Electronics Engineers (IEEE), 2003.

[23] *WEKA The University of Wailkato*. [Online]. Available: http://www.cs.waikato.ac.nz/~ml/weka/. Accessed: Jan. 26, 2017.