

AUTOMATIC DETECTION AND CATEGORIZATION OF SKIN LESIONS FOR EARLY DIAGNOSIS OF SKIN CANCER USING YOLO-V3 - DCNN ARCHITECTURE

S. OSWALT MANOJ¹, K. RAMA ABIRAMI², AKILA VICTOR^{✉,3}, MONIKA ARYA⁴

¹Assistant Professor, Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, Tamilnadu, India; ²Associate Professor, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru; ³Associate Professor Grade 1, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India. 632 014; ⁴Assistant Professor, Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg C.G. Chhattisgarh, 491001, India

e-mail: oswaltmanojibm@gmail.com; ramaabirami1988@gmail.com; a.akilavictor@gmail.com; arya.akshara@gmail.com

(Received July 22, 2022; revised December 1, 2022; accepted January 29, 2023)

ABSTRACT

Malignant melanoma is a type of benign skin cancer that is the most lethal due to its rapid development and affects a large number of people worldwide. Also, it is one of the deadliest diseases in the world. Moreover, existing research has stated that risk factors may be significantly decreased by making it nearly curable if diagnosed early on. This prompt identification and categorization necessitate using an automated system, even though the existing method is rather difficult. Hence our research employs the YOLO v3 - DCNN architecture to discover and categorize the deadliest kinds of skin cancer. Initially, YOLO v3 generates the feature map; simultaneously, colour features are extracted using colour moments with QuadHistogram, whereas Grey Level Co-occurrence Matrix (GLCM) with Redundant Contourlet Transform(RCT) generated texture features, and both (colour and texture) features get fused. Then, fused features are fed into the Deep Convolutional Neural Network (DCNN), which classifies the different types of skin cancer. Finally, our proposed approach is compared with the current works. Consequently, our proposed YOLO-v3 –DCNN has greater accuracy when contrasted with the baseline techniques.

Keywords: skin cancer, Deep Convolutional Neural Network (DCNN), deep learning, you only look once (YOLO-v3).

INTRODUCTION

The skin is the body's largest organ, shielding it from temperature, radiation, and illness. It also aids in body temperature control, including fat and water storage. The risk of skin cancer is one of the most severe concerns (Byrd *et al.*, 2018). This disease is generated by irregular cells which affect humans. Skin cancer starts with cells, which have been the fundamental components of the skin; skin cells proliferate and split to produce new cells. Skin cells deteriorate and die every day, and new ones replace them. This process can be misleading at times. New cells develop whenever the skin does not need them, while old cells perish when they don't. Excessive cells clump together to form a mass of tissue termed a tumour (Zhang *et al.*, 2020; O'Sullivan *et al.*, 2019; Hylands, 2019). A tumour could be either cancerous or benign. It can potentially spread to other body parts if not detected early (2014).

Skin cancer has been typically categorized as melanoma and nonmelanoma (Elgamal, 2013). Melanoma seems to be a severe, uncommon, and lethal kind of skin cancer. As per the American Cancer Society, melanoma skin cancer accounts for just 1% of all occurrences, but it does have a higher mortality rate (2021). Melanoma grows from melanocytes, which are skin cells. It begins whenever healthy melanocytes multiply excessively, resulting in the formation of a malignant tumour. It can potentially affect any part of the human body adversely. It is widespread in sun-exposed areas such as the hands, face, neck, and lips. Melanocytes are already responsible for producing dark colours on the skin, hair, eyes, and other body parts. As a consequence, most melanoma tumours appeared brown or black. Nonetheless, in a small percentage of instances, melanomas do not establish colour and instead look pink, red, or purple (Hodis, 2018).

Melanoma cancers may be controlled if identified early or spread to the rest of the body, eventually killing the sufferer (Khan *et al.*, 2019). Nonmelanoma tumours have become less complicated to cure than melanoma tumours. Melanoma research (Cascinelli *et al.*, 1987) was among the first to emphasize on Computer-Aided Diagnosis (CAD) (Parameshachari *et al.*, 2020; Arun *et al.*, 2020) via applying automated strategies for the earlier identification of all skin lesions. Such systems utilize classical computer vision technology to extract attributes like shape, colour, as well as texture to supply a classification model (Ercal *et al.*, 1994; Celebi *et al.*, 2007; Wighton *et al.*, 2011; Maglogiannis & Delibasis, 2015; Barata *et al.*, 2014).

Early identification of melanoma has been seen in studies to considerably lower the mortality percentage from melanoma cancer (Razmjoooy *et al.*, 2012). The fact that earlier identification of melanoma remains challenging, even for specialists, is a significant issue. As a result, utilizing a strategy to simplify the diagnosis might be beneficial to professionals (Cohen *et al.*, 2018; Khodaei *et al.*, 2018; Kulkarni & Mukhopadhyay, 2017; Narasimhan & Elamaram, 2016). Because of the wide variety of skin lesions in dermatology, an automated skin cancer diagnosis is difficult. Various automated approaches for detecting and diagnosing skin cancer have been developed.

The traditional way of medical picture analysis involves a succession of low-level pixel processing technologies. Because of the poor distinction between the surrounding skin and the lesion region, accurate lesion segmentation is also tricky. Recent advancements in this area have shown that deep learning has been the most beneficial machine learning approach applied to the issue and utilized in a variety of applications, including speech recognition (Rashid *et al.*, 2019), pattern recognition (Bisla *et al.*, 2019), as well as bioinformatics (Farag *et al.*, 2016). Deep learning methodologies have produced excellent outcomes in several areas compared to other traditional approaches to machine learning. In recent years, numerous deep-learning methodologies are also being investigated for computer-dependent skin cancer screening. Using these strategies speeds up the diagnostic procedure but also lessens human errors. It can affect the effectiveness and quickness with which physicians and radiologists identify melanoma. Even though the CNN design produces superior results, there are significant accuracy limitations in extracting features from skin lesions, necessitating the development of a new deep learning-dependent architecture. As a consequence of identifying skin cancer at an earlier phase, our

research proposed YOLO v3 – DCNN architecture. This research can be divided into the following stages:

Initially, You Only Look Once (YOLO)-v3 architecture uses a combination of YOLO v2, Darknet-53, and Residual networks to extract features from skin lesions. Simultaneously, colour-texture characteristics were retrieved by Grey Level Co-occurrence Matrix (GLCM) with Redundant Contourlet Transform(RCT) and colour moments with QuadHistogram. Then extracted features get fused.

In addition, the fused characteristics were fed into the DCNN for categorizing skin lesions, including malignant, benign, and basal cell carcinoma (bcc).

The remaining portions of the paper were assembled as follows: Section 2 reviews feature extraction and the categorization of skin cancer. Section 3 explains the proposed YOLO v3 - DCNN methodology. Section 4 elaborates on the experimental results, comparisons, and discussions, followed by Section 5, which concludes the paper.

LITERATURE SURVEY

Much study has been undertaken in latest days employing various artificial neural networks for early diagnosis of skin malignancies, which is addressed below:

Dorj *et al.* (Dorj *et al.*, 2018) established a strategy for categorizing skin lesion photographs into four groups. A pre-trained deep CNN termed AlexNet is being used to extract features, accompanied by an error-correcting output coding SVM functioning as a classifier to satisfy the need for intelligent and rapid skin cancer categorization.

Ünver *et al.* (Ünver *et al.*, 2019) proposed an effective pipeline for segmenting skin lesions in dermoscopic images. It combines the GrabCut algorithm with the You Only Look Once deep convolutional neural network (YOLO). The approach was tested on the PH2 and ISBI 2017 datasets, two extensively used public datasets (Skin Lesion Analysis Towards Melanoma Detection Challenge Dataset). The suggested method has a 93.39% accuracy rate. Numerous medical segmentation problems can also be solved using the recommended method as a classification step for diagnosing melanoma; the authors' future work may separate skin lesion images and integrate deep convolutional neural networks.

Nida *et al.* (Nida *et al.*, 2019) suggested an innovative methodology centred on RCNN and FCM clustering enabling accurate & automated Melanoma area segmentation among dermoscopic pictures. A deep Region-de-

pendent convolutional neural network (RCNN) effectively recognizes the numerous afflicted areas, mostly in bounding boxes utilizing Fuzzy C-mean (FCM) grouping, which facilitates localization.

Albahli *et al.* (Albahli *et al.*, 2020) presented a unique method for localizing and segmenting melanoma using YOLOv4 and dynamic contour segmentation. The three stages of our suggested framework are skin augmentation, melanoma localization, and, ultimately, melanoma segmentation. The melanoma lesion is efficiently and precisely detected and segmented using our suggested method, which achieves 94% accuracy. The suggested approach can also be used for other medical image segmentation issues.

Adegun *et al.* (Adegun & Viriri, 2020) introduced a unique paradigm for automated skin cancer detection, including segmentation and categorization of skin lesions. A deep learning-dependent CAD architecture comprised of a multi-scale encoder-decoder segmentation system and an FCN-dependent DenseNet categorization network has also been introduced to recognize and categorize skin lesion photos to recognize skin cancer illnesses.

Akram *et al.* (Akram *et al.*, 2020) suggested a unique methodology for skin lesion categorization, which utilizes deep feature data to create a far more discriminant feature vector whilst preserving the initial feature space. Following fine-tuning, they incorporated the obtained data from the chosen pre-trained systems, which considerably elevated categorization accuracy. In the recommended approach, the researchers utilized less than 3% of the total characteristics, enhancing categorization accuracy by eliminating redundancies and minimizing computing time. After execution of this approach, researchers make the following claims: (a) fusion of retrieved characteristics from a collection of pre-trained systems promotes overall accuracy, while (b) the inclusion of an attribute selection plus dimensionality reduction phase enhances categorization outcomes greatly.

Khan *et al.* (Khan *et al.*, 2020) used a unique technique that identifies skin cancer and classifies it. The suggested technique is focused on saliency assessment as well as the identification of the most discriminating deep characteristics. The suggested Gaussian system enhances lesion distinction, followed by a colour space transfer from RGB to HSV. By keeping that foreground and background as distinct as possible, the new colour space aids in building saliency maps that use inner and outer discontinuous windows. The inception of CNN ar-

chitecture on two basic output tiers is being used to capture in-depth information from segmented photos. Those retrieved characteristics are then fused utilizing the suggested decision-controlled parallel fusion approach before even being chosen employing the suggested window distance-controlled entropy characteristics selection technique. The most differentiating characteristics were then put into categorization.

Khan *et al.* (Khan *et al.*, 2020) developed a novel skin cancer localization and diagnosis strategy by merging a deep learning system with an iteration-controlled Newton-Raphson (IcNR) oriented attribute selection approach. This suggested system comprises three fundamental steps: lesion localization utilizing a faster region-dependent convolutional neural network (RCNN), deep characteristic retrieval, and feature selection using the IcNR approach. There is a localization step; a novel contrasted stretching strategy centred on the bee colony method (ABC) has been applied. The augmented images and underlying ground truths were then fed into Fast-RCNN, resulting in segmented pictures. A pre-trained model, DenseNet201, has been utilized to retrieve in-depth features via transfer learning, which would then be exposed to a decision stage through the suggested IcNR approach. The most discriminating characteristics are ultimately categorized by employing layered feed-forward neural networks.

Khan *et al.* (Khan *et al.*, 2021) developed an automated computer-aided diagnostic (CAD) solution relying on the deep learning framework. The initial dermoscopic images were once pre-processed employing the decorrelation formulation strategy, and also resulting images were then supplied to the MASK-RCNN for lesion segmentation. This MASK RCNN network is again trained to utilize the segmented RGB pictures obtained from the ground truth images. Incorporating a GrabCut algorithm and an Adaptive Neuro-Fuzzy Classifier (ANFC) framework.

Sikkandar *et al.* (Yacin Sikkandar *et al.*, 2021) propose a novel segmentation-dependent categorization framework for skin lesion identification. Initially, a Top hat filter, as well as in painting techniques, were used in the pre-processing stage. The pre-processed pictures are then segmented using the Grabcut method. The feature extraction approach is then utilized in collaboration with a deep learning-dependent Inception model. Ultimately, the dermoscopic pictures were categorized using an adaptive neuro-fuzzy classifier (ANFC) network.

Saeed *et al.* (Saeed & Zeebaree, 2021) proposed an advanced skin cancer detection and classification

method based on DCNN architectural variations. Transfer learning, fine-tuning, the ensemble technique, data creation, and augmentation are all viable methods for reducing labelled data insufficiency, which is prone to overfitting. It also helps to improve the overall effectiveness of skin lesion categorization in CAD systems.

As a result, to improve the accuracy and speed of skin lesion detection, there is a need to develop novel artificial intelligence-based approaches.

AUTOMATIC DETECTION FOR EARLY DIAGNOSIS OF SKIN CANCER AND CLASSIFICATION USING YOLO V3 – DCNN ARCHITECTURE

Skin cancer is among the most lethal diseases in the globe; therefore, early identification, as well as categorization, are necessary. Thus, this paper proposed a combined architecture of YOLO version 3, and deep CNN has been presented for detecting and classifying the skin lesion. It explored a You Only Look Once (YOLO)-v3 architecture, combining YOLO v2, Darknet-53, and Residual networks to extract features from skin lesions.

The darknet's 53 layers are layered with 53 additional layers for the detection head, giving YOLO v3 a total of 106 layers of the fully convolutional underlying architecture. This layer generates the feature map with high accuracy. Simultaneously texture and colour features are extracted; texture feature extraction can be done employing a Grey Level Co-occurrence Matrix (GLCM) with Redundant Contourlet Transform(RCT) to provide accurate information about an image's texture. Colour moments are used to extract colour characteristics with QuadHistogram, wherein the quad-tree decomposition is done to the pictures, and homogeneous blocks of distinct sizes are given. Then, the concatenated features were reduced using the PCA method. The YOLO v3 and PCA features are then combined to make accurate features, which would then be fed into a Deep Convolution Neural Network (DCNN) that's been trained with the unique ground truths to increase categorization accuracy. As a consequence, our proposed technique enhances the identification as well as categorization of skin lesions. Fig 1 depicts the framework of the proposed strategy.

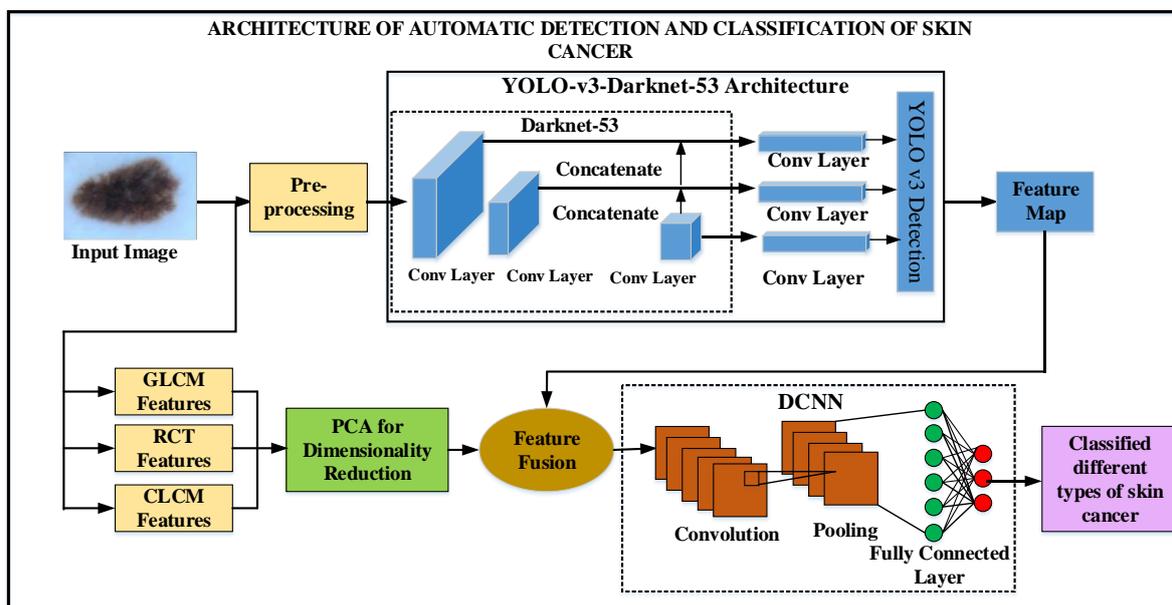


Fig. 1. Structure of the proposed approach

FEATURE EXTRACTION BY YOLO-V3

Early diagnosis of lesions is crucial for effective treatment. Early diagnosis detects neoplastic cells restricted to the epidermis during the radial growth stage.

Recently, there has been substantial interest in establishing computer-aided skin cancer diagnostics using artificial intelligence (AI). According to recent works, which we have mentioned in the literature survey, risk factors could be considerably minimized by detecting them

early on. This prompt identification and categorization necessitate using an automated system, even though the method is rather difficult. Thus our research introduces YOLOv3-DCNN architecture to increase the feature extraction and categorization accuracy. Firstly, the input data was loaded into the YOLO v3 - DCNN architecture. Features are extracted through the YOLO v3 simultaneously. The colour and texture features are extracted with Grey Level Co-occurrence Matrix (GLCM) with Redundant Contourlet Transform and colour moments with quad histogram. The obtained features then get fused.

Architecture of YOLO v3 and Darknet-106:

The proposed YOLOv3 utilizes YOLOv2, Darknet-53, and residual layers, as a novel deep neural network architecture shown in table 1.

Darknet-53: It has 53 convolutional layers, the majority of which are 3 x 3 and 1 x 1 filters, as well as the darknet's 53 layers are layered with 53 additional layers for the detection head, giving YOLO v3 a total of 106 layers fully convolutional underlying architecture to increase feature extraction accuracy. Instead of anticipating the coordinates for bounding boxes from fully connected layers, the proposed Darknet 53 uses anchor boxes. Darknet 53 helps the activations propagate through deeper layers without gradient diminishing.

YOLO v2: There are 448*448 high-resolution classifiers in our input pictures. The size of the anchor boxes in YOLO v2 is governed by the input picture shown in fig 2. It is decreased by a factor of 32. The network creates five bounding boxes containing five coordinates each: t_{zx} , t_{zy} , t_{zw} , t_{zh} , and t_{zo} . These coordinates expressed the width, height, x-coordinate, y-coordinate, and the class label of every annotation in the collection. The predictions b_{zx} , b_{zy} , b_{zw} , and b_{zh} are the predicted bounding box coordinates. The cell offsets (c_{zx} , c_{zy}) indicate the length of each cell in the picture, as well as the size of the bounding boxes (p_{zw} , p_{zh}). The predictions are determined as follows:

$$b_{zx} = \sigma(t_{zx}) + c_{zx} \quad (1)$$

$$b_{zy} = \sigma(t_{zy}) + c_{zy} \quad (2)$$

$$b_{zw} = p_{zw} e^{t_{zw}} \quad (3)$$

$$b_{zh} = p_{zh} e^{t_{zh}} \quad (4)$$

$$p_{zr}(obj) \times IOU(b_z, obj) = \sigma(t_{zo}) \quad (5)$$

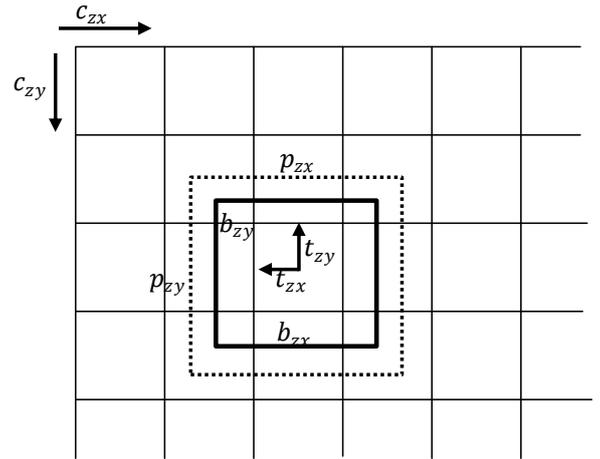


Fig. 2. Evaluation of bounding box offsets and dimensions

Where p_{zr} seems to be the conditional class probability, and IOU is the intersection of the predicted as well as ground truth bounding boxes, and $\sigma(t_{zo})$ is the change in class label. The input size is reduced by using anchor boxes.

The network's last layer comprises 1000 filters, accompanied by a final layer with a configurable count of filters. The quantity of filters for this layer has been evaluated by the count of classes discovered, as seen below:

$$N_{zfilters} = 5 \times (5 + N_{zclasses}) \quad (6)$$

For $N_{zclasses} = 2$, the count of filters in the final layer is 35

Whilst the YOLO approach extracts characteristics and detects objects using area information, the combination of feature maps containing popular texture and colour characteristics were utilized to enhance the algorithm's performance.

EXTRACTION OF TEXTURE AND COLOUR FEATURES:

The techniques for extracting texture and colour features are discussed as follows.

Gray Level Co-occurrence Matrix for Texture Feature Extraction: The GLCM with Redundant Contourlet Transform (RCT) is used to extract the image's texture features. The GLCM and RCT are widely used and provide accurate information about an image's texture. GLCM describes the texture of a picture by estimating the frequency of pairs of pixels in the picture

with defined values and a defined spatial relationship. $p_z(i_z, j_z)$ was utilized in this approach to obtain the image's pixel values at the spatial coordinates i_z & j_z . Every measurement generates a $416 * 416$ matrix that is utilized to describe each pixel in the picture.

Table 1. *Darknet-106 architecture adopted by YOLO-v3*

	Type	Filters	Size	Output
	Conv	32	3×3	224×224
	Conv	64	3×3/2	112×112
4×	Conv	32	1×1	112×112
	Conv	64	3×3	
	Re-sidual			
	Conv	128	3×3/2	56×56
6×	Conv	64	1×1	56×56
	Conv	128	3×3	
	Re-sidual			
	Conv	256	3×3/2	28×28
16×	Conv	128	1×1	28×28
	Conv	256	3×3	
	Re-sidual			
	Conv	512	3×3/2	14×14
16×	Conv	256	1×1	14×14
	Conv	512	3×3	
	Re-sidual			
	Conv	1024	3×3/2	7×7
8×	Conv	512	1×1	7×7
	Conv	1024	3×3	
	Re-sidual			
Avg-pool Con- nected Soft- max		Global 1000		

In our research, the Redundant Contourlet Transform (RCT) generates equal-size directional subband pictures. This redundancy gets produced by avoiding any downsampling procedure in the Laplacian pyramid approach; L lowpass approximations of the picture were constructed by employing L suitable low pass filters. When paired with the GLCM, RCT provides high-quality data on the whole texture of the picture. The GLCM and RCT features have a minimum level of complexity and extract higher accurate features.

Quad Histogram-based Colour Feature Extraction: The colour characteristics are then reconstructed utilizing colour consequences utilizing Quad Histogram. The images are quad-tree decomposed, and homogeneous blocks of varied sizes are provided. Aside from analyzing the global texture properties, an orthogonal combination of local binary patterns has also been proposed as a colour predictor for dermoscopic skin lesions.

The colour features extracted colour moments with the quad histogram are explained as follows: The implementation of a colour moments-based colour feature extraction approach to the three colour channels, R, G, and B, is expanded. In each of the three channels, single and multi-integrative co-occurrence matrices were also constructed. As a consequence, a feature vector of dimensions of $208*208*8$ is produced for each image. Every feature vector for one single channel is approximately $208*208$ in size. There seem to be three channels for a single co-occurrence matrix: R, G, and B. There have been three extra channels for multi-co-occurrence matrices: RG, GB, and BR. The original image's single and multi-co-occurrences are computed as two distinct channels, increasing the dimensions to 8. For proper concatenation, the feature vector is resized to $416*416*2$.

Several depths i_z of the quad-tree are thus generated, each containing range blocks of the same size. This research presents some feature histograms. Let L_z denote the greatest depth enforced in the (fractal) decomposition with quad-tree refinement, and l_z denotes the minimum depth. Ω_{l_z, L_z, k_z} the domain is defined as:

$$\Omega_{l_z, L_z, k_z} = \{(i_z, j_z) \in l_z N_z^2 \mid l_z \leq i_z \leq L_z, 1 \leq j_z \leq k_z\} \quad (7)$$

Here, i_z is the depth in the quad-tree structure and k_z is the number of feature bins selected. A function is defined as a histogram h_z on Ω_{l_z, L_z, k_z} ,

$$h_z: \Omega_{l_z, L_z, k_z} \rightarrow R_z \text{ with } h_z \geq 0 \quad (8)$$

If $(i_z, j_z) \in \Omega_{l_z, L_z, k_z}$ then $h_{zi_z, j_z} = h_z(i_z, j_z)$ is referred to as the value of h_z at (i_z, j_z) . If a histogram h_z on Ω_{l_z, L_z, k_z} It satisfies the required extra criteria and is referred to as a (weighted) quad-tree feature histogram.

$$h_{zi_z, j_z} = w_{zi_z} v_{zi_z, j_z} \quad (9)$$

Where, $\sum_{i_z=l_z}^{L_z} w_{zi_z} \leq 1$; $w_{zi_z} \geq 0$ and $v_{zi_z, j_z} \geq 0$; $\sum_{j_z=1}^{k_z} v_{zi_z, j_z} = 1, \forall i_z \in N_z$ with $l_z \leq i_z \leq L_z$

Equation 9 can be interpreted as follows: It w_{zi_z} can be interpreted to mean weighing the contents of the bins based on depth i_z . Also, each depth i_z we have k_z bins v_{zi_z, j_z} , the contents of which add up to 1. The colour moments with quad histogram colour features have a minimum level of complexity and minimum computation power. Finally, colour and texture features are obtained through GLCM with RCT and colour moments with quad histogram methods.

Principal Component Analysis: Then, the texture and colour features are merged into a single $416*416*7$

matrix and given as input to the PCA, reducing the concatenated features' dimensionality. The feature map has been generated using the PCA output of $7*7*1024$. Moreover, the feature map created by YOLO v3 has $7*7*1024$ in size. Both feature maps, including YOLO v3, colour, and texture features, are then compressed before being concatenated to establish a fused feature map. Our proposed approach extracts the features such as contrast, dis-similarity, homogeneity, energy, co-relation, and ASM features. This feature map was fed into the DCNN via the feature fusion block, which is then trained to generate the output.

CLASSIFICATION BASED ON DEEP CONVOLUTIONAL NEURAL NETWORKS (DCNN)

DCNN comprises three essential layers: the convolutional layer, the pooling layer, and the fully connected layer. Each convolution layer in DCNN has a large count of weights subsampled by the pooling layer to give output from the convolution layer whilst lowering the information ratio of the layer below. Ultimately, the outputs of the pooling layer were utilized to inject into the fully linked layers. Convolutional neuron layers, including data for various applications such as image categorization and numerous 2D matrices, are a crucial component of DCNN.

The proposed DCNN structure has an input layer, 2 convolutional layers, 2 pooling layers, and 1 max-pooling layer. The input is first loaded into the convolutional layer, which seems to have kernels that are $28*28*3$, the size of the pooling layer is $2*2$, and the learning rate is 0.00001. This approach may retrieve the regional attributes of the original picture depending on the retrieval of local features. This learning procedure's principal purpose is to build specific kernel matrices for creating more prominent features for the issue of skin cancer identification. In this case, the backpropagation (BP) model was adopted to get the lowest possible error value for the system. Sliding window-based convolution has

been employed for the network. In this research, the rectified linear unit (ReLU) would be employed as the activation function for the neurons via the function $f_z(x_z) = \max(x_z, 0)$. Max pooling can also reduce the output network size so that only the optimum levels are assessed as the sliding grid's next layer.

The BP strategy is a gradient descent-dependent system for eliminating neural network error that employs cross-entropy loss as the fitness function. This strategy may be described as follows,

$$L_z = \sum_{j_z}^{N_z} \sum_{i_z}^{M_z} -d_{z j_z}^{i_z} \log y_{z j_z}^{i_z} \quad (10)$$

Where N_z would be the sample number, $d_{j_z} = (0, \dots, 0, \underbrace{1, \dots, 1}_{k_z}, 0, \dots, 0)$ has been the needed output vector while $y_{z j_z}$, seems to be the obtained output vector of the m_z th class that may be derived utilizing the preceding formula:

$$y_{z j_z}^{i_z} = \frac{e^{f_{z j_z}}}{\sum_{i_z=1}^{M_z} e^{f_{z j_z}}} \quad (11)$$

The weight penalty is being used in the creation of the function L_z to incorporate a value to enhance the weight values:

$$L_z = \sum_{j_z}^{N_z} \sum_{i_z}^{M_z} -d_{z j_z}^{i_z} \log y_{z j_z}^{i_z} + \frac{1}{2} \eta \sum_{K_z} \sum_{L_z} \omega_{k_z}^2 \cdot l_z \quad (12)$$

Where, ω_{k_z} indicates the weight of the connection, L_z signifies the total count of layers, while K_z represents the layer l_z connections. This DCNN layer classifies the different types of skin cancers. Fig 2 illustrates the block diagram of DCNN for skin cancer detection.

Consequently, the proposed method accurately detects and classifies the different types of skin lesions. Moreover, the following section deals with the implementation results of this proposed strategy.

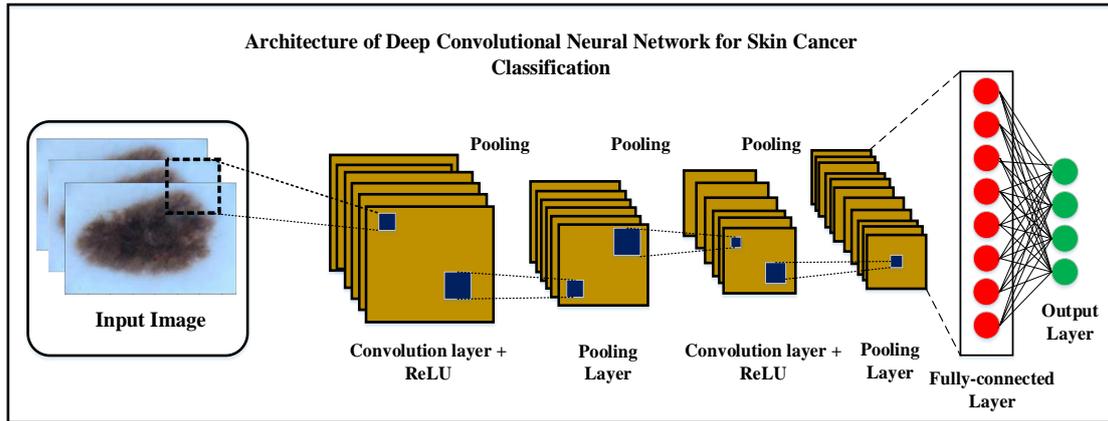
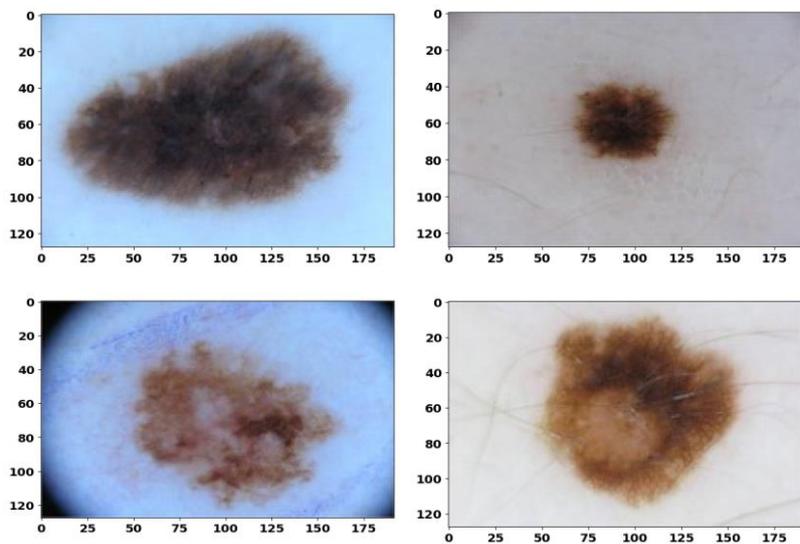
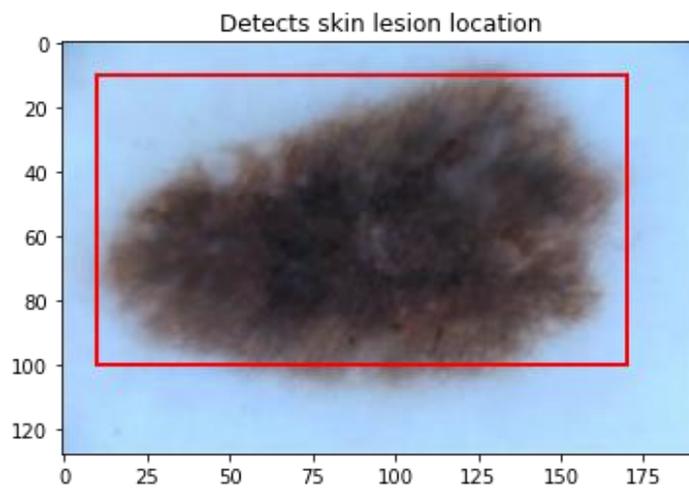


Fig. 3. Structure of DCNN in skin cancer detection



(a)



(b)

Fig. 4. Skin lesion detection from the input dataset

RESULT AND DISCUSSION

In this study, YOLO v3 and deep CNN have been presented for detecting and classifying skin lesions. The performances of our proposed methodology, as well as the implementation findings, are discussed in this section. After that, comparison results are presented, and our proposed approach is compared with the existing approach to show its effectiveness. The software for simulation is PYTHON 3, and the operating system is Windows 7 (64-bit) with 8GB of RAM.

Initially, the input images are fed into the YOLO v3-DCNN model. Then, we detect the skin lesion using YOLO v3; we extract the features using GLCM, RCT, and CLCM. Finally, we classify the skin lesions.

DATASET DESCRIPTION

In the existing research, collected dermatoscopic images are tiny and lack variety for the automatic detection of pigmented skin lesions, which makes the neural networks challenging to train. Therefore, this research utilized the dataset of the Human Against Machine with 8012 training images HAM10000 (Tschandl *et al.*, 2018). These dermatoscopic pictures are gathered from many populations, which have been recorded and preserved utilizing diverse modalities. Furthermore, the final dataset contains 10,015 dermatoscopic pictures, which have been utilized as a training and testing set for the proposed network. This dataset is split into training 70% and testing 30%. Additionally, the lesion id column

in the HAM10000 metadata file might well be utilized to track lesions with many pictures in the dataset.

Table 2. *Details of Dataset*

Dataset Name	Total number of images	Total number of classes	Number of images in the training set	Number of images in the test set
HAM10000	10,015	7	8012	2003

SKIN LESION DETECTION FROM THE INPUT DATASET

The sample input image from the MNIST HAM10000 dataset is shown in fig 4(a). The skin lesion is detected from the input image using our proposed YOLO-v3 and DarkNet-106 approach. Fig 4(b) illustrates the Region of Interest (ROI) used to filter the image. Based on the dimension, this research accurately detects the skin lesion area. Images that belong to the ROI value are set to 1, and ROI is set to 0 for the outside images. The skin lesion is then segmented from the generated region, shown in fig 5. The segmentation process is performed to extract the affected area.

Table 3. *Class distributions of the following 7 classes for the pigmented lesions: melanocytic nevus (nv), melanoma (mel), basal cell carcinoma (bcc), dermatofibroma (df), benign keratosis (bkl), vascular lesion (vasc), and actinic keratosis (akiec)*

Dataset	nv (%)	mel (%)	bcc (%)	df (%)	bkl (%)	Vasc (%)	akiec (%)	Total (%)
HAM10000	6705 (67)	1113 (11)	514 (5.1)	115 (1.1)	1099 (11)	142 (1.4)	327 (3.3)	10015

Fig 6 (a) and (b) indicate the results of skin lesion detection. The proposed YOLO-v3 and DarkNet-106 approach efficiently detect the skin lesion by generating an anchor box. The anchor box indicates the region of the lesions for which a high degree of accuracy can be predicted. Also, Darknet 106 assists the activations propagating through deeper layers without gradient diminishing. Then, we extract the colour and texture fea-

tures as bcc, bkl, Mel, akiec, df, and vasc. Then the extracted features get fused. Then, the extracted features are classified using the proposed DCNN approach, shown in the following fig. Furthermore, our results indicate that our proposed approach can detect skin lesions earlier and more accurately.

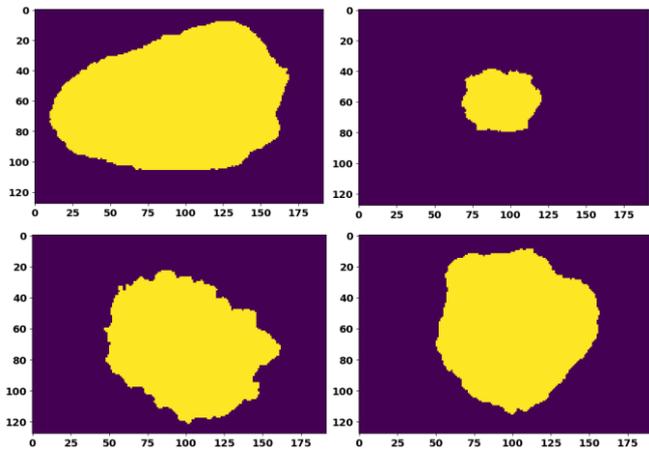


Fig 5. Segmentation of skin cancer from the filtered image

SKIN LESION DETECTION AND FEATURE EXTRACTION

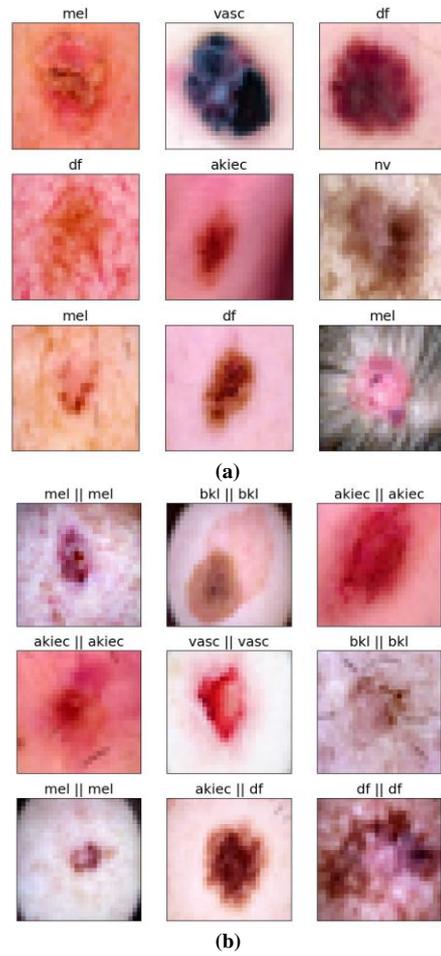


Fig. 6. Skin lesion detection

SKIN LESION CLASSIFICATION

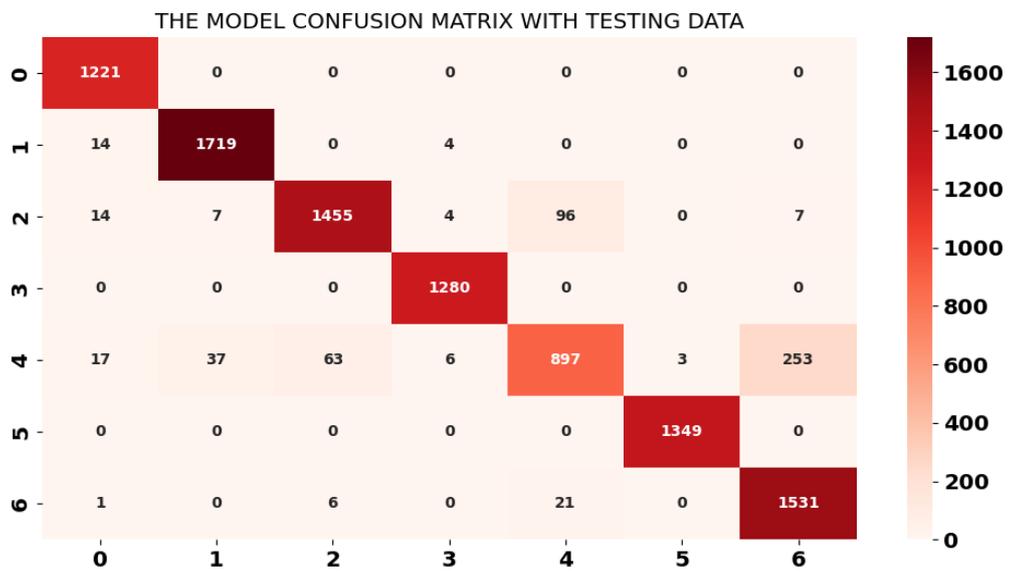


Fig. 7. Confusion Matrix

Fig 7 displays the skin lesion recognition confusion matrix. The proposed model produces the confusion matrix. The diagonal values indicate the count of tuples adequately identified by the models, and the off-diagonal values denote the count of tuples erroneously categorized by the models. The better the performance, the higher the diagonal value. In fig 7, the value 0, 1, 2, 3, 4, 5, and 6 indicate skin lesions as akiec, bcc, bkl, df, Mel, nv, and vasc. The classification results, such as accuracy and precision, are described in the performance parameters. Furthermore, our results indicate that our proposed approach can detect skin lesions earlier and more accurately.

PERFORMANCE PARAMETERS

This section describes the performance of our proposed technique; various parameters, such as accuracy

and precision, were employed to assess the unique technique's effectiveness in identifying and categorizing skin lesions.

Accuracy

The most basic intuitive performance metric is accuracy, which seems to be the ratio of precisely predicted observations to all observations. The accuracy is formulated as follows (13)

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (13)$$

Where TP- True Positive, TN-True Negative, FP- False Positive, FN-False Negative

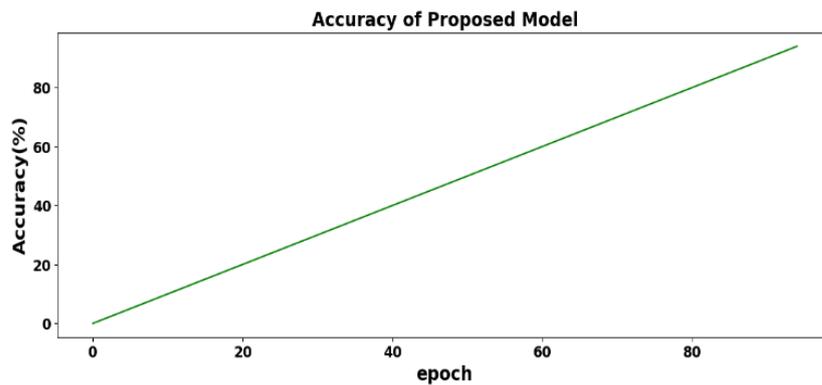


Fig. 8. Accuracy of the proposed model

Fig 8 depicts the accuracy of our proposed skin lesion detection approach. Our proposed architecture's accuracy improved when YOLO-V3 characteristics were integrated with GLCM, RCT, and CLCM features. The classification error value of the proposed approach is 4.6% which provides an accuracy of $94.572 \approx 95\%$.

Precision

Precision has been described as the ratio of correctly predicted positive observations to total predicted positive observations. The precision is calculated by equation (14),

$$\text{Precision} = \frac{TP}{TP+FP} \quad (14)$$

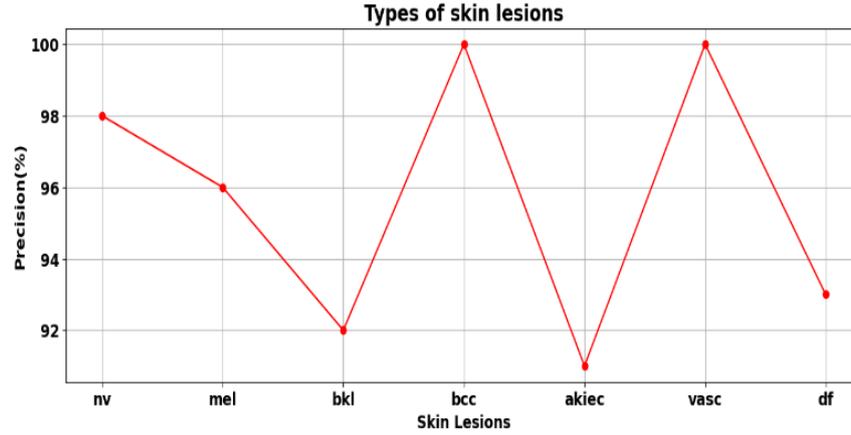


Fig. 9. Precision of the proposed model

The precision of the proposed skin lesion categorization approach is shown in fig 9. The DCNN technique improves the precision of the proposed architecture. The precision of our suggested technique attains the highest value of 100% and the lowest value of 91%, which describes the effectiveness of the classification approach.

Sensitivity and Specificity

Sensitivity measures how well a classifier model can recognize positive instances. It is often referred to as the true positive rate (TPR) or recall.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (15)$$

Specificity indicates the percentage of true negatives that the model accurately detects.

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (16)$$

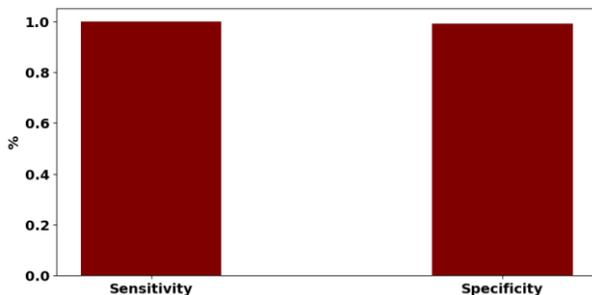


Fig. 10. Sensitivity and Specificity of the proposed approach

Fig 10 depicts the sensitivity and specificity of the proposed YOLO v3-DCNN-based approach. The obtained sensitivity value is 98.6%, and the specificity is 94%.

Training accuracy and validation accuracy of the proposed model

Fig 11 depicts the training and validation accuracy of our proposed approach. In fig 10, green colour graphical lines illustrate the training accuracy, whereas brown colour graphical lines illustrate validation accuracy.

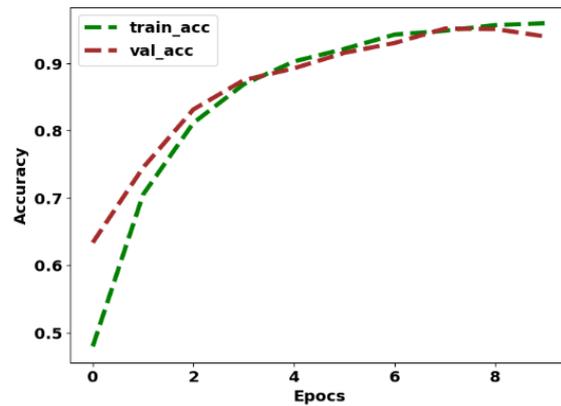


Fig. 11. The proposed model's training, as well as validation accuracy

The accuracy of training and validation is increasing over time. Our proposed model reaches 0.97 training accuracy and 0.975 validation accuracy over 9 epochs. Furthermore, the training accuracy exceeds the validation accuracy.

The proposed model's training, as well as validation losses

Fig 12 depicts our proposed model's training and validation losses. In fig 11, orange colour graphical lines illustrate the training loss, whereas violet colour graphical lines illustrate validation loss.

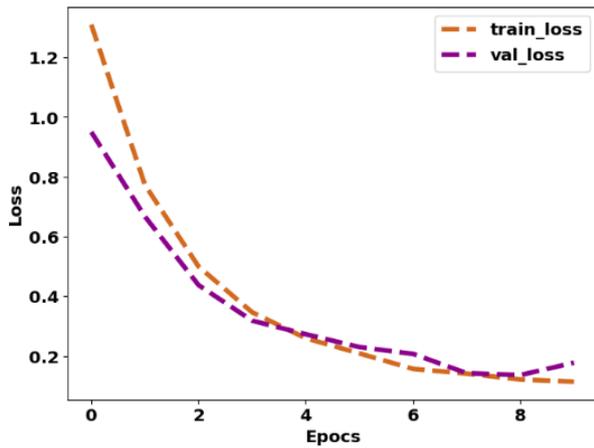


Fig. 12. The proposed model's training, as well as validation losses

Training, as well as validation loss, gradually reduces with time. Fig 12 shows that the proposed model obtained less than 0.1 train loss in 9 epochs. In 9.5 epochs, the proposed model obtained less than 0.1 validation loss. Furthermore, the training loss is less severe than the validation loss.

Comparison Results

This section explains the proposed technique's comparative outcomes, in which our novel technique is compared to the baseline approach, such as Grey Level Co-occurrence Matrix (GLCM) with Gabor and Color Level Co-occurrence Matrix (CLCM) (Nersisson *et al.*, 2021), Grey Level Co-occurrence Matrix (GLCM) with Gabor features (Nersisson *et al.*, 2021), Color Level Co-occurrence Matrix (CLCM) with Gabor features (Nersisson *et al.*, 2021), and Gabor with Color Level Co-occurrence Matrix (CLCM) (Nersisson *et al.*, 2021).

Table 4. Overall Accuracy

Methods	Accuracy (%)
GLCM + Gabor + CLCM [38]	94
GLCM + Gabor [38]	90
CLCM + Gabor [38]	91
Gabor + CLCM [38]	88
GLCM+RCT+CLCM (Proposed)	95

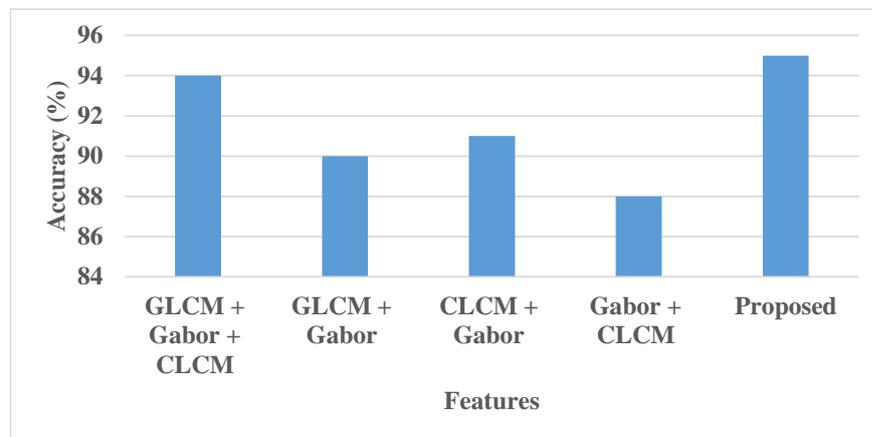


Fig. 13. Accuracy Comparison

Fig 13 and Table 4 illustrate the overall comparison of the feature fusion accuracy. The proposed technique attains higher accuracy by incorporating Grey Level Co-occurrence Matrix (GLCM) with Redundant Contourlet Transform(RCT) and Color moments with QuadHistogram. Our proposed approach compared with the baseline GLCM + Gabor + CLCM (Nersisson *et al.*, 2021), GLCM + Gabor (Nersisson *et al.*, 2021), CLCM + Gabor, and Gabor + CLCM (Nersisson *et al.*, 2021), such as 94%, 90%, 91%, and 88%. Consequently, our proposed approach has a superior accuracy of 95% to existing methodologies.

Table 5. Overall Precision

Methods	Precision (%)
GLCM + Gabor + CLCM [38]	92
GLCM + Gabor [38]	87
CLCM + Gabor	86
Gabor + CLCM	85
GLCM+RCT+CLCM (Proposed)	94

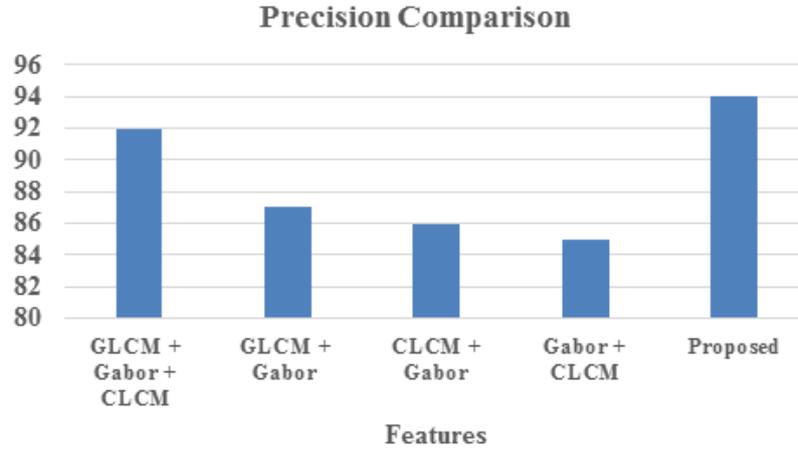


Fig. 14. Precision Comparison

Fig 14 and table 5 depict the precision for the overall feature fusion comparison. The proposed technique attains a higher precision by incorporating Grey Level Co-occurrence Matrix (GLCM) with Redundant Contourlet Transform(RCT) and Color moments with QuadHistogram. Our proposed approach compared with the baseline GLCM + Gabor + CLCM (Nersisson *et al.*,

2021), GLCM + Gabor (Nersisson *et al.*, 2021), CLCM + Gabor, and Gabor + CLCM (Nersisson *et al.*, 2021), such as 92%, 87%, 86%, and 85%. Consequently, our proposed technique has a greater accuracy of 94 % than existing techniques.

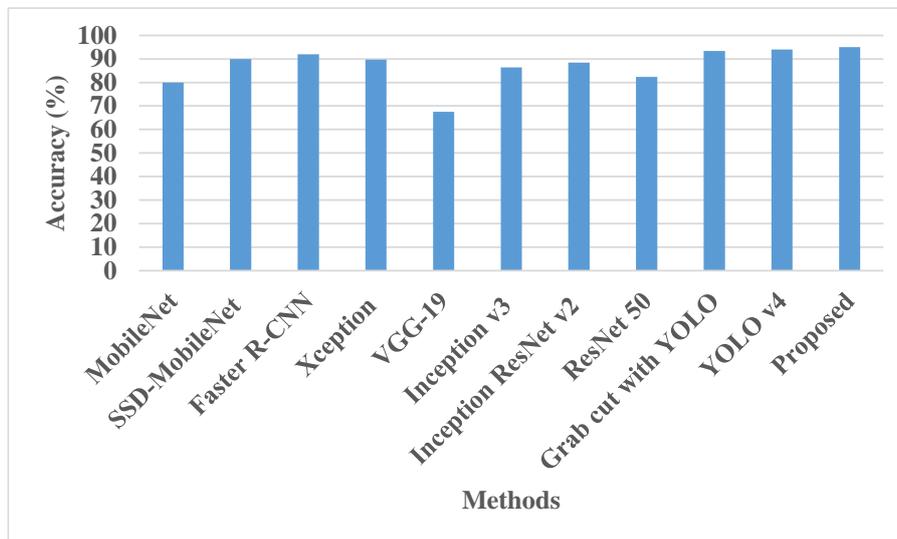


Fig. 15. Model Comparison

Fig 15 depicts an overall comparison of model accuracy. The proposed model attains higher accuracy by incorporating YOLO-v3 - DCNN. Our proposed approach compared with the baseline MobileNet (38), SSD-MobileNet (Nersisson *et al.*, 2021), Faster R-CNN (Nersisson *et al.*, 2021), Xception (Jain *et al.*, 2021), VGG-19 (Jain *et al.*, 2021), Inception v3 (Jain

et al., 2021), inception ResNet v2 (Jain *et al.*, 2021), ResNet 50 (Jain *et al.*, 2021), Grabcut algorithm with YOLO (Ünver *et al.*, 2019), and YOLO v4 (Albahli *et al.*, 2020) such as 80%, 90%, 92%, 89.66%, 67.54%, 86.4%, 88.4%, 82.32%, 93.39%, and 94%. Thus our proposed technique has obtained an accuracy of 95%, which is higher than the existing techniques.

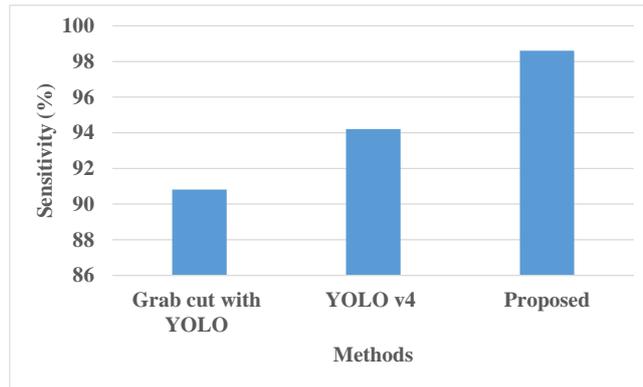


Fig. 16. Sensitivity Comparison

Fig 16 depicts an overall comparison of model sensitivity. The proposed model attains higher sensitivity by incorporating YOLO-v3 - DCNN. Our proposed approach compared with the baseline Grab cut algorithm with YOLO (Ünver *et al.*, 2019) and YOLO v4 (Albahli *et al.*, 2020), such as 90.82% and 94.2%. Thus, our proposed technique has a sensitivity of 98.6% higher than the existing techniques.

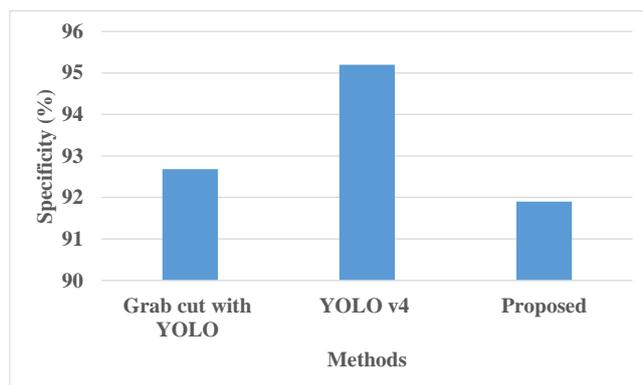


Fig. 17. Specificity Comparison

Fig 17 depicts an overall comparison of model specificity. The proposed model attains higher specificity by incorporating YOLO-v3 - DCNN. Our proposed approach compared with the baseline Grab cut algorithm with YOLO (Ünver *et al.*, 2019) and YOLO v4 (Albahli *et al.*, 2020), such as 92.68% and 95.2%. Thus, our proposed technique has a specificity of 91.9% higher than the existing techniques.

Fig 18 depicts an overall comparison of the dataset in terms of accuracy. The proposed approach used the HAM10000 dataset to classify skin cancer. The proposed YOLO-v3 – DCNN approach attains higher accuracy of 95% using the HAM10000 dataset. Moreover, the baseline datasets are ISBI 2016 (Yu *et al.*, 2018), ISIC Archive repository (Rokhana *et al.*, 2020), PH2 (Hu *et al.*, 2019), and Dermofit and MEDNODE (Mukherjee *et al.*, 2019) such as 86.81 %, 84.76%,

91.6%, and 90.58%. Thus our proposed technique has obtained an accuracy of 95% for the HAM10000 dataset, which is higher than the existing datasets.

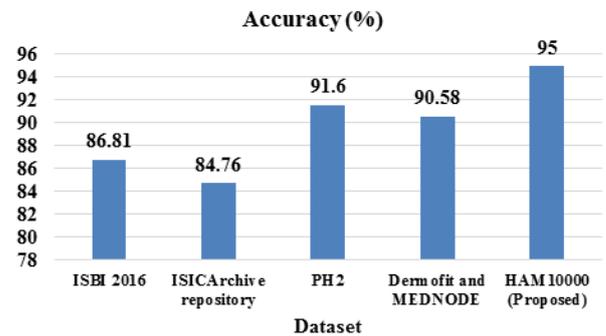


Fig 18. Dataset Comparison

CONCLUSIONS

Early disease diagnosis is essential for treatment planning in the healthcare and medical sectors. Treatment and further mitigation are essential for survival if illnesses are diagnosed early. The classification of surface-level skin cancers and diseases is evolving into a significant area of research due to the increase of several skin problems carried on by pollution and other factors. This research proposes a combination architecture of YOLO v3 and deep CNN to detect and classify skin cancer. It explored a You Only Look Once (YOLO)-v3 architecture; this layer generates the feature map with high accuracy. Simultaneously texture and colour features are extracted in texture feature extraction can be done by employing a Grey Level Co-occurrence Matrix (GLCM) with Redundant Contourlet Transform (RCT) to provide accurate information about an image's texture and Color moments. Then the extracted features are fed into a Deep Convolution Neural Network (DCNN), increasing categorization accuracy. Thus our proposed YOLO v3 - DCNN technique has obtained an accuracy of 95% by using the HAM10000 dataset, which is higher than the existing approaches.

REFERENCES

- "Defining cancer". National cancer institute (2014). Archived from the original on 25 June 2014. Retrieved 10 June 2014.
- Adegun AA & Viriri S (2020). FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images. *IEEE Access* 8:150377-96.

- Akram T, Lodhi HMJ, Naqvi SR, Naeem S, Alhaisoni M, Ali M & Qadri NN (2020). A multilevel features selection framework for skin lesion classification. *Human-centric Computing and Information Sciences* 10(1):1-26.
- Arun M, Baraneetharan E, Kanchana A & Prabu S (2020). Detection and monitoring of the asymptotic COVID-19 patients using IoT devices and sensors. *International Journal of Pervasive Computing and Communications*.
- Albahli S, Nida N, Irtaza A, Yousaf MH, Mahmood MT (2020). Melanoma lesion detection and segmentation using YOLOv4- DarkNet and active contour. *IEEE ACCESS* 8:198403–14.
- Barata C, Celebi ME & Marques JS (2014). Improving dermoscopy image classification using color constancy. *IEEE J Biomed Health* 19(3):1146-52.
- Bisla D, Choromanska A, Berman RS, Stein JA & Polsky D (2019). Towards automated melanoma detection with deep learning: Data purification and augmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* 0-0.
- Byrd AL, Belkaid Y & Segre JA (2018). The human skin microbiome. *Nat Rev Microbiol* 16(3):143-55.
- Cascinelli N, Ferrario M, Tonelli T & Leo E (1987). A possible new tool for clinical diagnosis of melanoma: the computer. *J Am Acad Dermatol* 16(2):361-7.
- Celebi ME, Kingravi HA, Uddin B, Iyatomi H, Aslandogan YA, Stoecker WV & Moss RH (2007). A methodological approach to the classification of dermoscopy images. *Comput Med Imag Grap* 31(6):362-73.
- Cohen VM, Pavlidou E, DaCosta J, Arora AK, Szyszko T, Sagoo MS & Szlosarek P (2018). Staging uveal melanoma with whole-body positron-emission tomography/computed tomography and abdominal ultrasound: Low incidence of metastatic disease, high incidence of second primary cancers. *Middle East African journal of ophthalmology* 25(2):91.
- Dorj UO, Lee KK, Choi JY & Lee M (2018). The skin cancer classification using deep convolutional neural network. *Multimed Tools Appl* 77(8):9909-24.
- Elgamal M (2013). Automatic skin cancer images classification. *IJACSA International Journal of Advanced Computer Science and Applications* 4(3):287-94.
- Ercal F, Chawla A, Stoecker WV, Lee HC & Moss RH (1994). Neural network diagnosis of malignant melanoma from color images. *IEEE T Bio-Med Eng* 41(9):837-45.
- Farag A, Lu L, Roth HR, Liu J, Turkbey E & Summers RM (2016). A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling. *IEEE T Image Process* 26(1):386-99.
- Hodis E (2018). *The Somatic Genetics of Human Melanoma* (Doctoral dissertation).
- Hylands P (2019). *Skin cancer: types, diagnosis and prevention*. *Heart Fail*. 10:00.
- Hu K, Niu X, Liu S, Zhang Y, Cao C, Xiao F, Yang W, Gao X (2019). Classification of melanoma based on feature similarity measurement for codebook learning in the bag-of-features model. *Biomed Signal Proces* 51:200–9.
- Jain, Satin, Udit Singhanian, Balakrushna Tripathy, Emad Abouel Nasr, Mohamed K. Aboudaif and Ali K. Kamrani (2021). *Deep Learning-Based Transfer Learning for Classification of Skin Cancer*. *Sensors* 21(23):8142.
- Key Statistics for Melanoma Skin Cancer. *Am. Cancer Soc*. Available online: <https://www.cancer.org/content/dam/CRC/PDF/Public/8823.00.pdf> (accessed on 8 February 2021).
- Khan MA, Akram T, Sharif M, Javed K, Rashid M & Bukhari SAC (2020). An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. *Neural Comput Appl* 32(20):15929-48.
- Khan MA, Akram T, Zhang YD & Sharif M (2021). Attributes-based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. *Pattern Recogn Lett* 143:58-66.
- Khan MA, Sharif M, Akram T, Bukhari SAC & Nayak RS (2020). Developed Newton-Raphson-based deep features selection framework for skin lesion recognition. *Pattern Recogn Lett* 129:293-303.
- Khan MQ, Hussain A, Rehman SU, Khan U, Maqsood M, Mehmood K & Khan MA (2019). Classification of melanoma and nevus in digital images for diagnosis of skin cancer. *IEEE ACCESS* 7:90132-44.
- Khodaei H, Hajiali M, Darvishan A, Sepehr M & Ghadimi N (2018). Fuzzy-based heat and power hub models for cost-emission operation of an industrial consumer using compromise programming. *Appl Therm Eng* 137:395-405.
- Kulkarni A & Mukhopadhyay D (2017). SVM classifier-based melanoma image classification. *Research Journal of Pharmacy and Technology* 10(12):4391-2.
- Mukherjee S, Adhikari A, Roy M (2019). Malignant melanoma classification using a cross-platform dataset with deep learning CNN architecture. In *Recent Trends in Signal and Image Processing*; Springer: Berlin/Heidelberg, Germany 31–41.
- Maglogiannis I & Delibasis KK (2015). Enhancing classification accuracy utilizing globules and dots features in digital dermoscopy. *Comput Meth Prog Bio* 118(2):124-33.

- Narasimhan K & Elamaran V (2016). Wavelet-based energy features for diagnosis of melanoma from dermoscopic images. *International Journal of Biomedical Engineering and Technology* 20(3):243-52.
- Nersisson R, Iyer TJ, Joseph Raj AN & Rajangam V (2021). A Dermoscopic Skin Lesion Classification Technique Using YOLO-CNN and Traditional Feature Model. *Arab J Sci Eng* 46(10):9797-808.
- Nida N, Irtaza A, Javed A, Yousaf MH & Mahmood MT (2019). Melanoma lesion detection and segmentation using deep region-based convolutional neural network and fuzzy C-means clustering. *Int J Med Inform* 124:37-48.
- O'Sullivan DE, Brenner DR, Demers PA, Villeneuve PJ, Friedenreich CM, King WD & ComPARE Study Group (2019). Indoor tanning and skin cancer in Canada: A meta-analysis and attributable burden estimation. *Cancer Epidemiol* 59:1-7.
- Parameshachari BD, Panduranga HT & liberata Ullo S (2020, September). Analysis and computation of encryption technique to enhance security of medical images. *IOP Conf Ser-Mat Sci* 925(1):012028, IOP Publishing.
- Rokhana R, Herulambang W, Indraswari R (2020). Deep convolutional neural network for melanoma image classification. In *Proceedings of the 2020 International Electronics Symposium (IES), Marrakech, Morocco, 24-26 March 2020* 481-6.
- Rashid H, Tanveer MA & Khan HA (2019, July). Skin lesion classification using GAN-based data augmentation. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 916-9. IEEE.
- Razmjooy N, Mousavi BS & Soleymani F (2012). A real-time mathematical computer method for potato inspection using machine vision. *Comput Math Appl* 63(1):268-79.
- Saeed J & Zeebaree S (2021). Skin lesion classification based on deep convolutional neural networks architectures. *Journal of Applied Science and Technology Trends* 2(01):41-51.
- Tschandl P, Rosendahl C & Kittler H (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data* 5(1):1-9.
- Ünver HM, Ayan E (2019). Skin lesion segmentation in dermoscopic images with combination of YOLO and GrabCut algorithm. *Diagnostics* 9:72.
- Wighton P, Lee TK, Lui H, McLean DI & Atkins MS (2011). Generalizing common tasks in automated skin lesion diagnosis. *IEEE T Inf Technol B* 15(4):622-29.
- Yu Z, Jiang X, Zhou F, Qin J, Ni D, Chen S, Lei B, Wang T (2018). Melanoma recognition in dermoscopy images via aggregated deep convolutional features. *IEEE T Bio-Med Eng* 66:1006-16.
- Yacin Sikkandar M, Alrasheadi BA, Prakash NB, Hemalakshmi GR, Mohanarathinam A & Shankar K (2021). Deep learning based an automated skin lesion segmentation and intelligent classification model. *J Amb Intel Hum Comp* 12(3):3245-55.
- Zhang N, Cai YX, Wang YY, Tian YT, Wang XL & Badami B (2020). Skin cancer diagnosis based on optimized convolutional neural network. *Artif Intell Med* 102:101756.