Multi-Class Multi-Level Classification Algorithm for Skin Lesions Classification using Machine Learning Techniques

Nazia Hameed[a,b,*], Antesar M. Shabut[c], Miltu K. Ghosh[b,d], M. A. Hossain[e]

[a] School of Computer Science, University of Nottingham
[b] Anglia Ruskin IT Research Institute (ARITI), Anglia Ruskin University, Chelmsford, UK
[c] School of Arts and Communication, Leeds Trinity University, Leeds, UK
[d] Department of Pharmacy, NSHM Knowledge Campus, India
[e] Healthcare Innovation Centre, Teesside University, UK

Emails:
Nazia Hameed: nazia.hameed@nottingham.ac.uk
Antesar Shabut: a.shabut@leedstrinity.ac.uk
Miltu K. Ghosh: miltu.pharm@gmail.com
M. A. Hossain: a.Hossain@tees.ac.uk
***Corresponding author:**
Nazia Hameed
Ingenuity Lab, Computer Science Building, Wollaton Road, Nottingham NG8 1BB
Phone: + 44 7740 311828

Abstract

Skin diseases remain a major cause of disability worldwide and contribute approximately 1.79% of the global burden of disease measured in disability-adjusted life years. In the United Kingdom alone, 60% of the population suffer from skin diseases during their lifetime. In this paper, we propose an intelligent digital diagnosis scheme to improve the classification accuracy of multiple diseases. A Multi-Class Multi-Level (MCML) classification algorithm inspired by the "divide and conquer" rule is explored to address the research challenges. The MCML classification algorithm is implemented using traditional machine learning and advanced deep learning approaches. Improved techniques are proposed for noise removal in the traditional machine learning approach. The proposed algorithm is evaluated on 3672 classified images, collected from different sources and the diagnostic accuracy of 96.47% is achieved. To verify the performance of the proposed algorithm, its metrics are compared with the Multi-Class Single-Level classification algorithm which is the main algorithm used in most of the existing literature. The results also indicate that the MCML classification algorithm is capable of enhancing the classification performance of multiple skin lesions.

Keywords: *skin lesion classification, computer-aided diagnosis, machine learning, deep learning, texture & colour features, melanoma classification, eczema classification*

1. Introduction

The human skin is the largest body organ and can be agonised from different factors like sun (Ultraviolet) radiations, tanning, lifestyle, smoking, alcohol usage, physical activities, viruses and working environment (Jaworek-Korjakowska & Kleczek, 2018; Salem, Azar, & Tokajian, 2018). These factors compromise its integrity and have a profound, devastating impact on its well-being. Illnesses that directly affect the skin is the fourth most frequent cause of all human diseases, affecting almost one-third of the world's population around 1.9 billion people at a time (British Association of Dermatologists, 2015), hence generating research interest across many different disciplines. Skin diseases contributed to approximately 1.79% of the global burden of diseases measured in disability-adjusted life years (DALYs) (Karimkhani, Dellavalle, & Coffeng, 2017). In the United Kingdom, 60% of the population suffer from skin diseases during their lifetime (British Skin Foundation, 2018). Skin diseases may be cancerous, inflammatory or infectious and affect people of all ages, especially the elderly and young children (British Association of Dermatologists, 2015). There are severe consequences of skin diseases such as death (in the case of melanoma), impairment of daily activities, loss of relationships, and damage to internal organs. Moreover, they also pose a real threat of mental illness leading to isolation, depression and even

suicide (Picardi, 2013). To decrease the associated consequences, cost, mortality and morbidity rate, skin diseases should be treated in their initial stages. Cancer and eczema are among the top five common skin disorders, according to Dr Macrene Alexiades-Armenakas (Fox News, 2015). Therefore, our main focus is to develop an intelligent digital diagnosis scheme that can diagnose and classify these diseases.

Melanoma, a type of skin cancer is caused by uncontrolled growth of melanin in the melanocyte's cells. It is the most common and rapidly increasing type of cancer. Melanoma is commonly classified into two types, benign and malignant melanoma (Nasir et al., 2018). In benign lesions (common nevi) melanin is normally present in the epidermis layer. Melanin is reproduced at a high abnormal stage in the malignant lesions. Malignant lesions are not life-threatening if the melanocytes and their associated cells remain in the epidermis layer but when they penetrate in the dermis and leave their deposits, the nature of the skin colour changes, and it became dangerous (Hameed, Hassan, & Hossain, 2016; Maglogiannis & Doukas, 2009). According to the World Health Organization, between 2 and 3 million non-melanoma skin cancers and 132,000, melanoma skin cancers occur globally each year (World Health Organization, 2018).

Approximately 99,550 cases are diagnosed in the USA, and approximately 13,460 are fatal (Siegel, Miller, & Jemal, 2018). However, melanoma is the most treatable cancer if detected at early stages. If skin cancer is detected in stage 1, the survival rate is almost 96%, whereas it is decreased to only 5%, if detected at stage-IV (Freedberg, Geller, Miller, Lew, & Koh, 1999; Nasir et al., 2018). Due to its life-threatening nature, it has gained remarkable attention from research and healthcare community, and their ultimate goal is to diagnose it in the early stages. However, it is challenging due to similarities in melanocytic and non-melanocytic skin lesions. Eczema is an inflammatory disease and caused by many factors. In the literature, most of the work is done on skin cancer classification, and limited work is available on the classification of other diseases.

In this research work, a multi-class multi-level (MCML) algorithm is proposed and developed to provide a multi-class classification of skin diseases. In MCML, the skin lesion classification problem is divided into sub-problems, and these sub-problems are solved in multiple steps instead of only one step to improve the classification performance. The MCML algorithm is implemented using two approaches: the traditional machine learning approach and the deep learning approach. In the traditional machine learning approach, improved techniques are proposed for removing black frames and circles which is another contribution of this research work. A universal set of features that can be extracted from every disease is also listed. To demonstrate the improved performance of the proposed algorithm, results are compared with the multi-class single-level algorithm. Comparison of traditional machine learning and deep learning approach is also performed in this research work. The remainder of the paper is organised as follows. Section 2 reviews the existing literature. Material and dataset are given in section 3, whilst methods are discussed in section 4. Experiments, results and limitations are explained in section 5. Section 6 concludes the research and highlight future directions.

2. Related Work

Over the last two decades, researchers have worked to provide intelligent diagnosis systems for the automated classification of skin disorders to assist the dermatologists, primarily in the early classification of skin cancer (E.Umbaugh, H.Moss, & V.Stoecker, 1992; Ercal, Moganti, Stoecker, & Moss, 1993; Nischik & Forster, 1997; Zhang, Stoecker, & Moss, 2000; Vasconcelos & Vasconcelos, 2017; Dorj, Lee, Choi, & Lee, 2018;). With the advancement of computer vision and image processing, continuous improvement is required to provide better accuracy. In the literature, intelligent diagnosis systems have been developed using a traditional machine learning approach (De Guzman, Maglaque, Torres, Zapido, & Cordel, 2015; De Guzman et al., 2015; Alam et al., 2016; Oliveira, Pereira, Manuel, & Tavares, 2016; Zakeri & Hokmabadi, 2018) and, more recently, an advanced deep learning approach (Esteva et al., 2017; Hameed, Shabut, & Hossain, 2018b; Vasconcelos & Vasconcelos, 2017).

Intelligent diagnosis systems based on traditional machine learning techniques mostly consist of pre-processing, segmentation, feature extraction and classification phases. In pre-processing, images are pre-processed to remove

the noise and to improve the segmentation accuracy (Oliveira, Marranghello, Pereira, & Tavares, 2016; Rosado, Jo, & Vasconcelos, 2015). There are many factors that cause noise to become part of images. These include the capturing environment, the capturing device, and lightening condition. All may affect the images in the form of a black frame, dermoscopic gel, air bubbles, circles, skin lines, hairs, and blood vessels (Hameed et al., 2016; Maglogiannis & Doukas, 2009). Pre-defined masks are used by Sultana et al. to remove the black frames (Sultana, Dumitrache, Vocurek, & Ciuc, 2014). Similar work is done by Abuzaghleh et al. to remove the black frame (Abuzaghleh, Barkana, & Faezipour, 2014). However, these techniques can only work when the dataset is small and consistent. To remove the hairs, DullRazor (Lee, Ng, Gallagher, Coldman, & McLean, 1997) is the widely used technique in literature, but it can detect only thick hairs. Several approaches have been used in the literature for noise removal and to enhance the quality of images include image resizing (Jain, Jagtap, & Pise, 2015), contrast adjustment (Rosado et al., 2015) , filtering (Abbas, Celebi, & Fondón, 2011; Celebi, Aslandogan, & Bergstresser, 2005; Maglogiannis, Ieee, & Delibasis, 2015), cropping and colour quantization (Celebi et al., 2005). Once the noise has been removed, the next step is the segmentation to extract the region of interest (ROI). Image segmentation techniques have been developed based on several techniques such as thresholding, clustering, region based, deformable model, soft computing and, pattern recognition to get the ROI (Ma, Tavares, & Jorge, 2009; Ma, Tavares, Jorge, & Mascarenhas, 2010; Ferreira, Gentil, & Tavares, 2014; Jodas, Pereira, & Tavares, 2016; Ma & Tavares, 2016; Oliveira, Pereira, Manuel, & Tavares, 2016;). Among these techniques, thresholding is the most popular because of its simple nature. Clustering such as k-means is also used, but it requires identifying the number of K before applying it (Ma et al., 2009). However, a single technique may not work, and in turn, a hybrid technique works better when dealing with images of a diverse nature. After identifying the ROI, a number of features are extracted to help perform the final classification. Geometric and shape features (Ma & Tavares, 2017; Oliveira, Pereira, & Tavares, 2018; Taufiq, Hameed, Anjum, & Hameed, 2017), colour features (J. Tavares & Jorge, 2012; Filho, Ma, & Tavares, 2015; Ma & Tavares, 2017; Nasir et al., 2018; Oliveira, Papa, Pereira, & Tavares, 2018; Oliveira, Pereira, et al., 2018;) and texture features (J. Tavares & Jorge, 2012;Oliveira et al., 2018; Zakeri & Hokmabadi, 2018) are all widely used for performing the classification. However, all these features cannot be extracted from every skin lesion because of many different factors such as disease nature, presence of moles, area affected by the disease, capturing distance etc. (Salem et al., 2018). For the classification process, several machine learning classifiers have been applied to the extracted features to achieve the best results. Frequently used techniques for skin lesion classification are support vector machine (Alam et al., 2016; C.-Y. Chang & Liao, 2011; Hameed et al., 2016; Hameed, Shabut, & Hossain, 2018a; Taufiq et al., 2017), artificial neural networks (ANN) (Rubegni et al., 2002; Shamsul Arifin, Golam Kibria, Firoze, Ashraful Amini, & Yan, 2012), k-nearest neighbours (Çataloluk & Kesler, 2012; Ganster et al., 2001) and decision trees (Salem et al., 2018; Victor & Ghalib, 2017).

Using the traditional machine learning approach, most of the research is done on the classification of skin cancer and limited research is performed on the classification of other diseases. Out of this limited research, some intelligent diagnosis systems are trained on the clinically extracted features instead of images (C. L. Chang & Chen, 2009; Güvenir & Emeksiz, 2000; Kumar, Kumar, & Saboo, 2016; Übeyli, 2009; Xie & Wang, 2011). Erythemato-squamous diseases consisting psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris are classified by the Derya et al. using SVM and neural network with an accuracy of 98.32% and 97.7% respectively (Übeyli, 2008, 2009). Similar kind of work is done by Guvenier and Emeksiz where they presented an expert system for classification of Erythemato-squamous diseases by incorporating nearest neighbour, naïve bayesian and voting feature algorithm. Voting feature algorithm outperforms with an accuracy of 99.2% (Güvenir & Emeksiz, 2000). Chang et al. used a hybrid technique comprising of the neural network features and decision trees to construct a predictive model for diagnosing Erythemato-squamous diseases using multi-variate variables. Their proposed predictive model achieves an accuracy of 92.62% (C. L. Chang & Chen, 2009). Abdi and Giveki proposed an automated detection of erythemato-squamous diseases using PSO–SVM based on association rules (Abdi & Giveki, 2013). Xie et al. (Xie & Wang, 2011), Kumar et al. (Kumar et al., 2016) and Nanni et al. (Nanni, 2006) also classified the erythemato-squamous disease using machine learning technique with an accuracy of 98.61%, 97.22% and 95.00% respectively. All of the work done for the classification of Erythemato-squamous diseases is performed on

clinically extracted multivariate features after biopsy (Ilter & Guvenir, 1998). Disease classification using clinically extracted features is not feasible as it is time-consuming and difficult to achieve. Moreover, it requires domain expertise and expert knowledge.

Other limitations found in the literature are single disease classification and training on limited data (Ganster et al., 2001; Nugroho et al., 2013; Giotis et al., 2015; Jain et al., 2015; Alam et al., 2016;). Continuing with the single disease classification, Dorj et al. used SVM classifier on the features obtained by the Convolutional Neural Network (CNN) to classify the skin cancer. They trained and tested their algorithm on 3753 images which were collected from the Internet and achieved an accuracy of 94.2% (Dorj, Lee, Choi, & Lee, 2018). Nasir et al. presented a strategy for the classification of melanoma with an accuracy of 97.5% when tested on the PH[2] dataset (Nasir et al., 2018). Although they have achieved good classification accuracy, their trained model lacks adaptability to new data because of the training on limited data. Zakeri et al. proposed a hybrid classifier for detecting the cancerous lesions with an accuracy of 96.8%, 97.3% and 98.8% for the melanoma, dysplastic and benign lesions on 792 images. Work done by Esteva et al. achieves dermatological level classification using the CNN, but they also worked only on the skin cancer. Vasconcelos et al. also performed experiments with different variations using deep learning for the melanoma image analysis. All the research mentioned above work is performed for classifying different types of skin cancers. Alam et al. presented a model for classification of healthy and eczema images. An accuracy of 90% is obtained when they trained and tested their model on 85 images. Their model also lacks generalisation because of limited data. The comparison of skin diseases classification work available in the literature is given in Table 1 where limitations are highlighted.

Table 1. Summary of studies in literature for skin disease classification

Reference	Classification Categories	Images	Results
Salem et al., 2018	**Melanoma**	369	Acc[*]: 76.17%
Dorj et al., 2018	**Melanoma**	3753	Acc: 94.2%
Oliveira et al., 2018	**Melanoma**	1104	Acc: 92.3%
Nasir et al., 2018	**Melanoma**	200	Acc: 97.7%
Alam et al., 2016	Healthy & Eczema	**85**	Acc: 90%
Übeyli, 2008, 2009	Erythemato-squamous Diseases[1]	**Non-image data**	Acc: 98.32%
Güvenir & Emeksiz, 2000	Erythemato-squamous Diseases	**Non-image data**	Acc: 99.2%
C. L. Chang & Chen, 2009	Erythemato-squamous Diseases	**Non-image data**	Acc: 92.62%
Abdi & Giveki, 2013	Erythemato-squamous Diseases	**Non-image data**	Acc: 98.91
Kumar et al., 2016	Erythemato-squamous Diseases	**Non-image data**	Acc: 98.61%

Acc[*] = Accuracy

---

[1] Erythemato-squamous Diseases includes psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris

3. Dataset

To conduct this research, a combination of images from different sources are used. Sources for collecting dataset include open-access dermatology repositories, organisations and researchers. Open-access dermatology repositories include the International Skin Imaging Collaboration (ISIC) Dermoscopic Archive (Codella et al., 2016) and the PH[2] (Mendonca, Ferreira, Marques, Marcal, & Rozeira, 2013). The ISIC Melanoma Project is an industry and academia partnership designed to facilitate the application of digital skin imaging to help reduce melanoma mortality. The ISIC Archive dataset[2] constitutes 13000 melanocytic lesion images that are biopsy proven and annotated as either benign or malignant. The PH[2] dataset[3] is also a publically available database comprising of 200 images which include the manual segmentation along with the clinical diagnosis performed by expert dermatologists. For healthy dataset, the "11K hand's" dataset is used (Afifi, 2017). Melanoma and eczema images are also obtained from different organisations such as the DermIS (DermIS, 2018), DermQuest (Derm101 Image Library, 2018) and DermNZ (DermNZ, 2018). These organisations provide classified images of different skin lesions, and they are freely available to use for academic purposes. A subset of the dataset is also obtained from Alam et al. (Alam et al., 2016). This dataset contains 85 images belonging to the healthy and eczema classes. The data collection sources for each class is illustrated in tabular form in Table 2.

Table 2. Image categories with their collection source

Class	Source
Healthy	1) 11K dataset (Afifi, 2017) 2) Alam et al. (Alam et al., 2016)
Eczema	1) DermIS 2) DermQuest 3) DermNZ 4) Alam et. al.
Benign	1) PH[2] dataset (PH[2] Database, 2013) 2) ISIC Image challenge (Gutman et al., 2016) 3) DermIS 4) DermQuest
Malignant	1) PH[2] dataset 2) ISIC Image challenge 3) DermIS 4) DermQuest

Some of the images along with their categories are shown in Figure 1.

---

[2] The ISIC dataset is available at: https://isic-archive.com/#images

[3] The PH[2] dataset is publically available at : https://www.fc.up.pt/addi/ph2%20database.html
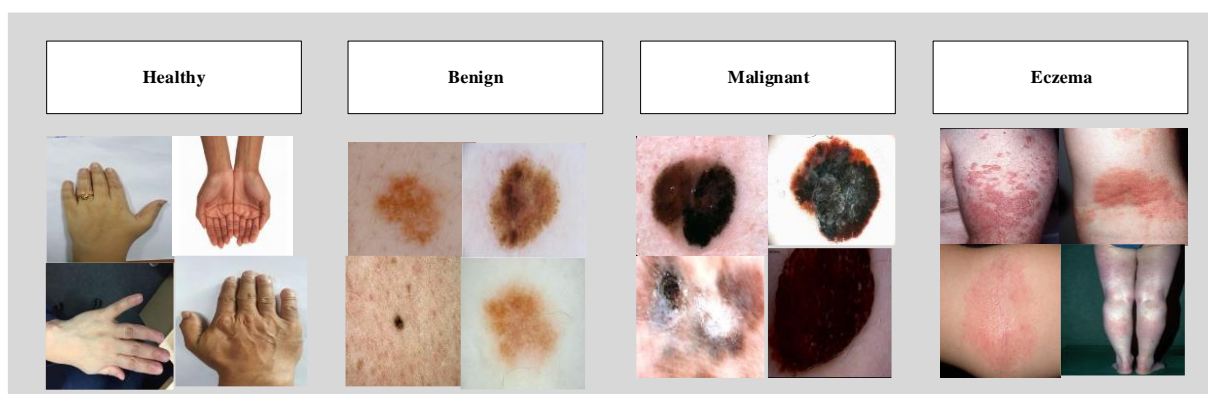
Figure 1. Example images from the dataset belonging to different categories

The number of images in each class are not consistent, resulting in a data imbalance issue. This issue is critical, as it may affect the classification results (Japkowicz & Stephen, 2002; Oliveira, Pereira, et al., 2018). In the imbalanced dataset, the accuracy of the evaluation result can be decreased as the classifier is biased and may give priority to a class having more image samples (Oliveira, Pereira, et al., 2018). Experiments are performed using the imbalance dataset, and the results are presented in the supplementary material. In the imbalanced dataset, the number of images in the healthy, benign, malignant and eczema classes are 3014, 3014, 918 and, 1235 respectively. Different approaches are proposed in the literature (Burdick, Marques, Weinthal, & Furht, 2018; Japkowicz & Stephen, 2002) to address this issue. In this research work, a random down-sampling approach is used. The random down-sampling approach removes images randomly from the majority class, i.e. from the healthy, benign and malignant class. Eczema has the minimum number of images, so other categories are downsized randomly to the number of images in the eczema category. The total number of images in each category are collected randomly and downsized to 918 to make them equal.. One is for training and testing, and another set is for comparing the performance of MCSL and MCML classification algorithms.For training and testing of the classification scheme, 860 images from each class are used for training and testing. The training and testing data was divided in the 70:30 ratio, i.e. 70% of the data is used for training the model, and 30% data is used for testing the models of MCSL and MCML classification algorithm. For comparing MCSL and MCML classification algorithm performance, 58 images from each class are used. In this research study, this comparison dataset is known as "Cross-Validation Data". We need the cross-validation data set because the performance of MCSL and MCML cannot be compared directly. The cross-validation data is not used in any training and testing phase. The total number of images in each class for the training, testing and cross-validation phase are presented in Table 3.

Table 3. Number of images in the training, testing and cross-validation dataset for the healthy, eczema, benign and malignant class

Image Class	Total	Training Images	Testing Images	Cross-Validation Images
Healthy	918	602	258	58
Eczema	918	602	258	58
Benign	918	602	258	58
Malignant	918	602	258	58
Total	**3672**	**2408**	**1032**	**232**

4. Methodology

Two approaches are used in this research work for classifying skin lesion images. The first approach makes use of traditional machine learning algorithms while deep learning is used in the second approach. Each approach is explained in the next subsections.

4.1. Traditional Machine Learning approach

In this paper, we define a traditional machine learning approach as a computational approach for the skin lesions classification that learns from a predefined bag of features and goes through a number of different steps. These steps are illustrated in Figure 2 which includes the following: (1) pre-processing, (2) segmentation, (3) feature extraction and (4) classification. Each step is detailed below.
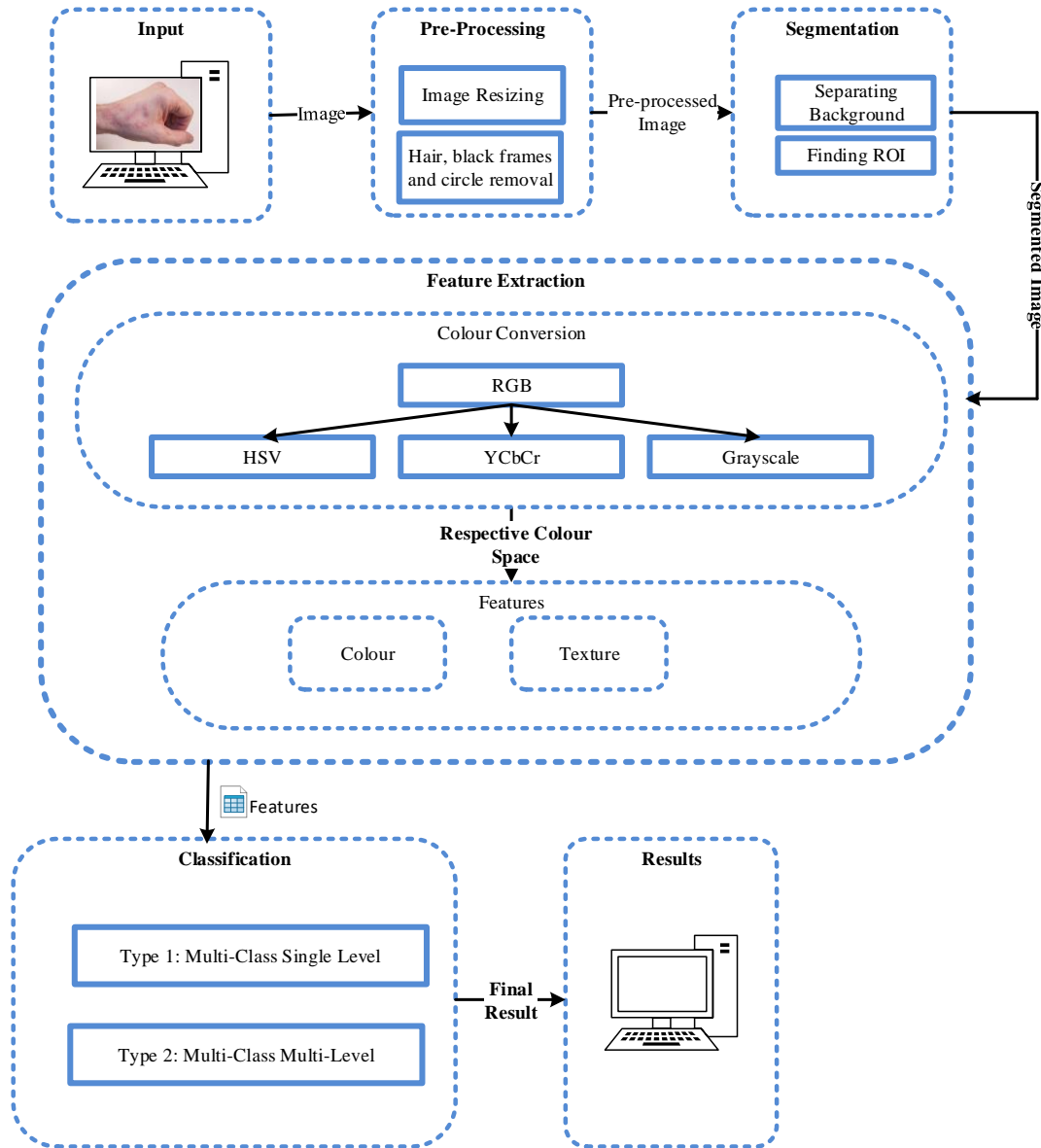
Figure 2. Overall methodology to classify skin lesions using the traditional machine learning approach

4.1.1. Pre-processing

The main purpose of the pre-processing step is to remove noise from the image. When capturing the image, many variables can affect the image such as skin nature, capturing environment, capturing device and lighting condition. Because of these variables; the skin images may contain some artefacts such as black frames, dermoscopic gel, air bubbles, skin lines, hairs, and blood vessels. These artefacts create a barrier to the segmentation process and result in accuracy loss and increased computational cost (Hameed et al., 2016; Maglogiannis & Doukas, 2009). Therefore, some pre-processing steps are required to remove these artefacts and in turn facilitate the segmentation and classification process. The tasks performed in the pre-processing step are shown in Figure 3.
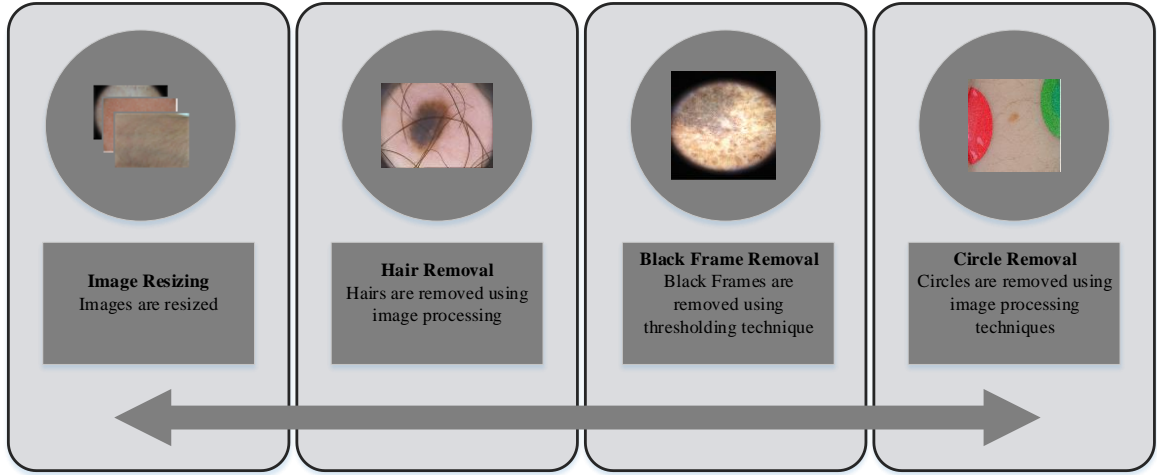
Figure 3. Pre-processing Step

4.1.1.1. Image Resizing

The data used in this research work is gathered from different sources hence the images are of different dimensions. For consistency, all the images are resized to 227 x 227 x 3.

4.1.1.2. Hair removal

In the hair removal algorithm, the RGB image is converted to the greyscale image using equation (1).

$$grey(x,y) = 0.3 * R(x,y) + 0.59 * G(x,y) + 0.11 * B(x,y) \qquad (1)$$

Where $grey(x,y)$ is the greyscale image, $R(x,y), G(x,y)$ and $B(x,y)$ are the red, green and blue colour channels of the RGB image of size M x N where $= 1, \dots, M$, $y = 1, \dots, N$ and pixel values of each channel are in [0,255] range. After the greyscale image conversion, the contrast enhanced image $(contrast(x,y))$ is constructed using the contrast-limited adaptive histogram equalisation method (Zuiderveld, 1994). After computing the contrast-enhanced image, the averaged image $(average(x,y))$ is constructed by applying the average filter. The $average(x,y)$ is computed using four filters of different sizes i.e. 3 x 3, 5 x 5, 7 x 7, and 9 x 9. To create the initial hair mask($hairmask(x,y)$); $contrast(x,y)$ and $average(x,y)$ are subtracted. The hair mask is converted to a binary image by applying the thresholding method. In this method, pixels within a defined range are selected as foreground, and pixels outside the range are selected as background. Let $Thresh(x,y)$ is a thresholded version of $hairmask(x,y)$ at some global threshold T and is calculated using equation (2).

$$Thresh(x,y) = \begin{cases} 1 & hairmask(x,y) \geq T \\ 0 & hairmask(x,y) < T \end{cases} \quad (2)$$

For thresholding, the experiments are performed using the ISODATA (Ridler, T.W. Calvard, 1978), and Otsu's (Otsu, 1979) thresholding methods. The resulting thresholded image has small objects which are removed by applying the morphological opening operation (Gonzalez & Woods, 2008b). Using morphological opening

operation, all the connected components that have fewer than "p" pixels are removed from the binary image, i.e. from $Thresh(x, y)$. In this research work, experiments are performed using different values of p i.e. using p=10, p=30 and p=50. For each combination i.e. combination of filter size, thresholding method and p value, the Jaccard index (JI) (Jaccard, 1901) is calculated. For calculating the JI, the ground truth hair mask is computed manually and then the JI is calculated using the ground truth image and the resulting image from each combination. The highest average JI is obtained using the combination of 5 x 5 filter size, Otsu's thresholding and p = 50. The effect of different filter size, thresholding method and different values of p along with the averaged JI for each combination is presented in the supplementary material[4].

To achieve a hair free image, the inpainting technique is used as it is the most used technique for hair removal in the literature (Abbas et al., 2011; Abbas, Garcia, Emre Celebi, & Ahmad, 2013; Ganster et al., 2001; Toossi et al., 2013). Image inpainting is a technique of modifying an image in an undetectable form (Bertalmio, Sapiro, Caselles, & Ballester, 2000). Different inpainting methods, i.e. harmonic (Shen & Chan, 2002), mum-ford shah (Esedoglu & Shen, 2002), Cahn-Hilliard (Bertozzi, Esedoglu, & Gillette, 2007) and transport (Bertalmio, 2001) are applied. The harmonic inpainting method is selected as it performs well and requires less computational time comparing with other methods. By the computational time, we mean that the time required to perform the inpainting method. The results achieved by the different inpainting methods are visually presented in Figure 4. The time required by each inpainting technique is given in Table 4.
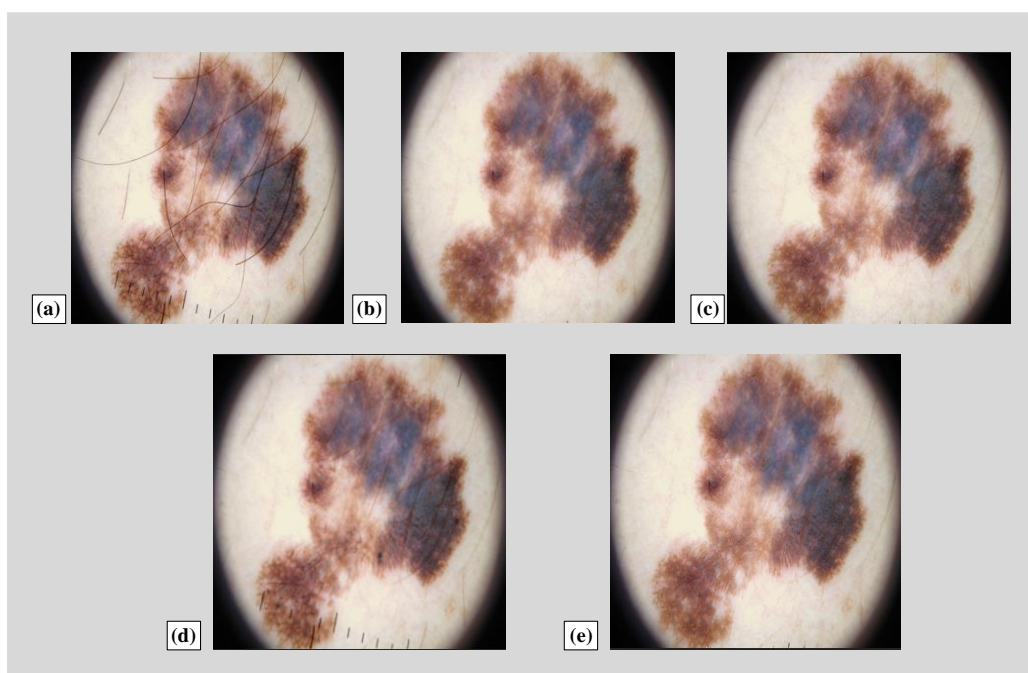

Figure 4. Effect of different inpainting techniques (a) Original image [5] (b) resulting image after applying Harmonic inpainting, (c) resulting image after applying Mumford-Shah inpainting, (d) resulting image after applying Cahn-Hilliard inpainting and (e) resulting image after applying Transport inpainting

---

[4] Effect of filters, thresholding method and different value of p is presented in Supplementary-Figure (6-8) and the Jaccard index of each combination is presented in Supplementary-Table 5.
[5] Image Courtesy: ISIC 2017 Skin Lesion Analysis Towards Melanoma Detection

Table 4: Time required by different inpainting techniques

Inpainting Technique Name	Computational Time (in sec)
Harmonic (Shen & Chan, 2002)	0.54
Mumford-Shah Inpainting (Esedoglu & Shen, 2002)	22.55
Cahn-Hilliard Inpainting (Bertozzi et al., 2007)	93.90
Transport Inpainting (Bertalmio, 2001)	39.71

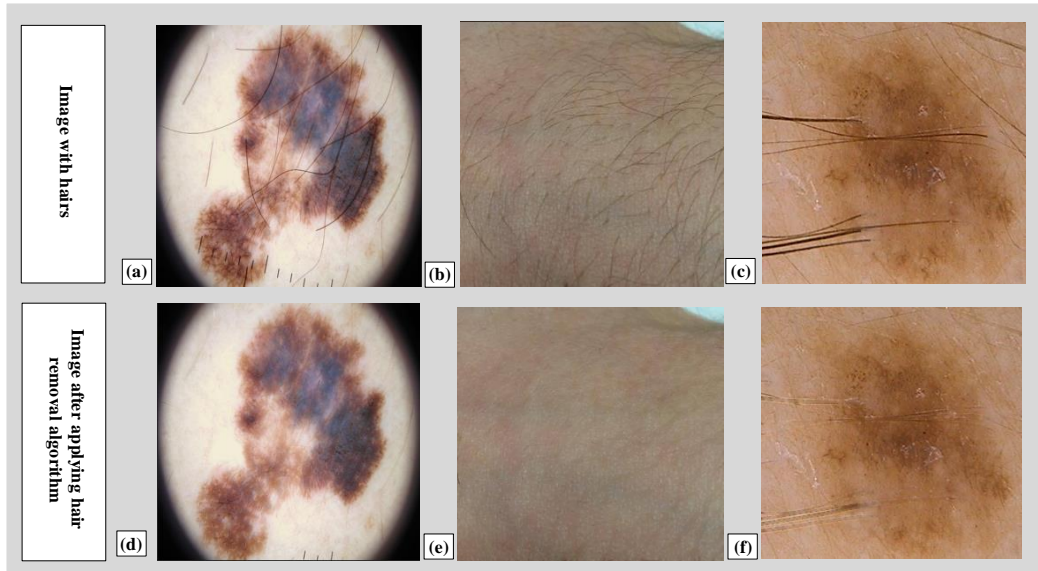The hair removal algorithm is applied to different skin lesion images, and the results are presented in Figure 5.

Figure 5. (a-c) Image before applying hair removal algorithm, (d-f) Image after applying hair removal algorithm

4.1.1.3. Black Frame Removal

Black frames are another noise that creates an obstacle in the segmentation phase. In the literature, there are some techniques proposed for removing black frames by creating a mask of different shapes (Abuzaghleh et al., 2014; Sultana et al., 2014). These techniques are suitable for a small number of images because one can create a mask for a few images, However, when you have thousands of images, this technique is not practical. In this research work, an enhanced technique is proposed to dynamically generate a black frame mask at runtime. The proposed technique is pixel-based and works on the RGB images. Let $R(x,y), G(x,y)$ and $B(x,y)$ be the red, green, and blue colour channels of the RGB image. The first step in the algorithm is to obtain the mean image ($mean(x,y)$), from the RGB image which is computed using equation (3).

$$mean(x,y) = \frac{1}{3} * \left(R(x,y) + G(x,y) + B(x,y) \right) \qquad (3)$$

The second step is to apply the thresholding method to calculate the black mask ($blackmask(x,y)$) using equation (4).

10

$$blackmask(x,y) = \begin{cases} 1 & mean(x,y) \geq T \\ 0 & mean(x,y) < T \end{cases} \qquad (4)$$

Experiments are performed with different values of T (10, 50 and 100) and the best result is achieved using $T = 50$. Results with different values of T are presented in Figure 6. The results are validated using JI. For computing the JI, the ground truth mask is computed manually and, then the JI is calculated using mask obtained by different values of T. The average JI obtained by different values of T are presented in Table 5.
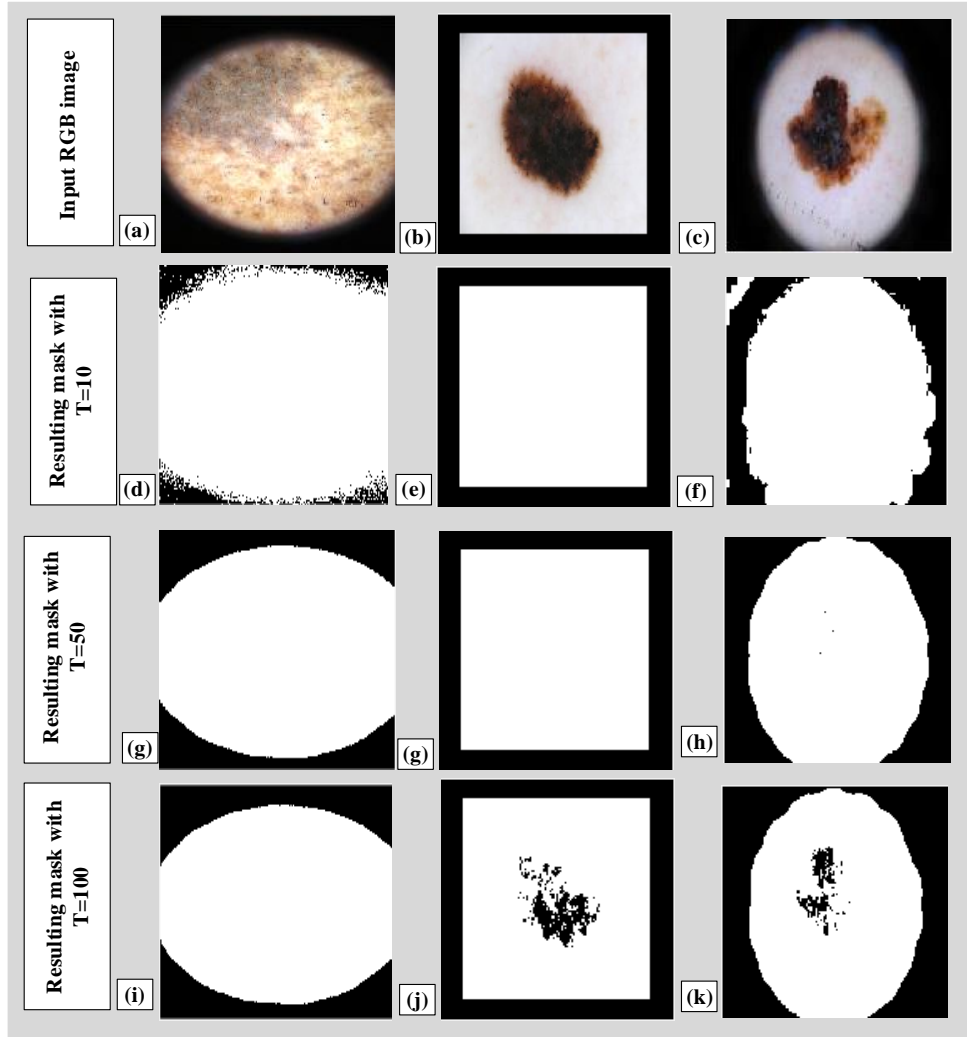
Figure 6. (a-c) RGB image, (d-f) resulting image using $T = 10$, (g-h) resulting image using $T = 50$,

(i-k) resulting image using $T = 100$

Table 5. Average JI for the mask with different threshold values

Average JI for $T = 10$	average JI for $T = 50$	Average JI for $T = 100$
0.80	0.98	0.96

Once the mask is calculated, it is used to obtain the image without a black frame using the morphological region filling (Gonzalez & Woods, 2008c). The results of the black frame removal algorithm are presented graphically in Figure 7.
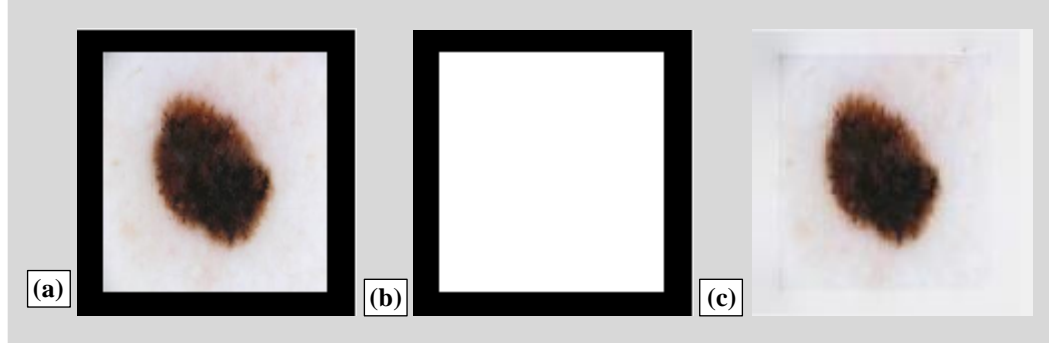

Figure 7. (a) RGB image, (b) black frame, (c) resulting image after applying region fill operation

The proposed technique has been applied to datasets collected from different sources, and the black frame is successfully generated for all the images. The proposed algorithm outperforms techniques proposed by Abuzagleh et al. and Sultana et al. (Sultana et al., 2014; Abuzaghleh, Barkana, & Faezipour, 2014) as the mask is dynamically calculated according to the input image at the runtime instead of a pre-defined mask.

4.1.1.4. Circle Removal

Some images in the ISIC challenge dataset contain circles of varied sizes as shown in Figure 8 and creates difficulty in the segmentation process.

Figure 8. Images with circle noise

To remove these circles, the input RGB image ($RGB(x, y)$) is converted into a greyscale image using equation (1). A binary image is then calculated by applying the Otsu's thresholding method which is denoted as $otsu(x, y)$. The obtained binary image after applying the Otsu's thresholding method contains unnecessary details which are removed by applying the morphological erosion operation. For the morphological erosion operation, the "disk" structuring element of size 1 is used in this algorithm. The resulting image contains the skin lesion which is not

12

part of the circles as shown in Figure 9 (k) and (l). These regions are removed and final mask denoted by mask(x, y) is computed using the hole filling method (Gonzalez and Woods, 2008). This method is used to fill the region in the thresholded image which is not part of the circles. In the morphological operations, a hole is defined as the set of background pixels that cannot be reached by filling in the background from the edge of the image (Wang, 2014). To obtain a circle free image, morphological region fill method is used to replace the pixels values using their neighbours. The results of the circle removal algorithm are presented in Figure 9.
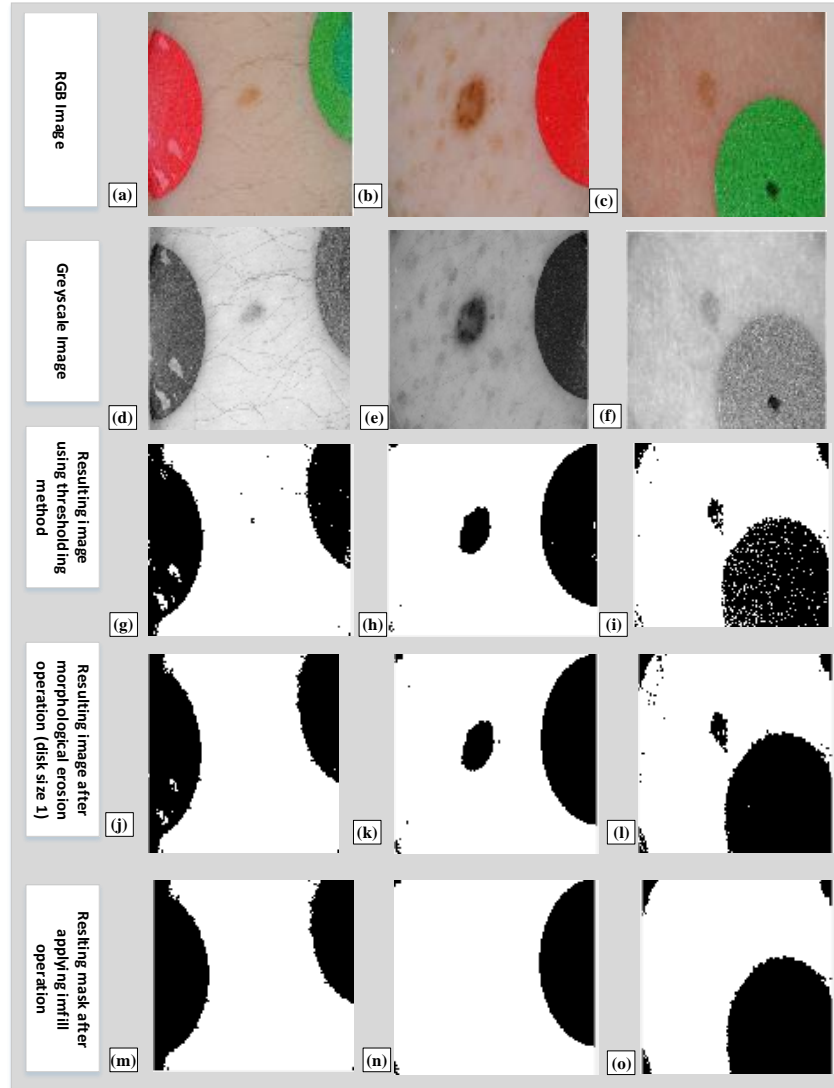

Figure 9. Steps involved in removing circles from skin images (a-c) RGB image with circles, (d-f) greyscale image, (g-i) resulting image after thresholding method, (j-l) resulting image after morphological erosion operation using disk structuring element of size 1, (m-o) resulting image after applying morphological hole operation.
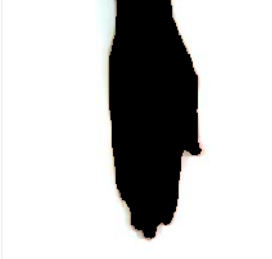
4.1.2. Segmentation

The main purpose of a segmentation step is to obtain the region of interest (ROI). The ROI area is expected to have more relevant information in the form of different features that can be used for lesion classification and diagnosis. For extracting the ROI, a hybrid technique is used. This hybrid technique utilises the capability of the K-means clustering (Hartigan & Wong, 1979), Otsu's thresholding (Otsu, 1979) and morphological erosion operation. In thresholding, the Otsu's method is applied. For this, the RGB image is first converted into a greyscale

image, and a global threshold level T is computed. This grey thresh is applied to the greyscale image to obtain the ROI. The resulting image after Otsu's thresholding is denoted by $otsu(x, y)$. For k-means clustering, the RGB image is first converted into L*a*b, and ab channel is used. The K-means clustering algorithm with k=2 is applied to the image, and the image is divided into two clusters, and the cluster with ROI is selected. As we want to divide our image into two clusters i.e. one for foreground and one for background, that is why k=2 is used in our experiments. For each cluster, number of black pixels are summed up and stored in N_b. The cluster having maximum numbers of N_b is selected as an optimal cluster. The resulting image obtained after k-means clustering is denoted as $cluster(x, y)$. The results for k-means clustering are presented in Table 6.

Table 6. Results obtained by applying the k-means clustering algorithm

Input Image	Resulting cluster 1 image after applying k-means	Resulting cluster 2 image after applying k-means	Resulting image after k-means clustering with a maximum value of N_b.
	N_b= 30635	N_b= 20900	
	N_b= 12079	N_b= 339450	
	N_b= 17662	N_b= 33867	

The segmented image ($segment(x, y)$), is obtained by combining $otsu(x, y)$ and $cluster(x, y)$. The $segment(x, y)$ is calculated using equation (5). This means that if a considered pixel value is non-zero in either of $otsu(x, y)$ or $cluster(x, y)$ it will be non-zero in $segment(x, y)$. The segmentation results are graphically presented in Figure 10.

$$segment(x, y) = \begin{cases} otsu(x, y), & otsu(x, y) \neq 0 \\ cluster(x, y), & cluster(x, y) \neq 0 \\ 0, & otsu(x, y) \text{ and } cluster(x, y) = 0 \end{cases} \quad (5)$$
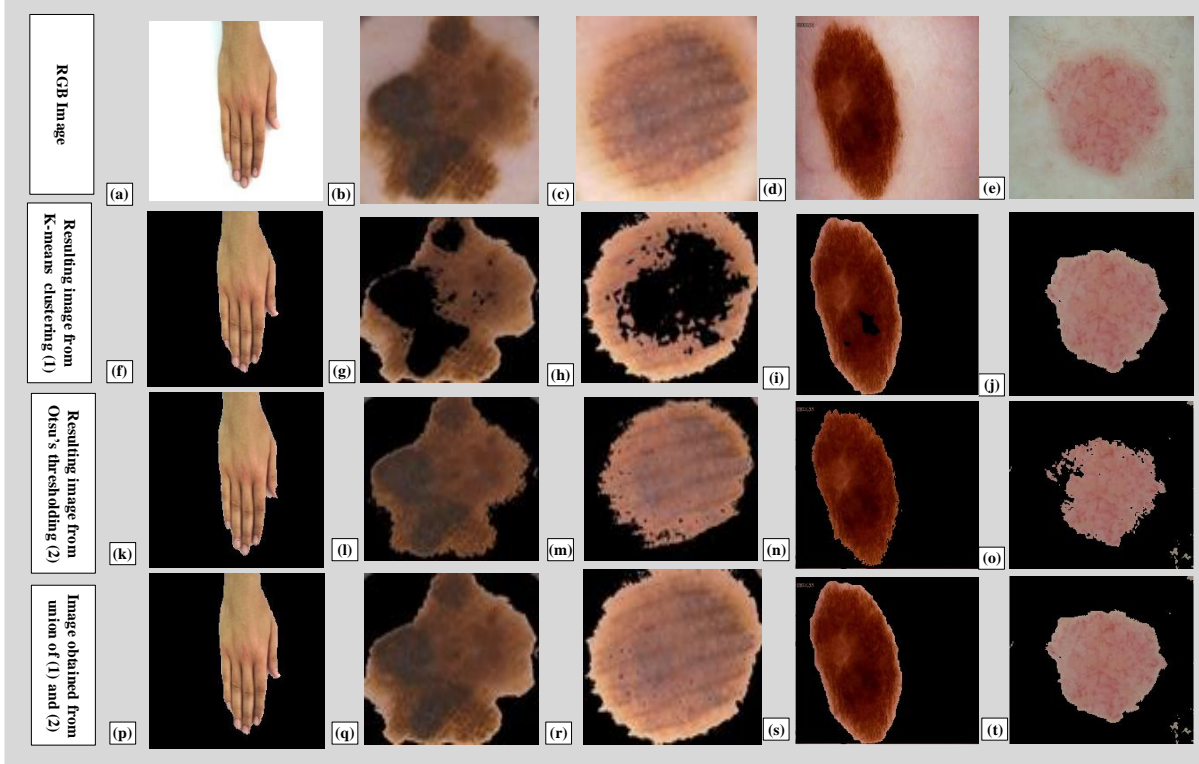

Figure 10. (a-e) Input RGB image, (f-j) Resulting image using k-means clustering technique, (k-o) Resulting image using Otsu's thresholding, (p-t) Resulting image from the union of thresholded and clustered image

There are some unnecessary objects in the final image which are removed by applying the morphological erosion operation with structuring element of disk size 1. The performance of the segmentation step is calculated using the JI. The average JI of Otsu's thresholding, k-means clustering, and hybrid technique are presented in Table 7.

Table 7. Average Jaccard score for different segmentation techniques

Average JI using Otsu's thresholding	Average JI using k-means Clustering	Average JI using hybrid technique
0.68	0.81	0.82

From Table 7, we can see that there is no much difference between the average JI of k-means clustering and hybrid technique. However, in most of the cases, the hybrid technique outperforms other two techniques. In 59% of the cases, the hybrid segmentation technique performs better than the Otsu's thresholding and k-means. However, very rarely (only 2%), the JI of hybrid and k-means is same. The percentage of cases where the hybrid technique outperforms others is presented graphically in Figure 11.
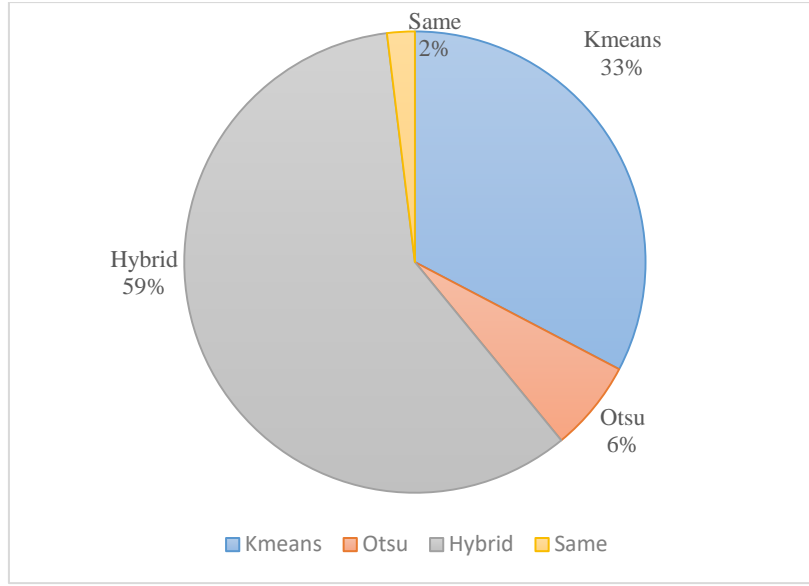
15

Figure 11. Pie-graph presenting the performance of Otsu thresholding, k-means clustering, and hybrid technique

4.1.3. Feature Extraction

One of the main challenges in this research work is to find the set of features that can be applied to all images, i.e. healthy, benign, malignant, and eczema, as the images of each class differ from others and they have their characteristics. For example, mostly the benign and malignant images have a clear boundary, and one can extract border and shape features while the healthy and eczema images may cover the full body part and do not have a clear boundary and it is difficult to extract the border features. Thus, a consistent set of features are required that can be extracted from images of all classes. In feature extraction, 36 features belonging to colour and texture categories are extracted and stored in the feature vector for the classification step.

From the segmented image, the colour histogram of R, G and B channel is calculated, and features are extracted from it. Colour histogram is a representation of the number of pixels at each intensity level of a colour channel. Colour histogram features state the global properties of the intensity level distribution for each colour channel. The histogram of an image can be represented as a probability distribution, $P(g)$, of the intensity levels as given in equation (6).

$$P(g) = \frac{N(g)}{M} \qquad (6)$$

Where g is the intensity level, $N(g)$ is the number of pixels at intensity level g, and M is the total number of pixels. The histogram features extracted from each colour channel are mean, mode, standard deviation, skewness, energy, entropy and kurtosis. The mean ($\bar{g}$) reflects the overall intensity level in the image. Standard deviation; also known as the variance and gives the spread of the data. A high contrast image has a high variance and a low contrast image has a low variance. The mathematical representation of the mean, and the standard deviation is given in equation (7) and (8).

$$M(\bar{g}) = \sum_{g=0}^{W-1} gP(g) = \sum_r \sum_c \frac{I(r,c)}{M} \qquad (7)$$

$$SD(\sigma_g) = \sqrt{\sum_{g=0}^{W-1} (g - \bar{g})^2 P(g)} \qquad (8)$$

Where W is the number of intensity levels, r is the number of rows and c is the number of columns in the image. The skewness computes the asymmetry of the probability distribution of the histogram. Therefore, it reveals information about the shape of the distribution. Skewness of the image is calculated using equation (9).

$$Skewness = \frac{1}{\sigma^3} \sum_{g=0}^{W-1} (g - \bar{g})^3 P(g) \qquad (9)$$

The energy measure is related to the colour span (i.e. the spread of the pixel values). The pixel colour energy decreases as the pixel values span a wider intensity range. The entropy measure describes the required amount of information to code the image data. In contrast to the energy measure, the entropy increases as the pixel values span a wider intensity range. The mathematical representation of energy and entropy is given in equation (10) and (11).

$$Energy = \sum_{g=0}^{W-1} [P(g)]^2 \qquad (10)$$

$$Entropy = -\sum_{g=0}^{w-1} [P(g) \, log_2 P(g)] \qquad (11)$$

Minimum and maximum values are extracted from each channel in the RGB colour space which is denoted as R_{min}, R_{max}, G_{min}, G_{max}, B_{min}, B_{max} for red, green, and blue channels respectively. For extracting other colour features the RGB image is divided into HSV, YC_bC_r, and grey scale colour spaces and consequently different features are extracted. These features include the mean value of H channel, V channel, Cb channel, Cr channel and greyscale image which are denoted as H_{mean}, V_{mean}, Cb_{mean}, Cr_{mean} and $Gray_{mean}$ respectively thus making 32 colour features in total.

For extracting texture features, the Gray Level Co-occurrence Matrix (GLCM) is computed first, and then 4 texture features are computed from it. GLCM is a matrix which shows the distribution of co-occurring pixels values at a given offset. The GLCM functions are used to characterise the texture of an image by calculating the occurrence of pair pixels, and their spatial relationship occurs in the images. The statistical measurements are then calculated (Alam et al., 2016; Oliveira, Papa, et al., 2018). The features extracted from this category are contrast, correlation, energy, and homogeneity. The contrast measures the local variations in the GLCM. To measure the joint probability occurrence of specified pair pixels, correlation is used. Contrast and correlation of GLCM are computed using a mathematical equation given in equation (12) and (13).

$$Contrast_{GLCM} = \sum_{i,j=0}^{W-1} P_{ij}(i-j)^2 \qquad (12)$$

$$Correlation_{GLCM} = \sum_{i,j=0}^{W-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2} \qquad (13)$$

where P_{ij} is the element of GLCM, μ is the GLCM mean and σ is the variance of the intensities of all reference pixels in the relationships that contributed to the GLCM. Energy is also known as the uniformity or the angular second moment, and it provides the sum of squared elements in the GLCM. The energy is calculated using the same formula given in (18), but here the input is the GLCM matrix. Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. The mathematical formula to calculate the homogeneity is given in equation (14). A total of 36 (32 colour and 4 texture) features are extracted and stored in a vector to be used in the classification step.

$$Homogeneity_{GLCM} = \sum_{i,j=0}^{W-1} \frac{P_{ij}}{1+(i-j)^2} \qquad (14)$$

4.1.4. Classification

To classify an image using extracted features, an appropriate learning algorithm should be selected. There are two types of algorithms available in machine learning, supervised learning algorithms and unsupervised learning algorithms. Supervised learning algorithms take the classified input data and train a model to generate the predictions for the response to the new unseen data. Unsupervised algorithms are used when classified data is not available. For this research, supervised learning algorithms are suitable because of the availability of the classified data. In this research work, for classification, we have used ANN. ANN is a machine learning algorithm inspired by the behaviour of neurons in the human brain and has been used successfully in skin lesions classification(Rubegni et al., 2002; Yasir, Rahman, & Ahmed, 2014; Alfed, Khelifi, Bouridane, & Huseyin, 2015;). In this research work, a multi-layered feed-forward neural network with a backpropagation algorithm is used. Backpropagation works by approximating the non-linear relationship between the input and the output by adjusting the weight values. ANN architecture consists of three layers, the input layer, hidden layer and output layer. ANN make use of a learning rule that allows the set of weights to be adaptively tuned during the training phase (Shenfield, Day, & Ayesh, 2018). In the classification phase, ANN receives the feature vector extracted from the feature extraction phase and starts the training process to adjust the weights values.

In training, three hidden layers are used. Experiments are also performed with one and two hidden layers, but the best results are obtained by using three hidden layers[6]. Sigmoid is used as an activation function as it is mostly used activation function for feedforward and backpropagation neural network (Khalid, Latif, & Adnan, 2017; Nielsen, 1992; Zhang, 2017). For evaluating the performance of the learned model (i.e. classifier), the holdout technique is used (Zheng, 2015). The holdout technique is the simplest validation technique in which the dataset is divided into two sets: the training set and testing set. The classifier learns the weights using the training set. These learned weights are then applied to the unseen set: test data, and in turn, the performance measures are calculated. The results depend on the data in the training and testing dataset, and the final evaluation may be significantly different depending on the data division; therefore, experiments are performed five times, and their average is computed. The main advantage of this method is that it takes less time for computation. Another reason for using holdout validation is that in this research work the results of the traditional machine learning approach are compared with the deep learning approach and using k-fold validation for deep learning is computationally expensive, and it will be not a fair comparison if different validation techniques are used for the two approaches while comparing them. In the classification step, MCML and MCSL algorithms are applied. In the MCSL classification, all the classes are categorized at the same level, i.e. healthy, benign, malignant, and eczema. The feature vector extracted in the feature extraction phase are given as input to the machine learning classifier. MCML works on the divide and conquers rule. In the MCML classification algorithm, the classification problem is divided into sub-classification problems, and at the end, the learned classifiers combine to get the final result. At level 1, binary classification is performed, and the images are classified into two categories, i.e. healthy and unhealthy. At level 2, unhealthy images are further classified into melanoma and eczema. Level 3 is the last level of MCML classifier, and at this level, the melanoma images are classified as either malignant or benign. The MCSL and MCML classification algorithms are graphically shown in Figure 12 and Figure 13.

---

[6] Training and testing confusion matrix obtained by using 1,2 and 3 hidden layers are given in supplementary material (Supplementary-Figure 9-10)
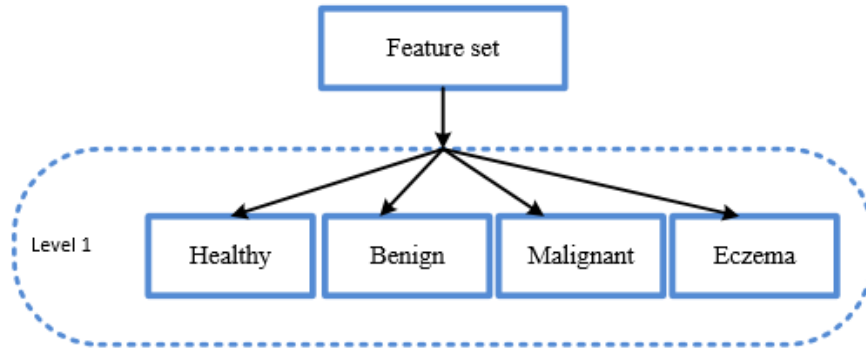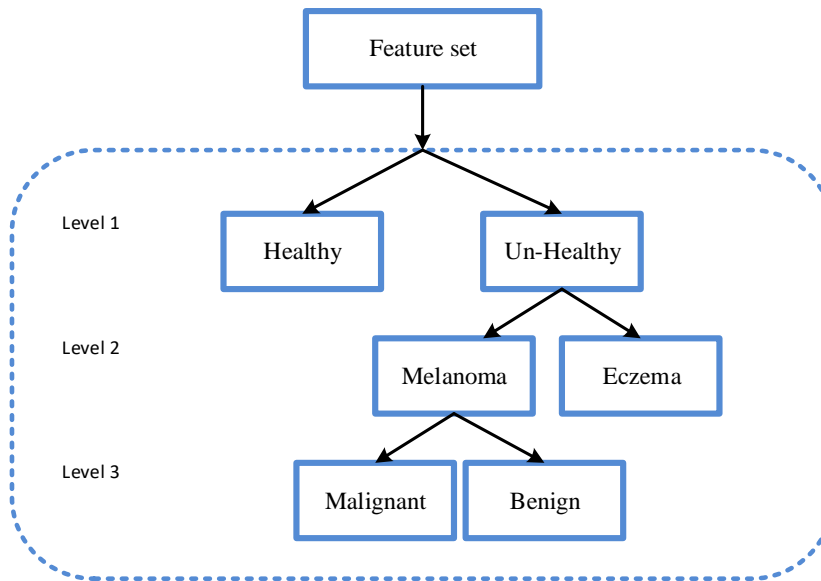
Figure 12.Multi-Class Single Level Classification

Figure 13. Multi-Class Multi-Level Classification

4.2. Deep Learning Approach

The deep learning approach is powered by the advances in computation and has been shown exceptional performance in object recognition and classification (Burdick et al., 2018; Esteva et al., 2017). Deep learning has produced results comparable to and in some cases superior to human experts. The deep learning algorithm is trained end-to-end directly from raw image pixels and the image label. For deep learning convolutional neural network (CNN) is used for image classification (Krizhevsky, Sutskever, & Hinton, 2012). For performing MCSL and MCML classification using the deep learning approach, the transfer learning approach is used, and a pre-trained model AlexNet (Krizhevsky et al., 2012) is modified, fine-tuned and re-trained on our dataset. AlexNET is a CNN that is trained on a million of images from the ImageNet database (Stanford Vision Lab, 2016). The AlexNET contains five convolutional layers (CL) and three fully-connected layers (FCL) (Krizhevsky et al., 2012). The architecture of AlexNET is presented graphically in Figure 14. Detail of each layer can be found in the supplementary material. The AlexNet model offers several benefits like well-known implementation, few training parameters and extreme validity (Dorj et al., 2018). AlexNet is trained on 1,000 object classes with a top-five error rate of 15.3%. For retraining AlexNet on skin image dataset, the images are resized to 277 x 277 x 3 because of model constraint. The parameters along with their description and values used in training our model using the deep learning approach are given in Table 8.

Table 8. Parameters used in training the deep learning algorithm

Parameter Name	Description	Value
Learning Rate	Step size to control the weights adjustment with respect to the loss gradient	0.001
Max Epoch	Full pass of the data	20
Mini Batch Size	Number of training examples utilised in one iteration	64
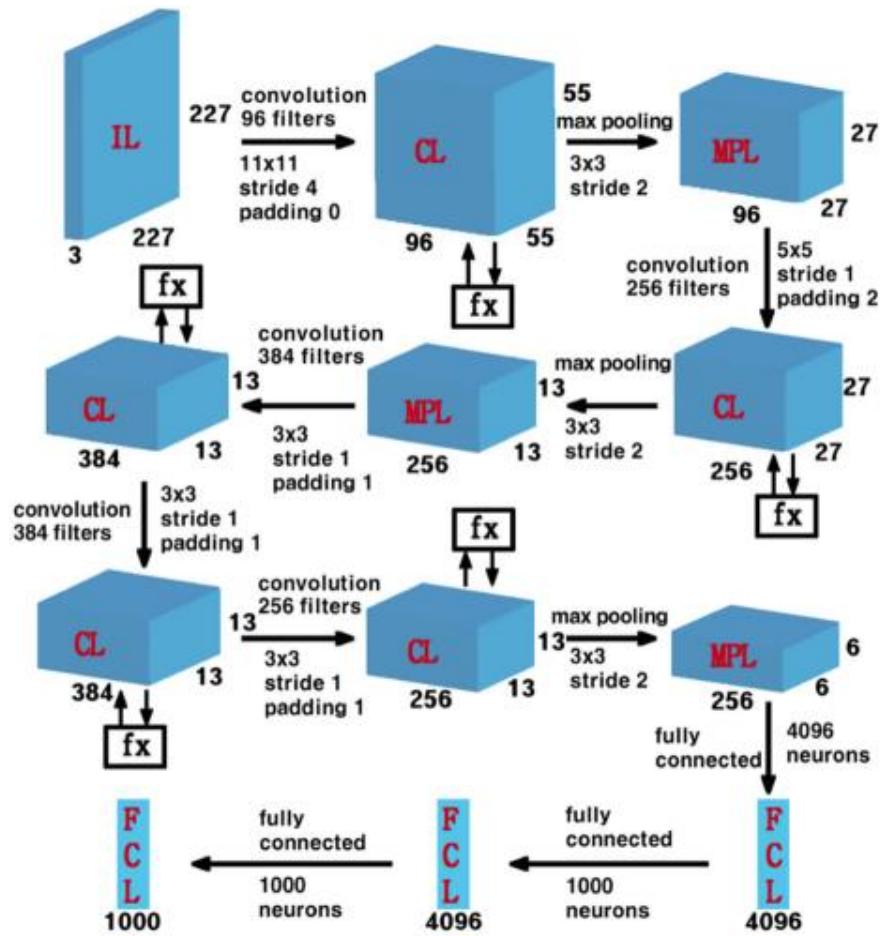
Figure 14. AlexNET Architecture. IL is the input layer; CL is the convolutional layers; MPL is the max-pooling layer, and FCL is the fully-connected layer (Dorj et al., 2018)

5. Results and Discussion

The proposed algorithms were developed using MATLAB 2018a. Algorithms are performed on an Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz with 16 GB of RAM, running Microsoft Windows 10 Enterprise 64 bits. Results of MCML and MCSL algorithms are calculated and compared using accuracy, sensitivity, specificity, and precision metrics. The accuracy, precision, sensitivity, and specificity are calculated using equations (15-18).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \qquad (15)$$

$$Precion = \frac{TP}{TP + FP} \times 100 \qquad (16)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \qquad (17)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \qquad (18)$$

The terms true positive, true negative, false positive and false negative are explained in Table 9.

Table 9.Terms used for measuring the performance metrics

Term	Meaning
True Positive (TP)	Healthy image classified as healthy
True Negative (TN)	Unhealthy image classified as unhealthy
False Positive (FP)	Unhealthy image classified as healthy
False Negative (FN)	Healthy image classified as unhealthy

While comparing the performance of a classification model, using the only accuracy can be misleading especially when there is huge class imbalance. In the imbalanced dataset, the classifiers tend to be biased towards the majority class and struggle to perform well on the minority class. Therefore, other performance measures such as sensitivity, specificity and precision are used along with the accuracy to evaluate the learned classifier. In MCSL classification, multiple skin lesions are classified at a single level. MCSL classification is performed using both traditional machine learning and deep learning. The dataset is divided into training and testing sets with a ratio of 70:30. When classification is done, the performance metrics are calculated for the training and testing phase separately for each class, i.e. healthy, benign, malignant, and eczema. As in MCSL algorithm, all classifications are performed at the same level, so multi-class confusion matrix is obtained. The results obtained for MCSL in the training and testing phase using the traditional machine learning approach are shown in Table 10 while the training and testing results using deep learning are given in Table 11.

Table 10. Sensitivity, specificity, precision and accuracy (in %) for healthy, benign, malignant and eczema class achieved in training and testing phase using traditional machine learning approach

Training Phase				
	Healthy	Benign	Malignant	Eczema
Sensitivity	97.17	69.20	64.74	77.40
Specificity	99.05	88.11	88.13	94.81
Precision	97.17	64.31	63.08	84.92
Accuracy	98.58	83.64	82.55	90.04

	Testing Phase			
Sensitivity	96.90	65.95	61.20	75.18
Specificity	98.69	87.22	87.44	93.78
Precision	96.15	60.24	61.20	82.17
Accuracy	98.24	82.39	81.02	88.65

Table 11. Sensitivity, specificity, precision and accuracy (in %) for healthy, benign, malignant and eczema class achieved in training and testing phase using deep learning approach

Training Phase				
	Healthy	**Benign**	**Malignant**	**Eczema**
Sensitivity	100	97.34	96.35	100
Specificity	100	99.11	99.22	99.56
Precision	100	97.34	97.64	98.69
Accuracy	100	98.67	98.50	99.67
Testing Phase				
Sensitivity	100	97.30	96.12	99.61
Specificity	99.87	98.97	99.23	99.61
Precision	99.61	96.92	97.64	98.85
Accuracy	99.90	98.55	98.45	99.61

In the MCML classification, the classification problem is divided into different levels. The results achieved by traditional machine learning approach for MCML classification for each level are given in Table 12, Table 13, and Table 14. While, the results achieved by the deep learning approach for each level of the MCML classification is shown in Table 15, Table 16, and Table 17.

Table 12. Sensitivity, specificity, precision and accuracy (in %) of healthy vs unhealthy using the traditional machine learning approach

	Training	Testing
Sensitivity	98.99	96.89
Specificity	99.14	98.79
Precision	97.51	96.51

	Training	Testing
Accuracy	99.10	98.30

Table 13. Sensitivity, specificity, precision and accuracy (in %) of melanoma vs eczema using the traditional machine learning approach

	Training	Testing
Sensitivity	93.33	91.55
Specificity	89.08	84.55
Precision	95.29	92.83
Accuracy	92.07	89.35

Table 14. Sensitivity, specificity, precision and accuracy of benign vs malignant (in %) using the traditional machine learning approach

	Training	Testing
Sensitivity	84.83	78.66
Specificity	77.50	73.98
Precision	75.63	72.87
Accuracy	80.81	76.18

Table 15. Sensitivity, specificity, precision and accuracy for healthy vs unhealthy (in %) using the deep learning approach

	Training	Testing
Sensitivity	99.83	99.61
Specificity	100	100
Precision	100	100
Accuracy	99.92	99.81

Table 16. Sensitivity, specificity, precision and accuracy for melanoma vs eczema (in %) using the deep learning approach

	Training	Testing
Sensitivity	99.50	99.61
Specificity	99.83	100
Precision	99.83	100
Accuracy	99.67	99.81

Table 17. Sensitivity, specificity precision and accuracy for benign vs malignant (in %) using the deep learning approach

	Training	Testing
Sensitivity	87.87	87.21
Specificity	98.17	98.45
Precision	97.96	98.25
Accuracy	93.02	92.83

Using the traditional machine learning approach, the average time consumed in pre-processing step for each image is around 3s, segmentation step takes 0.95s, 0.27s are taken by the feature extraction step and classification is performed in only 6s. Total training and testing time required for the traditional machine learning and deep learning approach is given in Table 18.

Table 18. Comparison of computational time for the traditional machine learning approach and deep learning

Training Time		
	Traditional Machine Learning	**Deep Learning**
MCSL	181.13 min	35.76 min
MCML	181.25 min	53.22 min
Testing Time		
	Traditional Machine Learning	**Deep Learning**
MCSL	3.077 s	0.032 s

MCML	3.282 s	0.16 s

The training and testing time of the traditional machine learning approach is higher than the deep learning approach because of pre-processing, segmentation, feature extraction and classification steps, whereas in deep learning, these steps are not required. Deep learning training time is also high because we performed our experiments on the CPU. Computational time can be reduced if experiments are performed on GPU. Although, MCML classification take more time but yields high accuracy.

As mentioned earlier, the cross-validation dataset is kept aside to validate the MCSL and MCML classification algorithm, and the cross-validation dataset is not used previously in the training and testing phase of the models. In the MCSL, only one model is trained whereas in MCML algorithm, different models are trained at different levels, and then all the models are combined to make the final classification. Hence the MCSL and MCML algorithm are compared using the cross-validation dataset. The comparison of MCSL and MCML algorithm is graphically shown in Figure 15. From the figure, it is clear that the accuracy of the traditional machine learning approach, achieved by the MCSL algorithm is 61.64% which is less than the accuracy achieved by MCML algorithm that reaches 63.79%. Using the deep learning approach, the MCML algorithm again performs better than MCSL. The classification accuracy achieved by MCML algorithm is 96.47%, and the classification accuracy achieved by the MCSL algorithm is 96.03%.
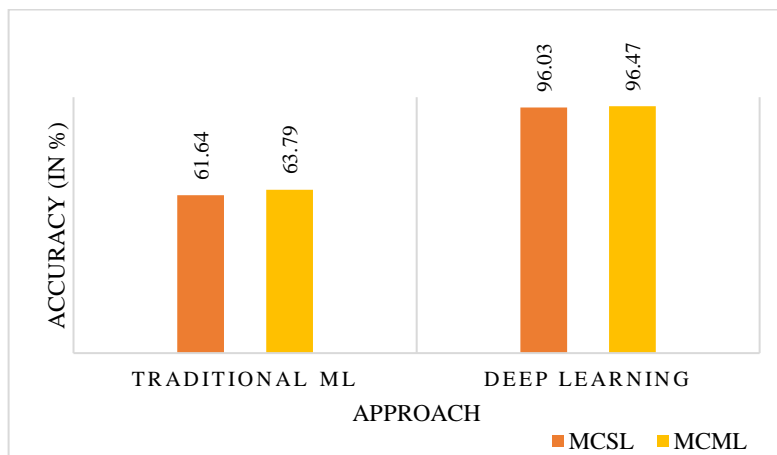

Figure 15. Comparison of MCSL and MCML using traditional machine learning and deep learning

In both approaches, the proposed MCML classification algorithm performs better. However, the training time of MCML classification is always more than the MCSL classification. There is a trade-off between the performance measures and the computational time. Given the nature of the field, diagnostic accuracy is more important (Salem et al., 2018). Moreover, it should be noted that the training is just performed for once. Once the model is trained, the classification of the new image is performed in a fraction of seconds. However, The adoption of the technique depends on the requirement of the deploying platform and the end user.

In MCML, the problem is divided into subproblems and then combined together; therefore it is more robust at classifying the skin diseases because the diseased images are scrutinised more than once in MCML, which ultimately results in generating less FN's and FP's. The MCML classification algorithm trained using deep learning approach is capable of generating less FN's for malignant melanoma class than the FN's produced by MCSL classification. The confusion matrix for MCSL and MCML using deep learning approach are presented in the supplementary material (Supplementary-Table (7-8)). Generating less FN's for malignant melanoma will

result in timely treatment of the fatal disease, and timely treatment of the fatal disease may save lives. Fewer numbers of FN's results in more sensitivity, which is the ultimate goal of the MCML classification algorithm.

Another advantage of MCML classification algorithm over the MCSL classification is reusability while retraining the model on the new data. If the model needs to be retrained on new data, the MCSL classification needs to be retrained on full dataset again, whereas in the MCML classification, only the specific model needs to be retrained and updated. For example, if we have more benign or malignant data, and we want our system to be re-trained, only benign vs malignant model has to be retrained in case of MCML algorithm, whereas, if we want to incorporate the new data in MCSL classification algorithm, retraining will be performed on the full dataset instead of a subset of the dataset.

The inconsistency of the dataset is one of the limitations of this research work. The dataset is gathered from different sources and hence has many inconsistencies. In online data repositories such as DermIS and DermQuest, different people can upload their skin lesion images. Therefore, the capturing device is different. Hence the format, quality, dimension, type of images differ for each image. In the dataset, some of the images cover the full body while other images contain the lesion area which limits the feature extraction step in traditional machine learning approach as we cannot extract shape and boundary features. The proposed classification model can only classify the skin image into one of the considered four non-overlapping classes. In the situation, where the rarer image comes for the classification, the proposed system will classify it to one of the provided class. Here, by rarer image, we mean any image other than these classes. In that case, the system will generate false positives, and the specificity and precision will be decreased, as the provided image is not from the considered classes, but it is classified as one of the class. Hence the system will generate FP's. In this research work, a small number of images are used for experiments. Better results can be achieved if a large amount of the dataset is used. Another limitation of this research work is the lack of using the feature selection algorithms. In our future work, we will extract more features and apply feature selection algorithms for finding the optimal set of features.

6. Conclusion

This paper presented an investigation into the development of an intelligent multi-class multi-level (MCML) classification algorithm to classify multiple skin diseases. The developed intelligent diagnosis scheme is expected to help the users and skin specialists in early skin lesion assessment. The proposed algorithm is implemented using two approaches, the traditional machine learning and deep learning. Pre-processing, segmentation, feature extraction and classification steps are involved in the traditional machine learning approach. Noise is removed in the pre-processing step, and a hybrid technique is implemented to get the region of interest in the segmentation step. Colour and texture features are extracted in the feature extraction step and the image is classified in the classification step. Transfer learning is used for the deep learning approach and learns directly from the images. The proposed algorithm is compared with the multi-class single-level classification algorithm, and high accuracy is achieved by MCML algorithm in both traditional machine learning and deep learning approach.

It is worth emphasising that previous studies have a limitation in the number of diseases they considered and the features they used for classification (Übeyli, 2008, 2009; Güvenir & Emeksiz, 2000; Abdi & Giveki, 2013; Kumar et al., 2016;). On the contrary, the present study focuses on the investigation into the development of an intelligent digital diagnosis system, where the limitation of the previous studies is overcome. The proposed algorithm is trained, tested and validated using 3672 images and 96.47% accuracy is achieved using the deep learning approach.

Future studies regarding multi-class skin lesion classification could be extended through incorporating more diseases with an objective to develop a mobile-enabled expert system for the remote areas where there is no, or very limited diagnosis facilities are available.

Conflict of interest.

The authors report no conflict of interest.

References

Abbas, Q., Celebi, M. E., & Fondón, I. (2011). Hair removal methods : A comparative study for dermoscopy images. *Biomedical Signal Processing and Control*, 6, 395–404. https://doi.org/10.1016/j.bspc.2011.01.003

Abbas, Q., Garcia, I. F., Emre Celebi, M., & Ahmad, W. (2013). A Feature-Preserving Hair Removal Algorithm for Dermoscopy Images. *Skin Research and Technology*, *19*(1), e27–e36. https://doi.org/10.1111/j.1600-0846.2011.00603.x

Abdi, M. J., & Giveki, D. (2013). Automatic detection of erythemato-squamous diseases using PSO-SVM based on association rules. *Engineering Applications of Artificial Intelligence*, *26*(1), 603–608. https://doi.org/10.1016/j.engappai.2012.01.017

Abuzaghleh, O., Barkana, B. D., & Faezipour, M. (2014). Automated skin lesion analysis based on color and shape geometry feature set for melanoma early detection and prevention. In *2014 IEEE Long Island Systems, Applications and Technology Conference, LISAT 2014*. Farmingdale, NY, USA: IEEE. https://doi.org/10.1109/LISAT.2014.6845199

Afifi, M. (2017). 11K Hands: Gender recognition and biometric identification using a large dataset of hand images. *arXiv Preprint arXiv:1711.04322*. Retrieved from http://arxiv.org/abs/1711.04322

Alam, N., Tabassum, T., Munia, K., Tavakolian, K., Mackinnon, N., & Fazel-rezai, R. (2016). Automatic Detection and Severity Measurement of Eczema Using Image Processing. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1365–1368). IEEE.

Alfed, N., Khelifi, F., Bouridane, A., & Huseyin, S. (2015). Pigment network - based skin cancer detection. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 7214–7217).

Bertalmio, M. (2001). *Processing of Flat and Non-Flat Image Information on Arbitrary Manifolds Using Partial Differential Equations*. UNIVERSITY OF MINNESOTA. Retrieved from nhttp://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.957

Bertalmio, M., Sapiro, G., Caselles, V., & Ballester, C. (2000). Image inpainting. *Proceedings of Conference on Computer Graphics and Interactive Techniques*, 417–424. https://doi.org/http://dx.doi.org/10.1145/344779.344972

Bertozzi, A. L., Esedoglu, S., & Gillette, A. (2007). Inpainting of binary images using the Cahn-Hilliard equation. *IEEE Transactions on Image Processing*, *16*(1), 285–291. https://doi.org/10.1109/TIP.2006.887728

British Skin Foundation. (2018). British Skin Foundation. Retrieved May 20, 2018, from http://www.britishskinfoundation.org.uk

Burdick, J., Marques, O., Weinthal, J., & Furht, B. (2018). Rethinking Skin Lesion Segmentation in a Convolutional Classifier. *Journal of Digital Imaging*, *31*(4), 435–440. https://doi.org/10.1007/s10278-017-0026-y

Çataloluk, H., & Kesler, M. (2012). A Diagnostic Software Tool for Skin Diseases with Basic and Weighted K-NN. In *2012 International Symposium on Innovations in Intelligent Systems and Applications*. Trabzon, Turkey: IEEE.

Celebi, M. E., Aslandogan, Y. a., & Bergstresser, P. R. (2005). Unsupervised border detection of skin lesion images. In *International Conference on Information Technology: Coding and Computing* (Vol. 2, pp. 1–6). https://doi.org/10.1109/ITCC.2005.283

Chang, C.-Y., & Liao, H.-Y. (2011). Automatic Facial Skin Defects Detection and Recognition System. In *Fifth International Conference on Genetic and Evolutionary Computing* (pp. 260–263). Xiamen, China: IEEE. https://doi.org/10.1109/ICGEC.2011.67

Chang, C. L., & Chen, C. H. (2009). Applying decision tree and neural network to increase quality of dermatologic diagnosis. *Expert Systems with Applications*, *36*(2), 4035–4041. https://doi.org/10.1016/j.eswa.2008.03.007

Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., … Halpern, A. (2016). Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). In *arXiv:1605.01397* (pp. 3–7).

De Guzman, L. C., Maglaque, R. P. C., Torres, V. M. B., Zapido, S. P. A., & Cordel, M. O. (2015). Design and Evaluation of a Multi-model, Multi-level Artificial Neural Network for Eczema Skin Lesion Detection. In *3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)* (pp. 42–47). IEEE. https://doi.org/10.1109/AIMS.2015.17

Derm101 Image Library. (2018). Derm101 Image Library. Retrieved January 12, 2018, from https://www.derm101.com/image-library/

DermIS. (2018). DermIS. Retrieved June 29, 2017, from http://www.dermis.net/dermisroot/en/home/index.htm

DermNZ. (2018). DermNZ-Image Library. Retrieved January 13, 2018, from https://www.dermnetnz.org/image-library/

Dorj, U., Lee, K., Choi, J., & Lee, M. (2018). The skin cancer classification using deep convolutional neural network. *Multimedia Tools and Applications*, 1–16. https://doi.org/https://doi.org/10.1007/s11042-018-5714-1

Esedoglu, S., & Shen, J. (2002). Digital inpainting based on the Mumford – Shah – Euler image model. *European Journal of Applied Mathematics*, *13*, 353–370. https://doi.org/10.1017/S0956792501004904

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Ferreira, A., Gentil, F., & Tavares, J. M. R. S. (2014). Segmentation algorithms for ear image data towards biomechanical studies. *Computer Methods in Biomechanics and Biomedical Engineering*, *17*(8), 888–904. https://doi.org/10.1080/10255842.2012.723700

Filho, M., Ma, Z., & Tavares, J. M. R. S. (2015). A Review of the Quantification and Classification of Pigmented Skin Lesions: From Dedicated to Hand-Held Devices. *Journal of Medical Systems*, *39*(177). https://doi.org/10.1007/s10916-015-0354-8

Fox News. (2015). 5 Most Common Skin Disorders. Retrieved June 1, 2018, from

http://www.foxnews.com/story/2009/12/15/5-most-common-skin-disorders.html

Freedberg, K. A., Geller, A. C., Miller, D. R., Lew, R. A., & Koh, H. K. (1999). Screening for malignant melanoma: A cost-effectiveness analysis. *Journal of the American Academy of Dermatology*, *41*(5 I), 738–745. https://doi.org/10.1016/S0190-9622(99)70010-1

Ganster, H., Pinz, A., Rohrer, R., Wildling, E., Binder, M., & Kittler, H. (2001). Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, *20*(3), 233–239. https://doi.org/10.1109/42.918473

Giotis, I., Molders, N., Land, S., Biehl, M., Jonkman, M. F., & Petkov, N. (2015). MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Systems with Applications*, *42*(19), 6578–6585. https://doi.org/10.1016/j.eswa.2015.04.034

Gonzalez, R. C., & Woods, R. E. (2008a). Hole Filling. In *Digital Image Processing* (Third, pp. 643–645).

Gonzalez, R. C., & Woods, R. E. (2008b). Opening and Closing. In *Digital Image Processing* (Third Edit, pp. 635–639).

Gonzalez, R. C., & Woods, R. E. (2008c). Region filling. In *Digital Image Processing*.

Gutman, D., Codella, N. C. F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., & Halpern, A. (2016). Skin Lesion Analysis toward Melanoma Detection : A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016 , hosted by the International Skin Imaging Collaboration (ISIC) (pp. 3–7).

Güvenir, H. A., & Emeksiz, N. (2000). Expert system for the differential diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *18*(1), 43–49. https://doi.org/10.1016/S0957-4174(99)00049-4

Hameed, N., Hassan, K. A., & Hossain, M. A. (2016). A comprehensive survey on image-based computer aided diagnosis systems for skin cancer. In *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)* (pp. 205–214). IEEE. https://doi.org/10.1109/SKIMA.2016.7916221

Hameed, N., Shabut, A., & Hossain, M. A. (2018a). A Computer-aided diagnosis system for classifying prominent skin lesions using machine learning. In *10th Computer Science and Electronic Engineering Conference*. Colchester: IEEE.

Hameed, N., Shabut, A., & Hossain, M. A. (2018b). Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine. In *12th International Conference on Software, Knowledge, Information Management and Applications*. Phnom Penh, Cambodia.

Hartigan, J. A., & Wong, M. A. (1979). A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108.

Hay RJ, Augustin M, Griffiths CE, S. W. (2015). The global challenge for skin health. *British Journal of Dermatology*, *172*(6), 1469–1472. https://doi.org/10.1111/bjd.13854

Ilter, N., & Guvenir, H. A. (1998). UCI Machine Learning Repository: Dermatology Data Set. Retrieved April 27, 2017, from https://archive.ics.uci.edu/ml/datasets/Dermatology

Jaccard, P. (1901). Distribution comparée de la flore alpine dans quelques régions des Alpes occidentales et orientales. *Bulletin de La Société Vaudoise Des Sciences Naturelles*, *37*, 241–272.

Jain, S., Jagtap, V., & Pise, N. (2015). Computer aided melanoma skin cancer detection using image processing. In *International Conference on Intelligent Computing, Communication & Convergence* (Vol. 48, pp. 735–740). India: Elsevier Masson SAS. https://doi.org/10.1016/j.procs.2015.04.209

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data*

Analysis, *6*(5), 429–449.

Jaworek-Korjakowska, J., & Kleczek, P. (2018). eSkin : Study on the Smartphone Application for Early Detection of Malignant Melanoma. *Wireless Communications and Mobile Computing*, *2018*, 1–11. https://doi.org/https://doi.org/10.1155/2018/5767360 Research

Jodas, D. S., Pereira, A. S., & Tavares, J. M. R. S. (2016). A review of computational methods applied for identification and quantification of atherosclerotic plaques in images. *Expert Systems with Applications*, *46*, 1–14. https://doi.org/10.1016/j.eswa.2015.10.016

Karimkhani, C., Dellavalle, R. P., & Coffeng, L. E. (2017). Global Skin Disease Morbidity and Mortality An Update From the Global Burden of Disease Study 2013. *Jama Dermatology*, *153*(5), 406–412. https://doi.org/10.1001/jamadermatol.2016.5538

Khalid, A., Latif, M. A., & Adnan, M. (2017). AN APPROACH TO ESTIMATE THE DURATION OF SOFTWARE PROJECT THROUGH MACHINE LEARNING TECHNIQUES. *Gomal University Journal of Research*, *33*(1).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *25th International Conference on Neural Information Processing Systems* (pp. 1097–1105). Nevada: Association for Computing Machinery. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Kumar, V. B., Kumar, S. S., & Saboo, V. (2016). Dermatological Disease Detection using Image Processing and Artificial Neural Network. In *International Conference on Artificial Intelligence and Pattern Recognition (AIPR)* (pp. 88–93). https://doi.org/10.1109/ICAIPR.2016.7585217

Lee, T., Ng, V., Gallagher, R., Coldman, A., & McLean, D. (1997). Dullrazor: A software approach to hair removal from images. *Computers in Biology and Medicine*, *27*(6), 533–543. https://doi.org/10.1016/S0010-4825(97)00020-6

Ma, Z., Tavares, J. D. S., & Jorge, R. (2009). A review on the current segmentation algorithms for medical images. In *First International Conference on Computer Imaging Theory and Applications* (pp. 135–140). Lisboa, Portugal. https://doi.org/10.5220/0001793501350140

Ma, Z., & Tavares, J. M. R. S. (2017). Effective features to classify skin lesions in dermoscopic images. *Expert Systems with Applications*, *84*, 92–101. https://doi.org/10.1016/j.eswa.2017.05.003

Ma, Z., Tavares, J. M. R. S., Jorge, R. N., & Mascarenhas, T. (2010). A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Computer Methods in Biomechanics and Biomedical Engineering*, *13*(2), 235–246. https://doi.org/10.1080/10255840903131878

Maglogiannis, I., & Doukas, C. N. (2009). Overview of advanced computer vision systems for skin lesions characterization. *IEEE Transactions on Information Technology in Biomedicine*, *13*(5), 721–723. https://doi.org/10.1109/TITB.2009.2017529

Maglogiannis, I., Ieee, S. M., & Delibasis, K. (2015). Hair Removal on Dermoscopy Images. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2960–2963).

Mathworks. (2018). Contrast-limited adaptive histogram equalization (CLAHE). Retrieved September 11, 2018, from https://uk.mathworks.com/help/images/ref/adapthisteq.html#bviewhb

MATLAB. (2018). imfill-MATLAB. Retrieved September 18, 2018, from https://uk.mathworks.com/help/images/ref/imfill.html

Mendonca, T., Ferreira, P. M., Marques, J. S., Marcal, A. R. S., & Rozeira, J. (2013). PH2 - A dermoscopic image database for research and benchmarking. In *Proceedings of the Annual International Conference of*

the IEEE Engineering in Medicine and Biology Society, EMBS (pp. 5437–5440). https://doi.org/10.1109/EMBC.2013.6610779

Nanni, L. (2006). An ensemble of classifiers for the diagnosis of erythemato-squamous diseases. *Neurocomputing*, *69*(7–9), 842–845. https://doi.org/10.1016/j.neucom.2005.09.007

Nasir, M., Attique Khan, M., Sharif, M., Lali, I. U., Saba, T., & Iqbal, T. (2018). An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. *Microscopy Research and Technique*, *81*(6), 528–543. https://doi.org/10.1002/jemt.23009

Nielsen, R. (1992). Theory of the Backpropagation Neural Network. *Neural Networks for Perception*, 65–93. https://doi.org/https://doi.org/10.1016/B978-0-12-741252-8.50010-8

Nugroho, H., Ahmad Fadzil, M. H., Shamsudin, N., & Hussein, S. H. (2013). Computerised image analysis of vitiligo lesion: Evaluation using manually defined lesion areas. *Skin Research and Technology*, *19*(1), 72–77. https://doi.org/10.1111/j.1600-0846.2011.00610.x

Oliveira, R. B., Marranghello, N., Pereira, A. S., & Tavares, J. M. R. S. (2016). A computational approach for detecting pigmented skin lesions in macroscopic images. *Expert Systems with Applications*, *61*, 53–63. https://doi.org/10.1016/j.eswa.2016.05.017

Oliveira, R. B., Papa, J. P., Pereira, A. S., & Tavares, J. M. R. S. (2018). Computational methods for pigmented skin lesion classification in images: review and future trends. *Neural Computing and Applications*, *29*(3), 613–636. https://doi.org/10.1007/s00521-016-2482-6

Oliveira, R. B., Pereira, A. S., Manuel, J., & Tavares, R. S. (2016). Computational methods for the image segmentation of pigmented skin lesions : A review. *Computer Methods and Programs in Biomedicine*, *131*, 127–141. https://doi.org/10.1016/j.cmpb.2016.03.032

Oliveira, R. B., Pereira, A. S., & Tavares, J. M. R. S. (2018). Computational diagnosis of skin lesions from dermoscopic images using combined features. *Neural Computing and Applications*, 1–21. https://doi.org/10.1007/s00521-018-3439-8

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

PH^2 Database. (2013). PH^2 Database. Retrieved from http://www.fc.up.pt/addi/ph2 database.html

Picardi, A. (2013). Suicide risk in skin disorders. *Clinics in Dermatology*, *31*, 47–56. https://doi.org/10.1016/j.clindermatol.2011.11.006

Ridler, T.W. Calvard, S. (1978). Picture Thresholding Using an Iterative Slection Method. *IEEE Transactions on Systems, Man and Cybernetics*, *8*(8), 630–632. https://doi.org/10.1109/TSMC.1978.4310039

Rosado, L., Jo, M., & Vasconcelos, M. (2015). From Dermoscopy to Mobile Teledermatology. *Dermoscopy Image Analysis*, (Mm), 1–34. https://doi.org/10.1201/b19107-13

Rubegni, P., Cevenini, G., Burroni, M., Perotti, R., Dell'Eva, G., Sbano, P., … Andreassi, L. (2002). Automated diagnosis of pigmented skin lesions. *International Journal of Cancer*, *101*(6), 576–580. https://doi.org/10.1002/ijc.10620

Salem, C., Azar, D., & Tokajian, S. (2018). An Image Processing and Genetic Algorithm-based Approach for the Detection of Melanoma in Patients. *Methods of Information in Medicine*, *57*(1), 74–80. https://doi.org/10.3412/ME17-01-0061

Shamsul Arifin, M., Golam Kibria, M., Firoze, A., Ashraful Amini, M., & Yan, H. (2012). Dermatological disease diagnosis using color-skin images. In *International Conference on Machine Learning and Cybernetics* (Vol. 5, pp. 1675–1680). https://doi.org/10.1109/ICMLC.2012.6359626

Shen, J., & Chan, T. F. (2002). Mathematical Models for Local Nontexture Inpaintings. *SIAM Journal on Applied Mathematics*, *62*(3), 1019–1043. https://doi.org/https://doi.org/10.1137/S0036139900368844

Shenfield, A., Day, D., & Ayesh, A. (2018). Intelligent intrusion detection systems using artificial neural networks. *ICT Express*, *4*(2), 95–99. https://doi.org/10.1016/j.icte.2018.04.003

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer Statistics , 2018, *66*(1), 7–30. https://doi.org/10.3322/caac.21332.

Stanford Vision Lab. (2016). ImageNet Database. Retrieved September 21, 2018, from http://www.image-net.org/

Sultana, A., Dumitrache, I., Vocurek, M., & Ciuc, M. (2014). Removal of artifacts from dermatoscopic images. In *10th International Conference on Communications*. Bucharest, Romania: IEEE. https://doi.org/10.1109/ICComm.2014.6866757

Taufiq, M. A., Hameed, N., Anjum, A., & Hameed, F. (2017). m-Skin Doctor : A Mobile Enabled System for Early Melanoma Skin Cancer Detection Using Support Vector Machine. In K. Giokas, L. Bokor, & F. Hopfgartner (Eds.), *eHealth 360°* (pp. 468–475). Hungry: Springer, Cham. https://doi.org/10.1007/978-3-319-49655-9_57

Tavares, J., & Jorge, N. (2012). Pattern Recognition in Macroscopic and Dermoscopic Images for Skin Lesion Diagnosis. In J. M. R. S. Tavares & R. N. Jorge (Eds.), *Lecture Notes in Computational Vision and Biomechanics* (pp. 504–514). https://doi.org/10.1007/978-94-007-4174-4

Toossi, M. T. B., Pourreza, H. R., Zare, H., Sigari, M. H., Layegh, P., & Azimi, A. (2013). An effective hair removal algorithm for dermoscopy images. *Skin Research and Technology*, *19*(3), 230–235. https://doi.org/10.1111/srt.12015

Übeyli, E. D. (2008). Multiclass support vector machines for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *35*(4), 1733–1740. https://doi.org/10.1016/j.eswa.2007.08.067

Übeyli, E. D. (2009). Combined neural networks for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *36*(3 PART 1), 5107–5112. https://doi.org/10.1016/j.eswa.2008.06.002

Vasconcelos, C. N., & Vasconcelos, B. N. (2017). Experiments using deep learning for dermoscopy image analysis. *Pattern Recognition Letters*, *0*, 1–9. https://doi.org/10.1016/j.patrec.2017.11.005

Victor, A., & Ghalib, M. (2017). Automatic Detection and Classification of Skin Cancer. *International Journal of Intelligent Engineering and Systems*, *10*(3), 444–451. https://doi.org/10.22266/ijies2017.0630.50

Wang, K. F. (2014). Quantitative detection of size and centroid of internal defect based on Radon transform and morphological operation in phase-shifting digital speckle pattern interferometry. *NDT and E International*, *63*, 50–52. https://doi.org/10.1016/j.ndteint.2014.01.008

World Health Organization. (2018). WHO | Skin cancers. Retrieved May 13, 2018, from http://www.who.int/uv/faq/skincancer/en/index1.html

Xie, J., & Wang, C. (2011). Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *38*(5), 5809–5815. https://doi.org/10.1016/j.eswa.2010.10.050

Yasir, R., Rahman, M. A., & Ahmed, N. (2014). Dermatological disease detection using image processing and artificial neural network. In *8th International Conference on Electrical and Computer Engineering* (pp. 687–690). Retrieved from http://proxy.lib.sfu.ca/login?url=http://dx.doi.org/10.1109/ICECE.2014.7026918

Zakeri, A., & Hokmabadi, A. (2018). Improvement in the diagnosis of melanoma and dysplastic lesions by

introducing ABCD-PDT features and a hybrid classifier. *Biocybernetics and Biomedical Engineering*, 1–11. https://doi.org/10.1016/j.bbe.2018.03.005

Zhang, Z. (2017). Artificial Neural Network. In *Multivariate Time Series Analysis in Climate and Environmental Research* (pp. 1–35). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-319-67340-0_1

Zheng, A. (2015). Evaluating Machine Learning Models. Retrieved May 19, 2018, from https://www.oreilly.com/ideas/evaluating-machine-learning-models/page/4/offline-evaluation-mechanisms-hold-out-validation-cross-validation-and-bootstrapping

Zuiderveld, K. (1994). *Contrast limited adaptive histogram equalization*. *Graphics gems IV*. San Diego, CA, USA: Academic Press Professional, Inc.