Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-021-05571-1

RESEARCH ARTICLE-ELECTRICAL ENGINEERING

A Dermoscopic Skin Lesion Classification Technique Using YOLO-CNN and Traditional Feature Model

Ruban Nersisson[1] · Tharun J. Iyer[1] · Alex Noel Joseph Raj[2] · Vijayarajan Rajangam[3]

Received: 17 July 2020 / Accepted: 18 March 2021
© King Fahd University of Petroleum & Minerals 2021

Abstract
Skin cancer is one of the most deadly diseases around the world, wherein one of the three cancers is skin cancer. Early detection of skin cancer is paramount for better treatment planning. This paper investigates a Convolutional Neural Network (CNN), specifically, You Only Look Once (YOLO), to extract features from the skin lesions. The features, obtained from the CNN, are concatenated with traditional features like texture and colour features extracted from the lesion region of the input images. Later, the concatenated features are fed to a Fully Connected Network, which is trained with the specific ground truths to achieve higher classification accuracy. The proposed method improves the detection and classification of skin lesions when compared with other models and YOLO without traditional features. The performance measures of the fusion network are able to achieve the accuracy of 94%, precision of 0.85, recall of 0.88, and area under the curve of 0.95.

Keywords Dermoscopic skin lesions · Convolutional neural network · YOLO · Feature fusion · Transfer learning

1 Introduction

Skin is the largest organ of the body, which is composed of epidermis and dermis. The structure of the skin can perspire and perceive the external temperature to protect the body. Multiple organs and tissues are protected from external invasions by the skin. It also reduces the impact on immune system, skin damage, virus infections, and damage due to chemicals. Despite this defence and barrier function, skin is not indestructible to various genetic and external aspects. Due to the extended exposure to harmful UV rays and chemical damage by acidic wellness products, cancerous substances make their way into the skin and cause various types of skin cancers [1–4]. The growing concern of skin lesions has attracted many group of researchers for skin cancer detection and classification algorithms towards effective treatment planning [5,6]. Over the past few decades, image processing techniques have made rapid development in medical image analysis. Dermoscopy also brought upon research into various methods of processing images from the skin surface [7,8]. For example, various types of melanoma were detected by Jain and Shivangi [9] through texture, colour and shape analysis using segmentation methods. Barata and Ruela [10] put forward a method to detect whether a skin lesion is benign or malignant based on texture and colour features, which is further investigated by Nezhadian and Rashidi [11]. Ashour and Amira [12] developed a novel optimized K-means algorithm using genetic algorithm for skin lesion classification. Abuzaghleh et al. [13] developed a tool to identify skin lesions in real time. Skin lesion classification has become more advanced using various machine learning algorithms to detect and classify skin lesions. Mhaske and Phalke [14] proposed a system to detect and classify melanoma using a neural network and support vector machine (SVM). The role of SVM is further analysed by many research teams such as Alquran et al. [15]. Aurora Saez et al. [16] proposed a machine learning method for the classification of melanoma from dermoscopic images. With the advent of deep learning, Deep Neural Networks (DNN) and CNN have found their way into skin lesion classification. Andre Esteva et al. [17] proposed a system for dermatologist level classification of skin cancer images using a Deep Convo-

✉ Alex Noel Joseph Raj
jalexnoel@stu.edu.cn

1 School of Electrical Engineering, Vellore Institute of Technology, Vellore, India

2 Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Department of Electronic Engineering, College of Engineering, Shantou University, Shantou 515063, China

3 Centre for Healthcare Advancement, Innovation and Research, Vellore Institute of Technology, Chennai, India

Published online: 07 April 2021

lutional Neural Network (DCNN). Tri-Cong Pharm et al. [18] proposed data augmentation to increase the instances of each image in the dataset that improved the classification parameters. DeVries and Ramachandran [19] proposed a Deep Multi-scale CNN based on the features known as Asymmetrical shape, Borders, Colour and Diameter (ABCD) to classify skin lesions. Peng and Saenko [20] proposed a novel fusion CNN based on two separate networks trained from texture and shape features. This CNN outperforms web image-based models, Computer-Aided Design (CAD)-based shape models, and weakly supervised models. Saba et al. [21] proposed a new automated method for skin lesion classification using an Inception V3 model fused with different features. When validated on PH2, ISBI 2016, and ISBI 2017 datasets, the proposed model outperforms several existing models. Muhammad Rashid et al. [22] proposed a technique where Scale Invariant Feature Transform (SIFT) features are used in a DCNN. The technique outperforms several methods when validated on the Caltech101, Barkley 3D, and Pascal 3D datasets. Serte and Demirel [23] proposed a novel Gabor wavelet-based DCNN for the detection of malignant melanoma. The Gabor-based approach provides directional decomposition where each sub-band presents decisions that can be fused for improved classification performance.

In this paper, CNN, based on the YOLO architecture, named as YOLO-CNN, is employed as the fundamental architecture for feature extraction. Traditional features (TF) like texture and colour features are concatenated with the features generated from YOLO-CNN (YF). The objective of the proposed method is to classify malignant and benign skin lesions. The proposed network is named as YOLO Traditional Features Model (YTFM).

The whole process can be divided into the following stages:

(a) Pre-processing of input images. The images are resized to 416 × 416,
(b) Extraction of texture features using Gray-Level Co-occurrence Matrix (GLCM) and colour features using Colour Level Co-occurrence Matrix (CLCM),
(c) Concatenation of texture–colour features and features extracted from the YOLO-CNN.

The concatenated features are fed to a Fully Connected Network (FCN) which is trained for better classification measures. The classification results of YTFM are compared with the classification outputs generated directly from the YOLO network without feature concatenation. Similar comparisons are presented for other fused and unfused DNN models. All the models are tested on the same dataset. The remaining sections of the paper are organized as follows: The CNN architecture, feature extraction techniques, and feature fusion method are presented in Sect. 2. The proposed methodology, YTFM, is presented in Sect. 3. The experimental results, comparisons, and discussions are elaborated in Sect. 3, which is followed by conclusions in Sect. 4.

2 Methodology

Over a decade, various Region-based Convolutional Neural Networks (R-CNN) like Fast R-CNN, Faster R-CNN, and YOLO are prevalent used for object detection and classification. While Fast R-CNN and Faster R-CNN are the algorithms well suited for classification, YOLO is more suited towards regression. The object detection process in Fast R-CNN and Faster R-CNN is twofold. First, the regions are selected within an image and further classified using CNN. The solution can be slow because it needs to run predictions for every selected region. However, in YOLO, the image is predicted in one forward pass of the algorithm instead of selecting interesting parts of the image. For skin lesion classification, Sorokin [24] proposed a segmentation based CNN for accurate segmentation of the lesion. Taqi et al. [25] proposed an SSD-MobileNet algorithm based on the Android platform for skin lesion detection. In [26], S.S Roy et al. employed YOLO for the diagnosis of melanoma by exploiting the processing power of the YOLO architecture. The authors reported an accuracy of 86% along with the least computation time, thus surpassing other state-of-the-art DNN networks. The YOLO algorithm predicts a class label with bounding box over the object of interest. For the proposed framework, YOLO was chosen since it presents both higher accuracy and processing speed.

2.1 YOLO-CNN

YOLO architecture [27] is more like a traditional Fully Convolutional Neural Network (F-CNN). The YOLO architecture splits the input image in an M × M grid. For each grid, two bounding boxes and class probabilities are generated. A single CNN is able to predict class probabilities and multiple bounding boxes for the Region of Interest(ROI). Detection performance is optimized by the YOLO-CNN which is trained on full images.

The main advantage of YOLO is three-folds. Firstly, YOLO is fast. The base network runs at 45 frames per second (fps), and the quickest version runs at more than 150 fps. Secondly, YOLO-CNN uses the full image for training and testing. This factor enforces the use of contextual information about classes and their appearance. Conversely, Fast R-CNN produces false alarms by considering background patches for objects due to reduced view of the field. Thus, in comparison, YOLO performs better than R-CNN, as it commits less than half the background errors. Thirdly, YOLO learns the general representations of objects. YOLO performs better than

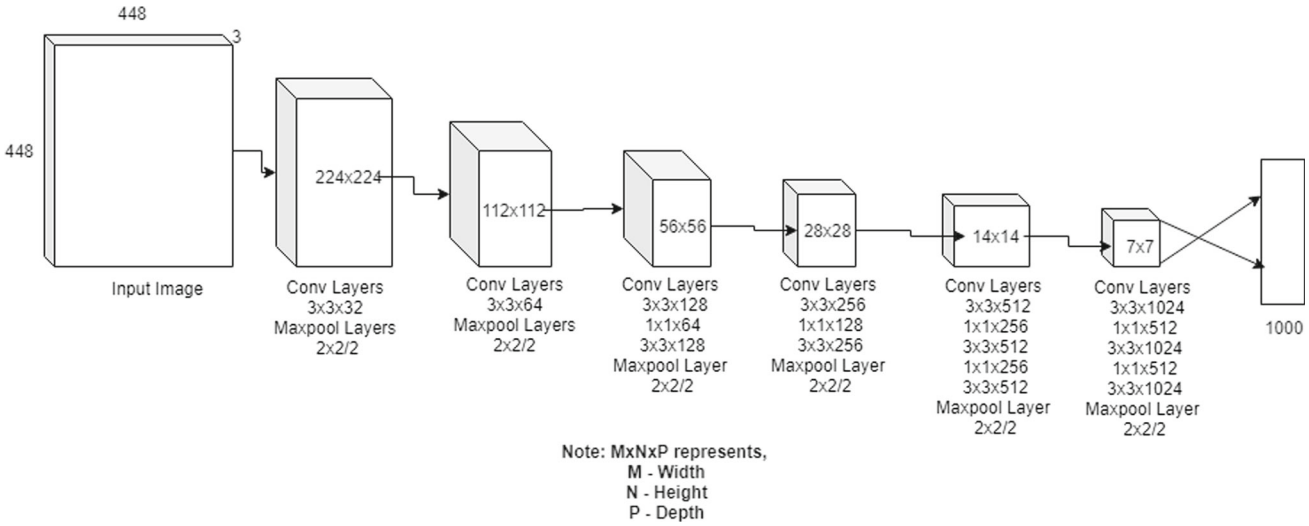

Fig. 1 YOLO v2 architecture

Deformable Parts Model (DPM) and R-CNN when trained on artwork and natural images. Since YOLO is highly generalized, it is less likely to present false alarms when applied to new domains or unexpected inputs.

The version of YOLO, used in this work, is the YOLO v2 based on Darknet-19, which consists of a 19 layers deep network. The architecture of the Darknet is shown in Fig. 1. For the proposed work, a few modifications in the architecture have been introduced to suit our dataset and needs.

It is proposed to use batch normalization on the convolutional layers to enhance the convergence and eliminate the need for other regularization methods. It uses 448 × 448 high-resolution classifiers. Darknet does not predict the coordinates for bounding boxes from the fully connected layers and instead uses anchor boxes. YOLO v2 uses five anchor boxes whose dimensions are related to the input image. It is reduced by a factor of 32. The network predicts five bounding boxes, each having five coordinates: t_x, t_y, t_w, t_h, and t_o. These coordinates are the parameters of each annotation in the dataset: the width, height, x-coordinate, y-coordinate, and the class label. The predictions b_x, b_y, b_w, and b_h are the coordinates of the predicted bounding box. The offsets of the cell (c_x, c_y) represent the length of each cell in the image, and the dimensions of the bounding boxes (p_w, p_h) are calculated as shown in Fig. 2.

The predictions are computed as given below:

$$b_x = \sigma(t_x) + c_x \tag{1}$$

$$b_y = \sigma(t_y) + c_y \tag{2}$$

$$b_w = p_w e^{t_w} \tag{3}$$

$$b_h = p_h e^{t_h} \tag{4}$$

$$p_r(object) * IOU(b, object) = \sigma(t_o) \tag{5}$$

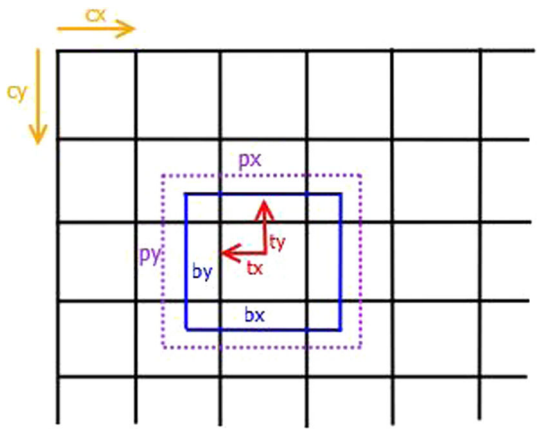

Fig. 2 Calculation of bounding box offsets and dimensions

where p_r is the conditional class probability; IOU is intersection over union between predicted bounding box and ground truth; and $\sigma(t_o)$ is change in class label.

By using anchor boxes, the input size is reduced to 416 × 416. The last layer of the network is structured with 1000 filters followed by a final layer with variable number of filters. The number of filters for this layer depends on the number of classes being detected as given below:

$$N_{\text{filters}} = 5 * (5 + N_{\text{classes}}) \tag{6}$$

For $N_{\text{classes}} = 2$, the number of filters is 35 in the last layer. The layer specification of Darknet-19 for YOLO v2 is presented in Table 1. While the YOLO algorithm uses region information to extract features and detect objects, the concatenation of the feature maps with popular texture and

Table 1 Layer specification of Darknet-19 architecture for YOLO v2

Type	Filter	Size/stride	Output
Convolutional	32	3 × 3	224 × 224
Maxpool		2 × 2/2	112 × 112
Convolutional	64	3 × 3	112 × 112
Maxpool		2 × 2/2	56 × 56
Convolutional	128	3 × 3	56 × 56
Convolutional	64	1 × 1	56 × 56
Convolutional	128	3 × 3	56 × 56
Maxpool		2 × 2/2	28 × 28
Convolutional	256	3 × 3	28 × 28
Convolutional	128	1 × 1	28 × 28
Convolutional	256	3 × 3	28 × 28
Maxpool		2 × 2/2	14 × 14
Convolutional	512	3 × 3	14 × 14
Convolutional	256	1 × 1	14 × 14
Convolutional	512	3 × 3	14 × 14
Convolutional	256	1 × 1	14 × 14
Convolutional	512	3 × 3	14 × 14
Maxpool		2 × 2/2	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	512	1 × 1	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	512	1 × 1	7 × 7
Convolutional	1024	3 × 3	7 × 7
Avgpool	1000	1 × 1	1000
Softmax		Global	

colour features is employed to increase the performance of the algorithm.

In addition to the object detection CNN, the features are concatenated with the feature map to obtain higher performance measures. Texture and colour features are extracted from the input image and fed to an FCN, which is trained to detect benign and malignant skin lesions. Texture and colour feature extraction techniques are elaborated in the following section.

2.2 Texture Feature Extraction

The texture features for the image are extracted using the GLCM and Gabor method. The GLCM and Gabor features are highly popular texture descriptors that provide accurate information about the texture of an image. GLCM characterizes the texture of an image by calculating the frequency of pairs of pixels with specific values and a specified spatial relationship in the image. Various statistical measures are then extracted from this matrix, which provides textural information of the image. The GLCM features are chosen as one of the texture features for this study due to less complexity in extraction and visually separable property. GLCM features are based on second order statistics and represents the image well in terms of texture features. Moreover, GLCM has been used in many different studies as a prime texture analysis method which is the reason for choosing it for this study. Also, it is easy to extract and gives good results for many different applications. The GLCM features provide texture information of the image in a spatial domain which will provide with high quality image information that can be used in a feature fusion technique. The statistical descriptors and their description are given in Table 2.

Here, $p(i, j)$ is used to access the pixel values of the image at the spatial coordinate i and j. Each measure returns a 416 × 416 matrix, which is used to describe each pixel of the image. Figure 3 illustrates GLCM feature extraction for a skin lesion.

The other feature used to describe the texture is the Gabor feature. The Gabor features are excellent texture descriptors used for the analysis of a specific frequency content along with the chosen directions of the region of analysis. This method is analogous to the human visual system in recognizing the frequency and orientation of an object. Here, the Gabor filter coefficients are calculated over four different orientations; 0°, 45°, 90°, and 135°. Gabor features are chosen for this study due to their broad applications in medical imaging as a texture feature extraction tool. When coupled with the GLCM, Gabor filters provide high-quality information on the entire texture of the image. Moreover, Gabor filters can be designed such that only relevant information can be extracted. This proves to be better for our study than other frequency based methods like Discrete Cosine Transform or Fourier Transform. In the discrete domain, which is often used for feature extraction, two-dimensional Gabor filters are given by:

$$G_c|i, j| = B.\mathrm{e}^{-\frac{(i^2+j^2)}{2\sigma^2}} \cos(2\pi f(i\cos\theta + j\sin\theta)) \quad (11)$$

$$G_c|i, j| = C.\mathrm{e}^{-\frac{(i^2+j^2)}{2\sigma^2}} \sin(2\pi f(i\cos\theta + j\sin\theta)) \quad (12)$$

Here, B and C are normalizing factors to be determined, and f defines the frequency of a texture. θ is the orientation of the Gabor filter. σ is the variable for changing neighbourhood distance. In this method, a neighbourhood distance of 1 is chosen. Visualization of the Gabor features is presented in Fig. 4.

2.3 Colour Feature Extraction

For extracting colour features of an image for classification of Skin lesions, many methods including colour moments [28,29] and colour histograms [30–32] have been proposed. In addition to colour features, a few methods have been proposed to exploit global texture features. Local Binary Pat-

Table 2 Layer specification of Darknet-19 architecture for YOLO v2

Statistic	Description	Justification	Formula
Contrast	This measure returns the intensity contrast between a pixel and its neighbour over the image	This feature can help in measuring the amount of local variations in the image	$\Sigma_{i,j}\lvert i-j\rvert^2 p(i,j)$ (7)
Homogeneity	This measure measures the similarity between the pixels	This measure helps in finding out the degree to which each pixel differs from the other	$\Sigma_{i,j}\frac{p(i,j)}{1+\lvert i-j\rvert}$ (8)
Correlation	This measure returns the correlation between a pixel and its neighbour over the whole image	This feature can help in measuring how much a pixel relates to the whole image	$\Sigma_{i,j}\frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j}$ (9)
Energy	This measure measures the sum of squared elements in neighbourhood. It is also known as uniformity	This measure helps in finding out the disorders in the texture of the image	$\Sigma_{i,j}p(i,j)^2$ (10)

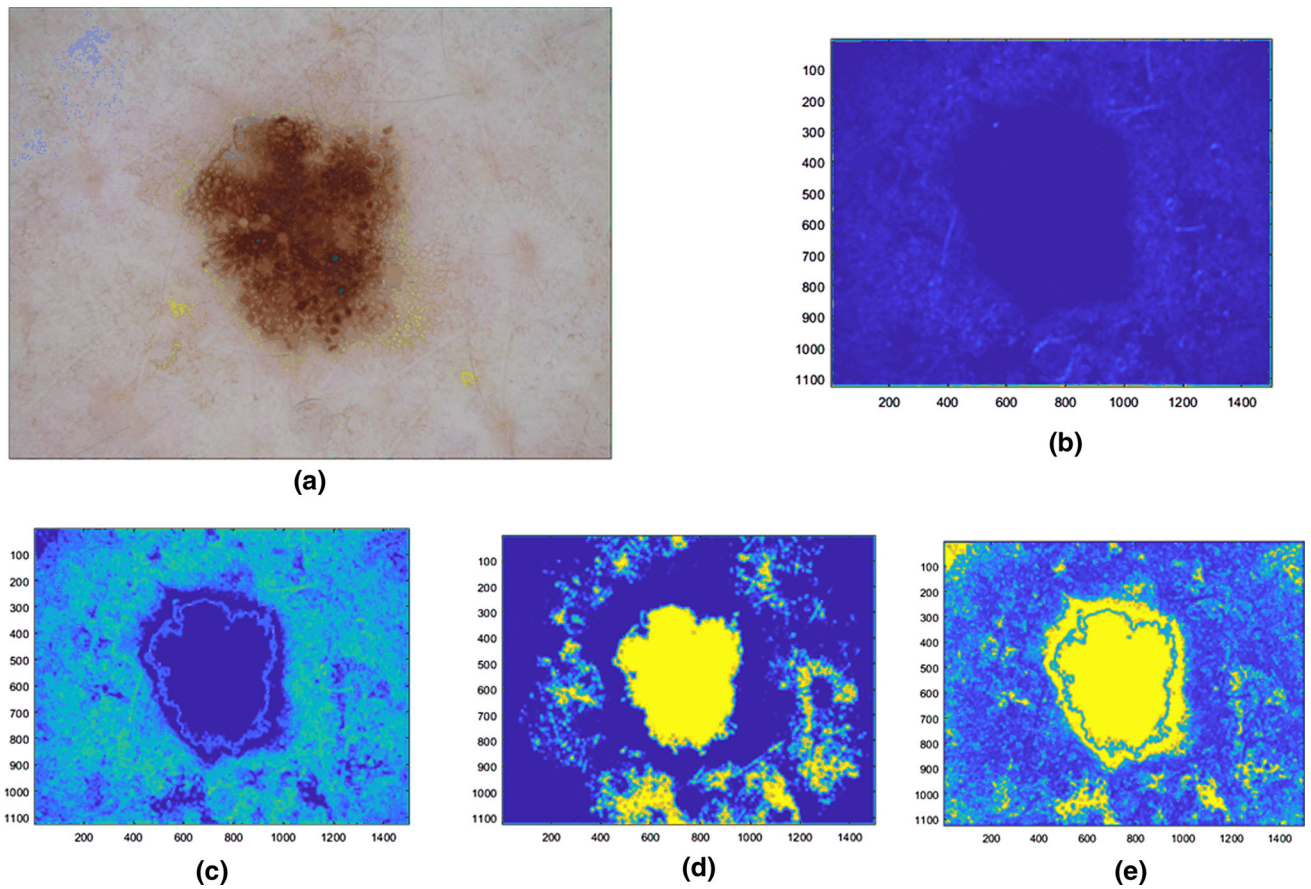

Fig. 3 **a** Original skin lesion image, **b** contrast GLCM matrix, **c** correlation GLCM matrix, **d** energy GLCM matrix, **e** homogeneity GLCM matrix

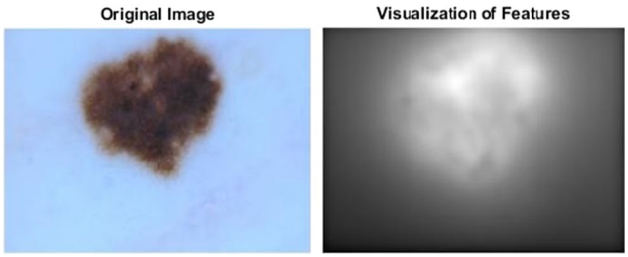

Fig. 4 Visualization of Gabor features

tern (LBP) is used as a global texture and colour descriptor. While the use of LBP has been widespread, more operators based on the concept of LBP have been proposed. Orthogonal Combination of Local Binary Pattern (OC-LBP) [33–35] has been proposed as a colour descriptor for dermoscopic skin lesions. The method of OC-LBP is shown to outperform the original LBP operator by 14.1% with an admissible trade-off between speed and accuracy. Other methods include a complete local similarity pattern extended for colour images [36]. A colour feature extraction method based on GLCM technique is employed in [37], wherein the original GLCM procedure is extended to the three colour channels, R, G, and B. For all the three channels, the single and multi-integrative co-occurrence matrices are extracted. The working procedure to extract the CLCM Feature Vector is shown in Fig. 5. Thus, a feature vector of dimensions, 208 × 208 × 8, is obtained for each image. Each feature vector for one channel is of the size 208 × 208. There are three channels: R, G, and B for single co-occurrence matrix. There are three other channels: RG, GB, and BR for multi co-occurrence matrices. The single and multi-co-occurrence for the original image is then calculated as two separate channels, which increases the dimensions to 8. The feature vector is then reshaped to 416 × 416 × 2 for proper concatenation. As compared to the OC-LBP methods, the CLCM is less complex and requires less computation power. It provides more information for the classifier to make a decision and will help in increasing the classification score.

2.4 Principal Component Analysis

PCA uses the idea of dimensionality reduction to standardize the data by computing a covariance matrix of the target matrix and its corresponding eigenvectors. Finally, the initial data are transformed into a linearly independent representation of arbitrary dimensions using a linear transformation, which then transforms multiple indicators (high dimensions) into a few major feature components.

In this paper, PCA is used to reduce the dimensionality of the concatenated features so as to further concatenate with the feature map obtained from YOLO. The texture and colour features are combined into one matrix of size 416 × 416 ×

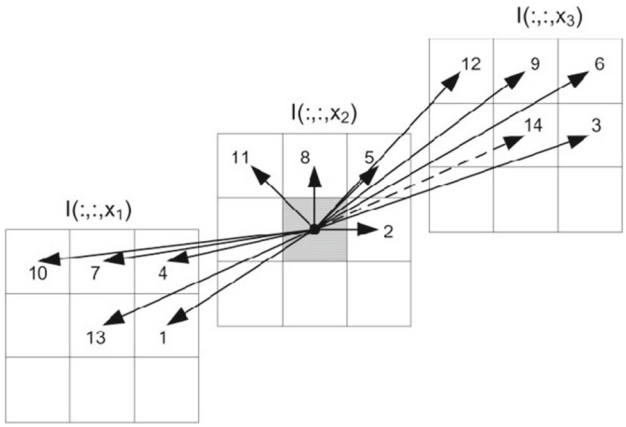

Fig. 5 Operating procedure of CLCM Vector

7. This is the input to the PCA, and the output from the PCA is 7 × 7 × 1024. The feature map obtained from YOLO is of size 7 × 7 × 1024. Both feature maps are flattened and then concatenated to form a fused feature map. This feature map is fed to the FCN through the feature fusion block which is trained to obtain the output.

2.5 Feature Concatenation

The proposed YTFM is divided into two parts: the YOLO-CNN and the texture-colour feature extraction. The CNN model is used to detect the lesion and assign a preliminary confidence score with one dimension. The texture and colour features are then combined to form the second dimension for the model. Then, feature vector fusion is performed in the concatenated layer. The final classification result is the output by using a softmax classifier.

2.6 Transfer Learning

In the field of healthcare, obtaining a properly labelled and accurate data is always an obstacle since it relies on the manual annotation by clinicians. With the data being insufficient for training, transfer learning becomes a good choice as the training time decreases, and the model can maintain a good generalization.

When it comes to transfer learning, a pre-trained model as a feature extractor is widely used. For this paper, two pre-trained models trained on the ImageNet [38], MS COCO [39], and PASCAL VOC datasets are used. ImageNet is a large database of images for the research of visual object detection and recognition. There are 15 million manually annotated images spread over 1000 classes. The PASCAL VOC 2007 dataset contains 9963 images spread over 20 classes. The MS COCO dataset contains 330,000 images spread over 80 classes.

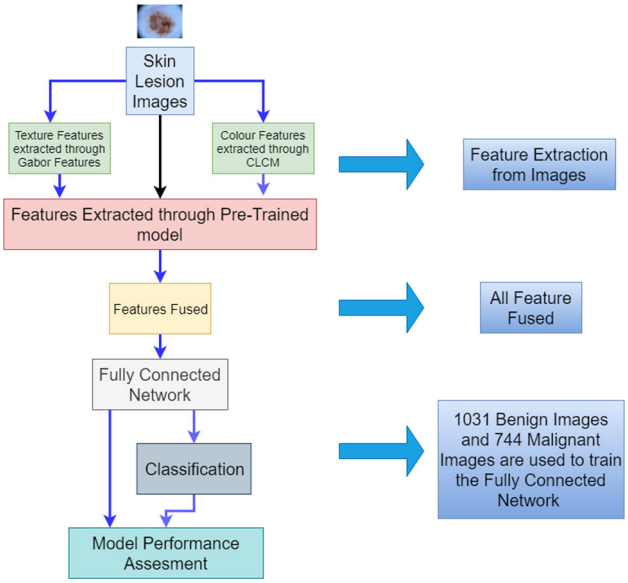

Fig. 6 Block diagram of methodology

In the Darknet-19 model, the input size is 416 × 416 × 3, with an output layer consisting of 1000 neurons. All the convolution layers employ 3 × 3 filters, and the number of filters in the last layer is 35. The pooling layer is a global average pooling layer.

3 Dermoscopic Skin Classification Using YTFM

In conventional detection and classification systems, region information is extracted by the network for detecting and classifying the object in question. From [20], it is understood that previous works have fused two different networks, extracting texture and shape features to classify objects. The objective of the proposed work is to increase the classification and detection accuracies of the stand-alone CNN by extracting the lesion information to concatenate with texture-colour features. The proposed method is a threefold process, as shown in Figs. 6 and 7.

1. Identify the region information from the input image and extract features using CNN.
2. Extract texture and colour information from the training images.
3. Concatenate all the features and feed FCN for training and classification of lesions.

4 Experimental Analysis

Two experiments were conducted to analyse the performance of the proposed method. The first was to run the YTFM and compare the results with the YOLO v2 model. The other experiment was designed for comparing the YTFM with other detection models.

4.1 Dataset

The dataset used in this work is the ISBI Melanoma dataset from 2016 ISIC skin lesion classification challenge. The dataset contains benign and malignant dermoscopic melanoma images [40]. There are 1031 benign images and 248 malignant images. Since the number of malignant images is much less than the number of benign images, image augmentation is used to increase the number of images to 744 images. Rotation and distortion were used to augment each malignant image for producing two other images. Training data account for 80%, and testing data account for 20%.

4.2 Configuration for Experiments

In the proposed model, the learning rate used was 0.001, and the batch size is 16. Batch Normalization is also added to the Fully Connected layer so that the input is distributed equally during training at each layer. The experiments were conducted using Google Colaboratory, an online hosted GPU to run Python Notebooks. The YOLO version used was Darkflow, an extension of Darknet modified for Tensorflow. The first experiment was trained using 1000 epochs of data. YOLO uses a squared sum error to calculate the loss. The loss function comprises a localization loss (returns error between the predicted bounding box and ground truth box and confidence loss objectiveness of the box) and classification loss. The loss function can be defined as,

$$\begin{aligned}
&\lambda_{\text{coord}} \Sigma_{i=0}^{s^2} \Sigma_{j=0}^{B} 1_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 - (y_i - \hat{y}_i)^2] \\
&+ \lambda_{\text{coord}} \Sigma_{i=0}^{s^2} \Sigma_{j=0}^{B} 1_{ij}^{\text{obj}} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 - (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\
&+ \Sigma_{i=0}^{s^2} \Sigma_{j=0}^{B} 1_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\
&+ \lambda_{\text{noobj}} \Sigma_{i=0}^{s^2} \Sigma_{j=0}^{B} 1_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\
&+ \Sigma_{i=0}^{s^2} 1_{ij}^{\text{obj}} \Sigma_{c \epsilon \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned} \tag{13}$$

where 1_i^{obj} denotes the object that appears in cell i. 1_{ij}^{obj} denotes that the jth bounding box predictor in cell i is responsible for that prediction, $\hat{p}_i(c)$ denotes the class probability for class c in cell i, λ_{coord} increases the weight for loss in bounding box coordinates, $\hat{C}_i$ is the confidence score of box j in cell i, λ_{noobj} weighs down the loss when detect-

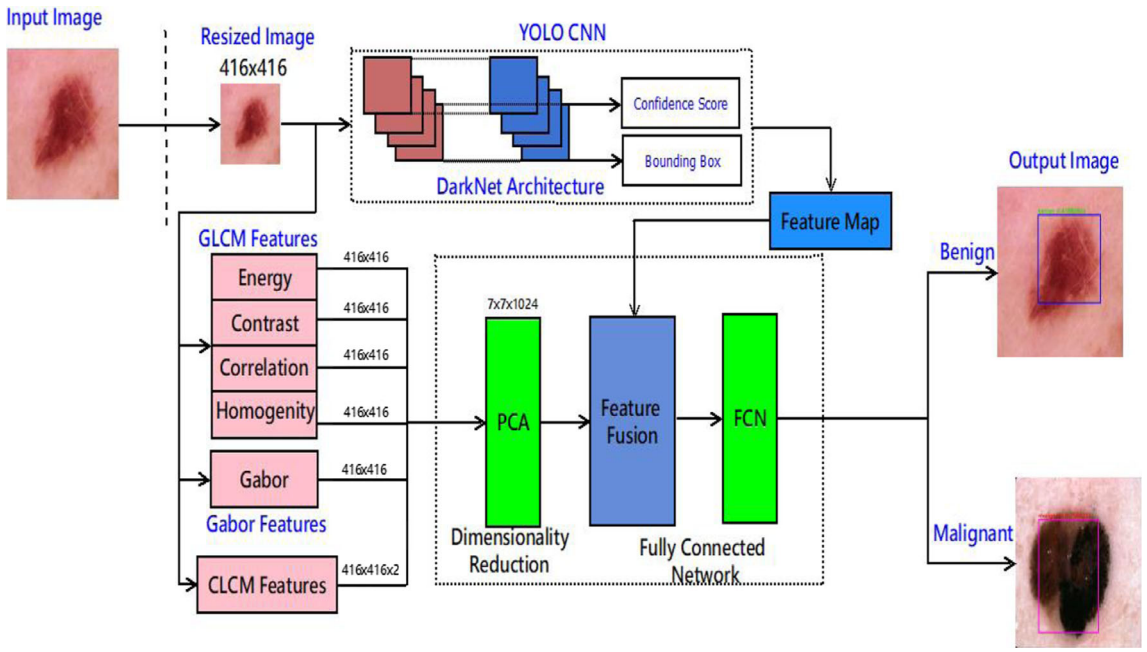

Fig. 7 Expanded block diagram

ing background. The FCN is trained for 1000 epochs of data with Adam optimizer and the same custom loss as defined above.

In the second experiment, pre-trained models of YOLO v2 (PASCAL VOC), MobileNet (ImageNet), Single Shot Detector (SSD)-MobileNet (MS COCO), and Faster R-CNN (ImageNet) are used to compare the performance and feature extraction ability. The above models were trained over 1000 epochs and also on the Google Colab platform. For Faster R-CNN, an SGD optimizer is used, and the RPN loss function, defined below, is used for optimization:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{\text{cls}}}\Sigma_i L_{\text{cls}}(p_i, p_i^*) + \lambda\frac{1}{N_{\text{reg}}}\Sigma_i p_i^* L_{\text{reg}}(t_i, t_i^*) \quad (14)$$

Here, N_{cls} and N_{reg} are the size of the batch and the number of anchor locations, respectively. μ is a weight parameter. i is the index of an anchor in the batch, and p_i is the predicted probability of the corresponding anchor being an image. p_i^* is the ground truth. t_i and t_i^* are coordinates of the vector representing predicted bounding box and ground-truth, respectively. The classification loss, L_{cls}, is the log loss over two classes: benign or malignant. The regression loss is defined as $L_{\text{reg}}(t_i, t_i^*) = R(t_i - t_i^*)$, where R is the robust loss function. SSD-MobileNet and MobileNet use binary cross-entropy as the loss function and an Adam optimizer. The binary cross-entropy loss function is defined as

Table 3 Four possibilities of the classification algorithm

Label	Prediction	
	Benign	Malignant
Benign	True positive (TP)	False positive (FP)
Malignant	False negative (FN)	True negative (TN)

$$L(p, y) = -(y\log(p) + (1-y)\log(1-p)) \quad (15)$$

where p is the predicted class, y is the ground truth, and log is the natural log function.

4.3 Performance Measures

Accuracy, sensitivity, F1-score, and area under the curve (AUC) are used as performance measures for the classification algorithms. For labelled classes and predicted classes, four possibilities can occur as shown in Table 3.

Accuracy is the number of correctly classified images to the total number of images as given by

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \quad (16)$$

Precision and recall are the indices which represent accuracy of a binary classification model. Precision is defined as the

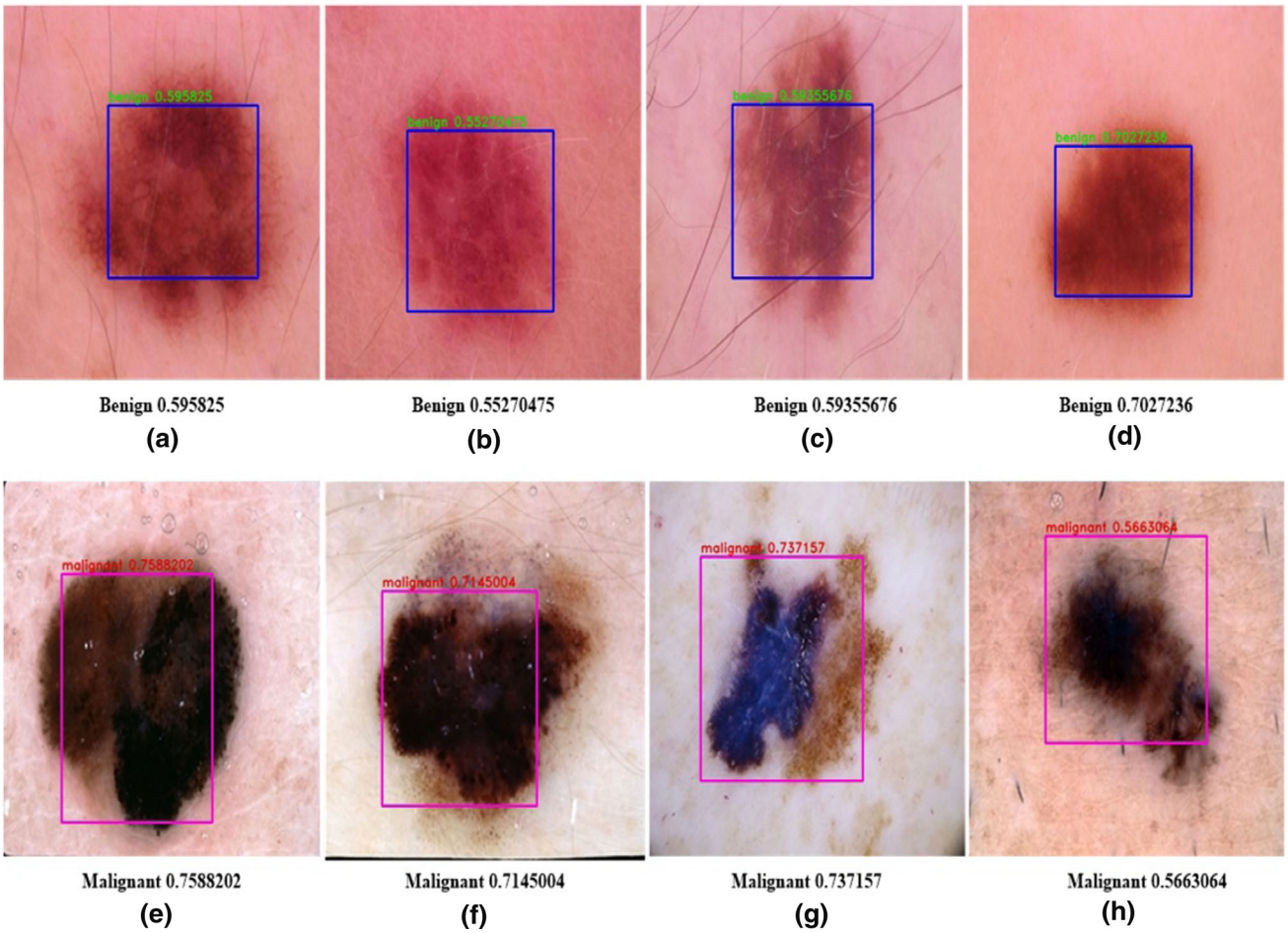

Fig. 8 Results from YTFM **a**–**d** Benign and **e**–**h** Malignant

number of positive predictions that actually belonging to the positive class. Recall is defined as the number of positive predictions made out of all the positive cases in the dataset.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{17}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

The Receiver Operating Characteristic (ROC) curve is drawn by pitting the True Positive Rate (TPR) on the y axis and the False Positive Rate (FPR) on the x axis. AUC is then calculated by calculating the area under this curve using the trapezoidal rule. The curve is divided into multiple closed sub intervals. The AUC is then calculated by the summation of the area of the trapezoids formed from these sub intervals.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{19}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{20}$$

AUC value is an index defined to measure the quality of a binary classification model.

4.4 Results

The proposed YTFM is experimented for the classification of benign and malignant lesions for the ISBI Melanoma dataset. The detection and classification results from the YTFM are presented in Fig. 8. The bounding boxes represent the region of the lesions with the accuracy of prediction. The extracted and traditional features from the bounding boxes are concatenated for the classification problem.

The experimental results are shown in Table 4. YF is the model trained without the traditional features, and YTFM is the model trained with the traditional features. It is observed that the proposed model delivers an accuracy of 94%, which is 5% higher than the YF model.

Table 5 illustrates the performance of the proposed feature fusion method over other models as mentioned. Faster R-CNN is an object detection algorithm that works similar to YOLO or SSD. The main concept behind Faster R-CNN

Table 4 Performance of feature concatenation on YOLO

Model	Option	Accuracy (%)	Precision	Recall	AUC
YOLO v2 (PASCAL VOC)	YF	89	0.9	0.85	0.91
	YTFM	94	0.92	0.88	0.95

Table 5 Performance of feature fusion on other models

Model	Accuracy (%)	Precision	Recall	AUC
MobileNet (ImageNet)	80	0.79	0.83	0.87
SSD-MobileNet (MS COCO)	90	0.7	0.79	0.85
Faster R-CNN (ImageNet)	92	0.78	0.71	0.88

Table 6 Ablation study of features in the feature fusion model

Features	Accuracy (%)	Precision	Recall	AUC
GLCM + Gabor + CLCM	94	0.92	0.88	0.95
GLCM + Gabor	90	0.87	0.9	0.91
CLCM + Gabor	91	0.86	0.89	0.88
Gabor + CLCM	88	0.85	0.9	0.89
Gabor	87	0.88	0.87	0.84
CLCM	86	0.85	0.82	0.87
GLCM	90	0.88	0.85	0.86

is to extract features through a CNN, propose regions to classify and pooling to classify the content in a bounding box or discard it as background information. Unlike YOLO being a SSD, Faster R-CNN extracts the features and classifies the regions in separate runs rather than one single run. MobileNet is a contrasting model to YOLO and Faster R-CNN. It was designed to be a low-power and compact model for efficient object detection. MobileNet was designed to infer images quickly at the expense of accuracy. It is proposed to use the MobileNet trained on two different weights from ImageNet and MS-COCO datasets to find the variance in performance. Since ImageNet offers better generalisation over MS-COCO, it has a better performance overall. Among the other models, Faster R-CNN performs better in terms of accuracy and AUC. MobileNet trained on ImageNet shows better performance than MobileNet trained on MS COCO in terms of precision, recall, and AUC. This is due to better generalization in ImageNet over MS COCO. Over ImageNet, Faster R-CNN performs better in terms of accuracy showing a 12% difference. But MobileNet performs better in terms of recall while precision and AUC are quite similar. This could be caused due to the dataset, and further research on more data would explain this difference.

4.5 Ablation Study

To provide more information on the effectiveness of our features, an ablation study has been performed using the features extracted as shown in Table 6. The ablation study on the YTFM model is performed using the texture and colour features.

Table 7 Confusion matrix—training

	YOLO without TF		YTFM	
	Benign	Malignant	Benign	Malignant
Benign	537	77	382	33
Malignant	92	721	52	953

From the ablation study given in Table 6, it is observed that the fusion of the three feature types performs better against other combinations. But, amongst the features, GLCM proves to be the best performing feature followed by Gabor Features. GLCM alone provides as good performance as a combination of CLCM and Gabor features. This proves that GLCM offers more information than Gabor Features and CLCM. But the fusion of all three features performs better than the traditional YOLO v2 model.

From Table 7, the distribution of TP, TN, FP, and FN in the dataset can be seen as detected by the proposed method. Ideally, the number of FP and FN, also known as Type I and II errors, should be zero with the number of TP and TN being maximum. With this method, the number of TP and TN is significantly greater than the number of FP and FN, which highlights the good performance of the method. Confusion matrix shows the performance of the model in simple terms. As can be seen, there is a decrease in the number of FP and FN as detected by the YTFM over just the YOLO without TF

Table 8 Confusion matrix—testing

	YOLO without TF		YTFM	
	Benign	Malignant	Benign	Malignant
Benign	136	15	175	15
Malignant	24	180	24	140

in the training dataset. From Table 8, the Confusion Matrix for the testing dataset is quite similar between feature fusion and the CNN without feature fusion. With more data provided to the network, better and detailed performance can be researched upon.

5 Discussions

As can be seen from the obtained results, feature fusion proves to improve the classification of dermoscopic melanoma images. The concatenation of CNN features with traditional features significantly improves the information available for classification. Due to this, the classification and parameters affecting the classification have improved. While the accuracy of the system may seem similar to previous works, the classification indices are better due to the effect of the feature concatenation. This is because pre-trained models use a singular method to extract features from an image, which is then used for classification. The complexity of colour images and the fact that extra information can be obtained through different methods and feature concatenation proves to be a better model to deal with skin cancer images containing hair, sweat, etc. Also, the algorithm is invariant to hair or any external effects.

6 Conclusion

In the medical and healthcare industry, the early diagnosis of diseases is vital for treatment planning. If the early-stage diagnosis of diseases is made, then treatment and subsequent prevention are key to survival. With various skin diseases appearing due to pollution and other factors, the classification of surface-level skin cancers and diseases is becoming an essential field of research. It is also hard to obtain labelled data, which would make it easier for classification models. Also, the poor image quality, blurred regions, and noise like hair or sweat in dermoscopic images may affect training the models from scratch or using transfer learning to fine-tune the models for achieving good results. The proposed method of extracting colour–texture features from dermoscopic images and concatenating with the CNN features for subsequent training of FCN shows that the performance of the classifier is improved. The performance of the proposed method in terms of accuracy and the classification indexes is better than other models. The YTFM model is able to achieve the accuracy of 94%, precision of 0.85, recall of 0.88, and area under the curve of 0.95.

Acknowledgements The authors would like to thank Memorial Sloan-Kettering Cancer Center for their contribution in establishing the ISIC Archive. We would also like to thank Gutman et al. [40] for making their database openly available. The authors would like to thank Vellore Institute of Technology, Vellore, for providing the required support to carry out this research work.

Funding This research was financially supported by the Scientific Research Grant of Shantou University, China, Grant No: NTF17016.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with animals performed by any of the authors. All procedures performed in studies involving human participants were in accordance with the ethical standards. The dataset used for experimentation is ISBI Melanoma dataset obtained from the 2016 International Skin Imaging Collaboration (ISIC) skin lesion classification challenge. The dataset is publicly available at https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery, and many researchers have used these images to experimentally verify and compare their works. As required we have also included the reference [40] which is related to their original publication. Further, we have also recognized their support by thanking them under the acknowledgment section.

References

1. Baldwin, L.; Dunn, J.: Global controversies and advances in skin cancer. Asian Pac. J. Cancer Prev. **14**(4), 2155–2157 (2013)
2. Pfeifer, G.P.; Besaratinia, A.: UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. Photochem. Photobiol. Sci. **11**(1), 90–97 (2012)
3. Craythorne, E.; Al-Niami, F.: Skin cancer. Medicine **45**(7), 431–434 (2017)
4. Tracey, E.H.; Vij, A.: Updates in melanoma. Dermatol. Clin. **37**(1), 73–82 (2019)
5. Jerant, A.F.; Johnson, J.T.; Sheridan, C.D.; Caffrey, T.J.: Early detection and treatment of skin cancer. Am. Fam. Phys. **62**(2), 357–368 (2000)
6. Sreelatha, T.; Subramanyam, M.V.; Prasad, M.N.G.: Early detection of skin cancer using melanoma segmentation technique. J. Med. Syst. **43**(7), 190 (2019)
7. Massone, C.; Di Stefani, A.; Soyer, H.P.: Dermoscopy for skin cancer detection. Curr. Opin. Oncol. **17**(2), 147–153 (2005)
8. Wolner, Z.J.; Yélamos, O.; Liopyris, K.; Rogers, T.; Marchetti, M.A.; Marghoob, A.A.: Enhancing skin cancer diagnosis with dermoscopy. Dermatol. Clin. **35**(4), 417–437 (2017)
9. Jain, S.; Pise, N.; et al.: Computer aided melanoma skin cancer detection using image processing. Procedia Comput. Sci. **48**, 735–740 (2015)

10. Barata, C.; Ruela, M.; Francisco, M.; Mendonça, T.; Marques, J.S.: Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Syst. J. **8**(3), 965–979 (2013)
11. Nezhadian, F.K.; Rashidi, S.: Melanoma skin cancer detection using color and new texture features. In: 2017 Artificial Intelligence and Signal Processing Conference (AISP), pp 1–5. IEEE (2017)
12. Ashour, A.S.; Hawas, A.R.; Guo, Y.; Wahba, M.A.: A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images. Signal Image Video Process. **12**(7), 1311–1318 (2018)
13. Abuzaghleh, O.; Barkana, B.D.; Faezipour, M.: Noninvasive real-time automated skin lesion analysis system for melanoma early detection and prevention. IEEE J. Transl. Eng. Health Med. **2015**, 1–12 (2015)
14. Mhaske, H.R.; Phalke, D.A.: Melanoma skin cancer detection and classification based on supervised and unsupervised learning. In: 2013 International conference on Circuits, Controls and Communications (CCUBE), pp. 1–5. IEEE (2013)
15. Alquran, H.; Qasmieh, I.A.; Alqudah, A.M.; Alhammouri, S.; Alawneh, E.; Abughazaleh, A.; Hasayen, F.: The melanoma skin cancer detection and classification using support vector machine. In: 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), pp 1–5. IEEE (2017)
16. Saez, A.; Sanchez-Monedero, J.; Gutiérrez, P.A.; Hervás-Martínez, C.: Machine learning methods for binary and multiclass classification of melanoma thickness from dermoscopic images. IEEE Trans. Med. Imaging **35**(4), 1036–1045 (2015)
17. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
18. Pham, T.-C.; Luong, C.-M.; Visani, M.; Hoang, V.-D.: Deep CNN and data augmentation for skin lesion classification. In: Asian Conference on Intelligent Information and Database Systems, pp. 573–582. Springer (2018).
19. DeVries,T.; Ramachandram, D.: Skin lesion classification using deep multi-scale convolutional neural networks. arXiv:1703.01402 (2017)
20. Peng, X.; Saenko, K.: Combining texture and shape cues for object recognition with minimal supervision. In: Asian Conference on Computer Vision, pp. 256–272. Springer (2016)
21. Saba, T.; Khan, M.A.; Rehman, A.; Marie-Sainte, S.L.: Region extraction and classification of skin cancer: a heterogeneous framework of deep CNN features fusion and reduction. J. Med. Syst. **43**(9), 289 (2019)
22. Rashid, M.; Khan, M.A.; Sharif, M.; Raza, M.; Sarfraz, M.M.; Afza, F.: Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and sift point features. Multimed. Tools Appl. **78**(12), 15751–15777 (2019)
23. Serte, S.; Demirel, H.: Gabor wavelet-based deep learning for skin lesion classification. Comput. Biol. Med. **113**, 103423 (2019)
24. Sorokin, A.: Lesion analysis and diagnosis with mask-rcnn. arXiv:1807.05979 (2018)
25. Taqi, A.M.; Al-Azzo, F.; Awad, A.; Milanova, M.: Skin lesion detection by android camera based on SSD-mobilenet and tensorflow object detection API. Am. J. Adv. Res. **1**, 3 (2019)
26. Roy, S.S.; Haque, A.U.; Neubert, J.: Automatic diagnosis of melanoma from dermoscopic image using real-time object detection. In: 2018 52nd annual conference on information sciences and systems (CISS), pp. 1–5. IEEE (2018)
27. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
28. Situ, N.; Yuan, X.; Chen, J.; Zouridakis, G.: Malignant melanoma detection by bag-of-features classification. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3110–3113. IEEE (2008)
29. Barata, C.; Celebi, M.E.; Marques, J.S.J.S.: A survey of feature extraction in dermoscopy image analysis of skin cancer. IEEE J. Biomed. Health Inform. **23**(3), 1096–1109 (2018)
30. Sivic, J.; Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: Null, p. 1470. IEEE (2003)
31. Sultana, N.N.; Puhan, N.B.: Recent deep learning methods for melanoma detection: a review. In: International Conference on Mathematics and Computing, pp. 118–132. Springer (2018)
32. Kavitha, J.C.; Suruliandi, A.: Texture and color feature extraction for classification of melanoma using SVM. In: 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), pp. 1–6. IEEE (2016)
33. Sotoodeh, M.; Moosavi, M.R.; Boostani, R.: A novel adaptive lbp-based descriptor for color image retrieval. Expert Syst. Appl. **127**, 342–352 (2019)
34. Zhu, C.; Bichot, C.-E.; Chen, L.: Multi-scale color local binary patterns for visual object classes recognition. In: 2010 20th International Conference on Pattern Recognition, pp. 3065–3068. IEEE (2010).
35. Singh, G.; Chhabra, I.: Effective and fast face recognition system using complementary OC-LBP and HOG feature descriptors with SVM classifier. J. Inf. Technol. Res. (JITR) **11**(1), 91–110 (2018)
36. Li, J.; Sang, N.; Gao, C.: Completed local similarity pattern for color image recognition. Neurocomputing **182**, 111–117 (2016)
37. Benco, M.; Hudec, R.; Kamencay, P.; Zachariasova, M.; Matuska, S.: An advanced approach to extraction of colour texture features based on GLCM. Int. J. Adv. Robot. Syst. **11**(7), 104 (2014)
38. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
39. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L.: Microsoft coco: common objects in context. In: European Conference on Computer Vision, pp. 740–755. Springer (2014)
40. Gutman, D.; Codella, N.C.F.; Celebi, E.; Helba, B.; Marchetti, M.; Mishra, N.; Halpern, A.: 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). arXiv:1605.01397 (2016)