



# Deep learning in skin lesion analysis for malignant melanoma cancer identification

M. Senthil Sivakumar<sup>1</sup> · L. Megalan Leo<sup>2</sup> · T. Gurumekala<sup>3</sup> · V. Sindhu<sup>1</sup> · A. Saraswathi Priyadharshini<sup>4</sup>

Received: 11 February 2023 / Revised: 5 May 2023 / Accepted: 4 July 2023 /  
Published online: 15 July 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

The higher rate of skin diseases caused by infections, allergies, lifestyle changes, increased use of chemicals, unhealthy food habits, and hereditary reactions, among other factors demand effective diagnostic mechanisms. These mechanisms should facilitate and expedite the precise diagnosis of malignant melanoma. In addition, the existing technology-assisted tests require elaborate procedures to accurately diagnose the skin disease to be treated. The deep learning-enhanced inspection involves neural networks that diagnose skin diseases by analyzing data for skin lesions such as rashes, boils, and cancer growth. This paper proposes an automated high-precision diagnostic model with the web application to identify Malignant Melanoma Cancer. The diagnostic model framework employs a convolutional neural network (CNN) with ResNet50 for data collection, preprocessing, segmentation, enhancement, feature extraction, classification, and malignant melanoma identification to improve accuracy. The proposed model eliminates unwanted noise, enhances spectral image information, and improves classification accuracy through preprocessing and mixed hybrid pooling phases. Performance analysis shows that the proposed Malignant Melanoma Cancer detection model achieves the highest accuracy of 94% and an F1-score of 93.9%. Results from a training and testing dataset of images from the International Skin Image Collaboration (ISIC) demonstrate a significant improvement over traditional methods. Further, the web application developed for this model accelerates the precise diagnosis of Malignant Melanoma Cancer with accurate information compared to other methods in practice. The potential of applying the proposed skin cancer classification model automates the malignant and benign image classification process, speeds up diagnosis and treatment, and reduces the risk of misdiagnosis.

**Keywords** Artificial neural network · Convolutional neural network · Machine learning · ResNet50 · Benign · Malignant

## 1 Introduction

Over the last few decades, lifestyle and dietary changes have significantly impacted human well-being, making them susceptible to numerous internal and external diseases and causing a lack of immunity. Such unhealthy conditions affect every part of the human body, depending on weakness. However, the skin protects the human body from fungal and harmful bacterial infections. Therefore, skin diseases should be attended to and treated at an early stage.

Skin diseases can develop due to hereditary factors, employment, lack of nutrition, daily habits, and exposure to chemicals, among other factors. Skin cancer is one such disease that destroys tissues and spreads uncontrollably in the body. Malignant cells formed in skin tissues are identified as skin cancer. Early diagnosis and identification of malignant cells can help prevent irreversible damage [7]. Digital image processing significantly contributes to the field of health care in diagnosing malignant cells early and precisely, helping in the treatment of such conditions to avoid life-threatening damage. Digital image processing uses spectral information to represent images and categorize discrete pixels. The widely used image classification methods are unsupervised, supervised, and object-based investigations.

The unsupervised image classification initially clusters the spectral classes based on numerical information of images and matches them accordingly. In contrast, supervised image classification requires a labeled sample of training image collection to generate pixel-level details of images for classification using algorithms. The most common method of supervised image recognition is deep learning, where the features of the images are extracted and stored using neural networks [5, 8]. A Convolutional Neural Network (CNN) is a form of deep neural network that functions with convolutions specifically for image-related tasks. A CNN-based automatic classification model was developed to classify medical images of the DermNet database to diagnose conditions such as acne, keratosis, eczema herpeticum, and urticaria. Detecting variations in skin tones, locating diseases, and determining diagnoses are the main challenges of automatic classification. An up-to-date survey of all investigations on framing efficient medical image classification models that can automatically and accurately identify melanoma from skin lesion images is presented in [4, 14].

In general, the accuracy of CNN models improves with an increased size of training image collection. Image classification functions in an object-based method by utilizing both spectral and spatial information. The image pixels are classified based on their spectral properties, shape, color, and spatial relationship with neighboring pixels in this phase. Thus, the study proposes a CNN with Resnet50-based image classification procedure to detect and diagnose skin cancer at an early stage by processing diseased skin images using machine learning methods for higher precision and rapid diagnosis.

The main objective of this work is to develop a high-precision and low-loss Malignant Melanoma Identification model to classify skin cancer and to produce a web-based tool that assists specialists and technicians to predict whether a given skin lesion is benign or malignant. Early skin lesion detection allows the doctor to detect the critical stages of a patient's condition and start treatment as early as possible in order to control risk factors. Implementing this system allows for foreknowledge of skin samples, resulting in a more effective and immediate diagnosis. This paper introduces a novel approach using CNN that offers a better solution with improved accuracy and reduced loss to

skin disease diagnosis. Web application developed in this proposed model simplifies the diagnostic procedure for the user.

## 1.1 Related works

Advanced digital image processing, machine learning techniques, and deep learning models are helping to analyze medical images, especially in identifying skin diseases such as acne, keratosis, eczema herpeticum, urticaria, benign (noncancerous), and malignant (cancerous). Among the diseases, the malignant tumor grows uncontrollably faster and destroys body tissues. Thus, the presence of malignant melanoma in cells has a detrimental effect on other organs [13]. Skin cancer is currently the most prevalent malignancy in the human body. Numerous research works have been developed recently to analyze skin lesions using computer vision.

Blue-white structure (BWS) helps in detecting skin cancer cells by diagnosing malignant Melanoma from extracted features of the labeled datasets [21]. To diagnose skin cancer, the classification system follows the most significant dermoscopic criteria for BWS detection. In [1], a fully convolutional network (FCN)-based skin disease detection system was proposed, which uses multi-scale encoder-decoder segmentation, conditional random field (CRF), and FCN-based DenseNet in classification to produce contour refinement, lesion boundaries in localization, and improved computing efficiency. An efficient classification algorithm Error-Correcting Output Codes derived Support Vector Machine (SVM) algorithm [12], was attempted to classify skin cancer images of actinic keratoses, basal cell carcinoma, squamous cell carcinoma, and melanoma with the help of a pre-trained AlexNet model. In addition, an intelligent Region of Interest (ROI) based system is proposed in [3] for identifying melanoma and discerning life-threatening nevus cancer. The use of an improved k-mean clustering algorithm to extract ROIs from the images of Melanoma cells classifies the cancer image from normal skin diseases precisely.

A study was conducted to identify and predict the variables of Non-melanoma skin cancer using a decision tree (DT) based Cox (proportional hazards) regression analysis [20]. A threshold and edge-based classification were introduced for image classification by apprising the confidence level and object contour estimation [9]. However, it was found that the DT classification models derived randomly during model building did not validate independently and did not show statistically significant differences between overlapping data sets. Also, it failed to classify the malignant lesion region of skin images from smooth other surrounding regions. Another approach was introduced in [11], which was a k-nearest neighbor (KNN) approach that efficiently classified malignant skin lesions from normal skin images with highly accurate test results. However, it was noted that the KNN-based classification approach did not provide stable test results for large datasets due to sensitivity to noise. Also the proposed model requires an improvement to guarantee even higher accuracy and faster results since the model was tested only with limited number of images [6].

An artificial neural network (ANN) was utilized in [17] to classify melanoma skin cancer using hybrid texture features and a gradient difference-based pre-processing technique to improve accuracy. In order to improve the accuracy of skin lesion prediction, pre-processing of dermoscopic samples was performed using gradient intensity and thresholding algorithms to process the skin lesion. However, it was noted that ANN generally does not provide optimal results in classification due to the trial and error experience in architecture development. A deep neural network (DNN) with TensorFlow implementation was

presented in [15] to improve the accuracy of skin lesion classification. Deep residual learning of DNN for image recognition composed of multiple layers of interconnected neurons that extracts the features precisely, leading to improve accuracy. The paper also discussed the current deep learning models and Residual Network (ResNet) models for image classification in detail. ResNet is a powerful and popular architecture for deep learning models in computer vision, and it has been the basis for many subsequent improvements and variations in deep learning research.

Convolutional Neural Network (CNN) is one of the most popular architectures used for image classification. An analysis of CNN for image classification has been presented in [22]. The study [10] focuses on analyzing the performance of three popular networks: AlexNet, GoogLeNet, and ResNet50. AlexNet is one of the first deep convolutional neural networks to achieve state-of-the-art performance on the ImageNet dataset. It consists of eight layers, including five convolutional layers and three fully connected layers, and introduced several innovations such as the use of rectified linear units (ReLU) as activation functions, overlapping pooling, and data augmentation techniques. GoogLeNet, also known as Inception-v1 allows the network to learn both local and global features at multiple scales. The GoogLeNet architecture consists of 22 layers and was designed to be computationally efficient, allowing for faster training and inference times. ResNet50, introduced in 2015, is a deep residual neural network architecture that has been shown to outperform other architectures on a variety of computer vision tasks. It introduced the concept of residual blocks, which allow the network to learn residual functions instead of attempting to learn the original mapping directly. This architecture consists of 50 layers, including residual blocks, and has been widely adopted in many computer vision applications.

An automatic classification model based on CNN was developed to diagnose various skin conditions such as acne, keratosis, eczema herpeticum, and urticaria using medical images from the DermNet database [19]. The model is designed to learn complex and heterogeneous lesion features that encode the abrasive appearance of cancer cells. The use of dense blocks in the model helps to optimize features and identify image lesions with reduced complexity and improved accuracy. However, the complex architecture of deep learning algorithms presents challenges due to the large number of parameter tunings required, identification difficulties due to the abrasive appearance of medical images, and the difficulty of training large datasets to improve accuracy. Detecting variations in skin tones, locating diseases, and determining diagnoses are also major challenges in automatic classification.

In [18], a CNN architecture using VGGNet was introduced for lung CT image analysis. This architecture accurately identified CT images with damaged and irregular lesions as true positive samples while correctly identifying unaffected samples as true negatives. The hyper-parameters of CT image samples were extracted and analyzed to observe precision, accuracy, loss, sensitivity, recall, and specificity, and the results were compared for performance analysis. The analysis revealed that the VGGNet-based CNN architecture had lower accuracy than the ResNet-based CNN architecture in medical image classification.

In [2], a recent survey of investigations on efficient medical image classification models capable of accurately identifying melanoma from skin lesion images is presented. The challenges in identifying melanoma cells are discussed, including various pre-processing and segmentation techniques involved in melanoma cell detection. Moreover, the article also discusses the use of deep learning models for classifying skin lesions into different categories such as malignant and benign, the performance parameters used for comparative analysis, exploration and identification of state-of-the-art techniques, and the challenges of using large datasets. Table 1 presents a summary of existing skin cancer classification

**Table 1** Summary of existing skin cancer classification algorithms

Architecture	Disease	No. of Images	Size	Kernel Size	Demerit
DT [9]	Melanoma skin cancer	4556	–	–	<ul style="list-style-type: none"> <li>- Does not perform validation independently</li> <li>- Not designed specifically to track cancer incidence so does not allow to identify cancer accurately</li> <li>- Weaken the signal of the protective effect</li> </ul>
KNN [6]	Skin lesions	40	–	–	<ul style="list-style-type: none"> <li>- Improvement required to guarantee even higher accuracy and faster results</li> <li>- Tested with limited number of images</li> </ul>
ANN [17]	Melanoma skin cancer	3388	–	–	<ul style="list-style-type: none"> <li>- ANN is hardware dependence</li> <li>- Developed in trial and error approach which does not provide an optimum result</li> <li>- Lack of a quantitative scale or scoring system for image interpretation</li> </ul>
DilatedVGG16, DilatInceptV3 [10]	Melanoma-Cancer	1796	128×128×3	3×3×3	<ul style="list-style-type: none"> <li>-Large number of parameters to be trained</li> </ul>
Kernel-Extreme-Learning-Machine (KELM) [2]	Melanoma-Cancer	2750	224×224×3	3×3×64	<ul style="list-style-type: none"> <li>-Limited segmentation accuracy</li> <li>-More computational time</li> </ul>
FCN [19]	Melanoma-Cancer	2000	128×128×64	1×1×64	<ul style="list-style-type: none"> <li>-Space limitations,</li> <li>-Larger computational time</li> <li>-Performance is too slow.</li> </ul>
FCN [1]	Melanoma-Cancer	–	128×128×3	3×3×64	<ul style="list-style-type: none"> <li>-Sensitive to noise</li> <li>-Uncertain in the data processing</li> </ul>
ECOC SVM, and CNN [1]	Skin-Cancer (Carcinoma, Melanoma)	3753	227×227×3	13×13×256	<ul style="list-style-type: none"> <li>-Biased result</li> <li>-Uncertain information</li> </ul>
CNN-HRFB [19]	Melanoma-Cancer	2000	128×128×64	1×1×64	<ul style="list-style-type: none"> <li>-Feature map is much smaller</li> <li>-Multiple up-samplings required</li> </ul>
CNN [15]	Skin diseases (Acne, Keratosis, Eczema-herpeticum, Urticaria)	174	227×227×3	3×3×256	<ul style="list-style-type: none"> <li>-Large filter size is required</li> <li>- Tested with a limited number of images</li> </ul>
ROI-based CNN [13]	Melanoma-Cancer	3738	227×227×3	5×5×48	<ul style="list-style-type: none"> <li>-Large count of ROI degrades the performance</li> </ul>

algorithms. However, based on comparative analysis, medical image classification for identifying skin cancer still lags behind in terms of accuracy, loss, noise, and computational resources, and therefore requires further improvement [16].

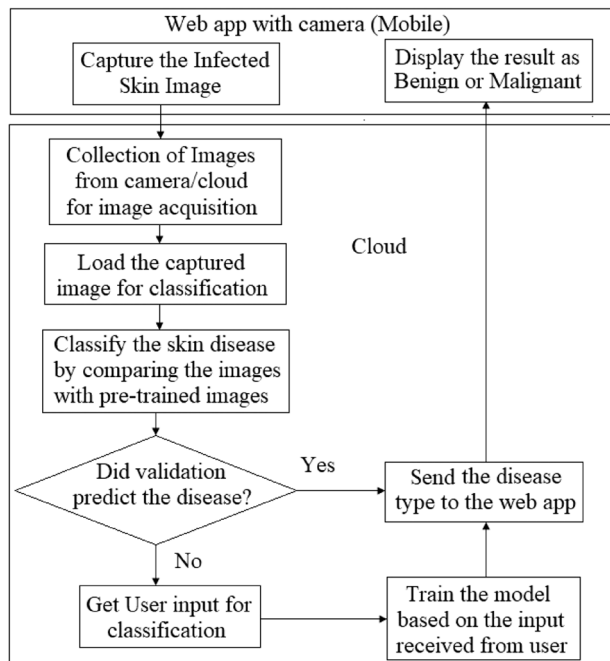
## 2 Proposed methodology

This paper presents a novel approach for identifying Malignant Melanoma using a CNN with higher accuracy and lower error rates. The proposed approach utilizes the RetNet50 CNN architecture to classify skin lesion images by identifying benign and malignant lesions. The following sections detail the design workflow, procedures for building a CNN-based ML model, training and retraining of the model, and evaluation standards.

### 2.1 Workflow of the design

The workflow of the proposed system for identifying malignant melanoma skin cancer is shown in Fig. 1. The classification model comprises two main modules. The first module consists of an image-capturing camera, which captures images of skin diseases for acquisition, and a web application that displays the graphical view to the user regarding the uploaded image and classified result. The system accepts data directly from the camera and from third-party libraries such as the Pi Camera Python library. In the second module, the pre-trained classification model performs malignant and benign classification based on the images captured and fed into it. The image classification module evaluates the captured or loaded image by comparing extracted features with pre-trained images for disease

**Fig. 1** Flow diagram of skin cancer diagnosis





**Fig. 2** Processing steps used in melanoma cancer diagnosis

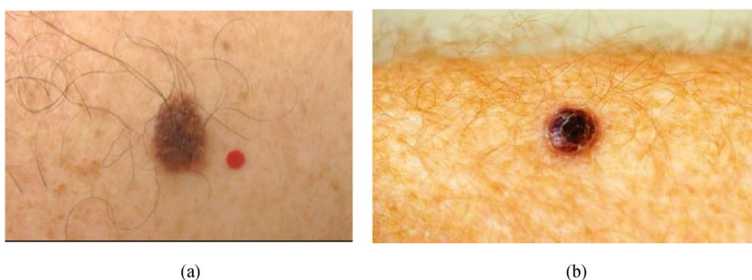
identification. The pre-trained learning model delivers information automatically from the cloud to identify the disease type, and human intervention is required only when the loaded image is unpredictable. In such a condition, the system accepts user input on unpredictable images and self-trains for future prediction.

The proposed classification system uses a CNN architecture with ResNet50 to identify cancer cells. A CNN with ResNet50 is effective for image classification and object detection in many computer vision tasks. ResNet50 blocks are used to allow the network to learn residual functions and enable it to better handle vanishing gradients during training. The CNN with ResNet50 can be fine-tuned for different computer vision tasks such as image classification, object detection, and semantic segmentation by adjusting the architecture and the training data. Fine-tuning refers to the process of using a pre-trained model as a starting point and then training it further on a new dataset specific to the task at hand, allowing the model to learn from a smaller dataset and improve its accuracy for the specific task.

## 2.2 Building a CNN-based ML model

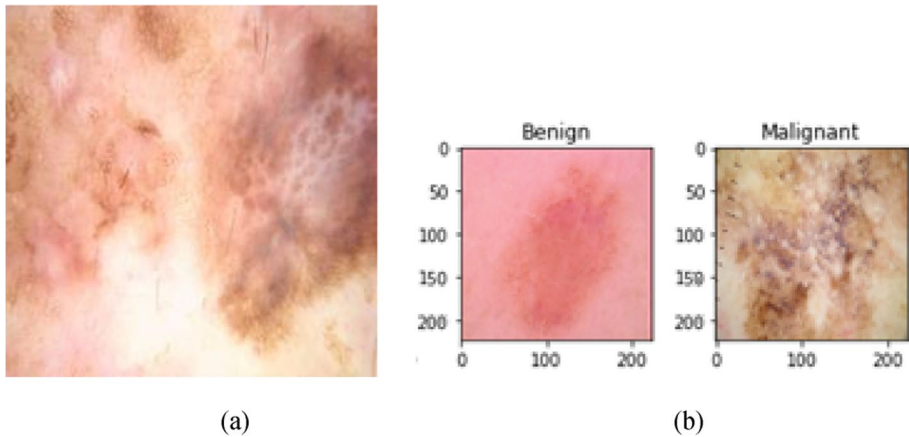
The processing steps used in building a CNN-based skin cancer classification model is illustrated in Fig. 2. The proposed methodology has six steps namely data collection, image preprocessing, image segmentation, image enhancement, training of CNN model, testing of CNN model, and building the web application.

**Data collection** Melanoma detection was initiated by the International Skin Imaging Collaboration (ISIC) organization to facilitate the early classification of skin diseases and avoid serious impacts on the human body. A CNN model was trained using the ISIC’s authenticated dataset. The proposed classification model utilizes 3300 images of both benign and malignant skin lesions to train and test the presence of malignant skin. Sample images of benign and malignant lesions are shown in Fig. 3.



**Fig. 3** Sample image of skin (a) benign (b) malignant

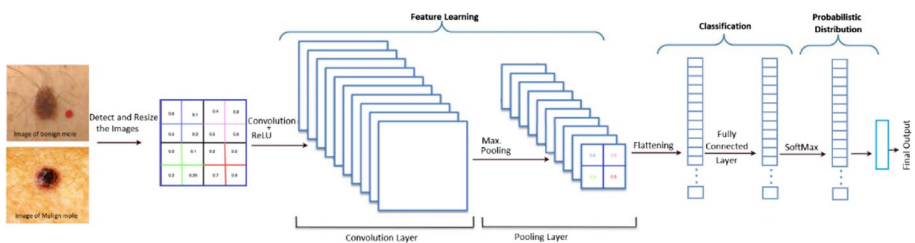




**Fig. 4** Image preprocessing (a)normal image (b) resized  $224 \times 224$  image

**Image pre-processing** Pre-processing of images is perhaps the most challenging part in building a classification model. This process ensures that the data is fit for use by filtering out unwanted noise, eliminating low-quality or unreliable pixels, and resizing images as per the requirements. Rescaling, resizing, and data augmentation are the main subclasses of image pre-processing. Rescaling and resizing the images provide reliable shapes to minimize the loss function in data training, and to provide independent, flexible, and precise pipelined data processing. In the proposed model, the images are transformed into  $224 \times 224 \times 3$  by manual or automatic cropping to increase model quality without any resampling. Figure 4 shows a sample of resized images in image preprocessing. Data augmentation performs random transformations of images, such as vertical and horizontal flips and shifts, clipping and cropping, padding, random rotation, and random zooming, to enhance the diversity of the trained data. Task-specific data augmentation introduced during preprocessing improves the abstraction levels, performance, and accuracy of skin lesion classification.

**CNN- model** CNN architectures generally uses large datasets to improve classification accuracy. However, the process of pooling and training large datasets can be time-consuming. Figure 5 shows the CNN architecture used in this work for Melanoma Malignant Diagnosis. Neurons that extract features from input images allow real-time monitoring and training of those images via learnable weights and biases. An effective learnable CNN



**Fig. 5** CNN architecture used for melanoma cancer diagnosis



model was developed in Tensorflow from scratch using Keras open-source neural-network library. The core of the proposed CNN model consists of convolutional layers, normalization technique, Rectified Linear Unit (ReLU), pooling, zero padding, flattening layer, dropout, dense layers, fully connected layer, softmax, and backward propagation layers. These layers combine to form a full functional convolution operation. The Keras Sequential API supports adding one layer at a time to the network, starting with Conv2D. The pre-trained convolution layer extracts the most relevant features from the image to define changes in the behavior of neurons. The benign and malignant images are identified from extracted features by convolution size of (224, 224, 64). Kernel size is selected in Keras to decide the portion of the image to be transformed by the kernel filters. In the CNN model, the size of the filter and image must be symmetric to reserve spatial arrangement in the image. Normalization is the next layer used to improve the performance of the network, learning rate, and the weights and biases of the activation function. Using batch normalization, the proposed model normalizes the images by adjusting and rescaling undesirable changes to the input layers before moving on to the next distribution. The next layer is a non-linear layer called ReLU, which improves the learning rate of the normalized images without decreasing the feasibility, speed, and stability conditions.

Pooling is the next layer, which reduces the size of the input in the process and noise in the evaluation parameters to avoid overfitting. The proposed model uses a mixed hybrid pooling technique that boosts the invariance properties to yield state-of-the-art performance. In mixed hybrid pooling, maximum pooling improves classification accuracy and speed, while average pooling smoothens feature extraction. A  $2 \times 2$  filter used in the image process offers a maximum number of input pixels ( $W1 \times H1 \times D1$ ) and reduced pixels in output ( $W2 = (W1 - \text{size}) / \text{stride} + 1$ ,  $H2 = (H1 - \text{size}) / \text{stride} + 1$  and  $D2 = D1$ ) which is about 25% of the input pixels. Zero-padding controls the image dimensions in a generic form after applying filters to eliminate the loss of features at the boundaries. The flattening process mixes all the resulting 2-dimensional arrays into a single long feature vector without much interpretation in the feature map and is less prone to overfitting. Dropout regularizes the learning process to reduce overfitting by adjusting input units to zero or removing neurons temporarily in Keras. In the dense layer, neural networks are deeply connected to receive all input features from all neurons for effective prediction. A fully connected layer is a feed-forward neural network layer that connects neurons in non-linear combinations of receptive fields for recognizing and classifying images effectively with a CNN model. Softmax is an activation function that identifies an object according to a multinomial probability distribution. Backpropagation tracks the changes in output due to the change in weight and bias to minimize the loss function. The proposed CNN model efficiently extracts features through convolution, normalization, ReLU, pooling, padding, flattening, dropout, dense, and fully connected layers to provide high-level knee joint image classification results on the output side.

To improve classification accuracy, CNN was further developed as a complex-valued CNN (CV-CNN), and CNN with ResNet50. CV-CNN performs complex-valued operations with the compactness of convolution, the nonlinear activation functions, residual learning, and the noise robustness of residual blocks. ResNet50 can classify images with large datasets like training 1.2 million images and testing 50,000 images from ImageNet without degrading the performance of the classification model. Residual learning in image recognition provides high-level feature extraction for effective classification. Also, it provides easy and fast computation for feature extraction and classification with improved precision in a neural network. The residual Network Equation is derived as follows,

$$Y_i = h(X_i) + F(X_i, W_i) \quad (1)$$

$$X_{i+1} = f(Y_i) \quad (2)$$

Where, F-stacked non-linear layer, f-Rely activation function.

It is noted that when both  $f(Y_i)$  and  $h(X_i)$  become identity mappings, it is possible for direct propagation of signals in both forward and backward directions from one unit to other units of the network. In addition, minimal error rates are achievable when  $f(Y_i)$  and  $h(X_i)$  benefit from identity mappings. Now, the next step is to train and evaluate the CNN model.

### 2.3 Training and evaluation of classification models

A CNN model was trained and evaluated using a data set of 3300 images of benign and malignant lesions. The dataset was divided into two sets with an 80:20 ratio for split validation. Eighty percent of the images in the dataset were used for training the CNN model, while 20% of images were used for testing and validation. Precisely resized and pre-processed images provided high accuracy with 20% of the testing and validation sets. K-fold cross-validation was utilized to split the data in order to evaluate skin cancer with limited data samples. This approach improves models performance by ensuring the best fit. In the proposed model, 3-fold cross-validation was utilized to optimize performance parameters and maintain faster training. Split validation prevents the classification model from overfitting during the training process, thereby improving performance and providing high-precision image classification during evaluation. Table 2 shows the data splitting of training and testing datasets.

The CNN classification models were developed and trained on Google Colab, which is a cloud platform providing a suitable environment to develop and train classification models with the support of standard libraries. This cloud platform provides up to 25GB of RAM, and a Tesla K80 GPU. The extracted skin images were divided into two sets, with 80% of the data frame comprising a training set and 20% of data constituting a testing set for Kellgren-Lawrence (KL) grades. The fully connected MaxPool2D and Conv2D layers of a CNN were used to extract features from knee joint images for classification. The model architecture and summary are presented in Fig. 2. The input X-ray images are transformed and fed as input data to the network. Every 50 epochs, the entire dataset is trained in the network until the classification model reaches its global optimum. The global optimum is selected as a region where the model provides insignificant variation in loss and accuracy functions. To train the proposed model, the transfer learning paradigm is also employed, which facilitates the high-precision classification of generic types of medical images. Furthermore, transfer learning enables the CNN model to achieve high accuracy in testing and

**Table 2** Training and testing dataset

Type of Images	Train	Test and validation	Total
Benign	1440	360	1800
Malignant	1200	300	1500
Total	2640	660	3300

validation with limited datasets. Additionally, it customizes the model to extract relatively useful spatial features for accurate validation through feature extraction and fine-tuning. Fine-tuning enables parallel training for previously and newly added layers to improve accuracy and resource-saving. After training, the model was normalized by adjusting the weights and biases to assess the reliability of the model on updated and different sets of sample data, which supports assessing the fitness of the model. Retraining was also performed to improve the model's performance by utilizing the knowledge of existing classifiers. Transfer learning enables models to be retained on specific datasets in order to significantly enhance classification precision. The CNN, CV-CNN, and ResNet50 classification models were trained and retrained on the same datasets for performance comparison. The testing and training of all three classification models were done using ImageNet datasets for accuracy improvement and fed for fifty epochs until the global optimum was reached. The convolutional weights of the models were kept constant after the training, retraining, and normalization processes to operate the model in evaluation mode. A web application was also developed using the Flask framework on Visual Studio IDE to get the final version of the benign and malignant testing layout. The user-friendly styling and layout were implemented using HTML, CSS, Javascript, and Bootstrap.

## 2.4 Evaluation standards

The proposed malignant melanoma cancer diagnosis models should be evaluated for accuracy and loss functions. A balanced view of the classification models should be tested through the F1-score, calculated from precision and recall metrics. The performance metrics are expressed as follows.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$\text{Loss} = \sum_{i=0}^M \sum_{j=0}^N (y_{i,j} \cdot \log(y_{i,j})) \quad (6)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Where, TP - True Positive, TN - True Negative, FP - False Positive, FN - False Negative,  $y$  is the predicted value.

Fit statistics is a value indicating the difference between two population means. It can be observed through regression (R) analysis by determining the mean between two populations. The coefficient of determination, R-square ( $R^2$ ), can be used to determine the differences in one variable from the second variable, as well as to calculate the likelihood of future predictions occurring. R-square gives a variation between the variables within the range of 0 to 1.

$$R - square = 1 - \frac{SSE}{SST} = 1 - \frac{MSE(n - p - 1)}{MST(n - 1)} \quad (8)$$

Where, SSE is the sum of squares error, SST is the sum of squares total, MSE is the mean square error, MST is the mean square total,  $p$  is number of predictors in the regression equation, and  $n$  is the total number of samples.

Root Mean Square Error (RMSE) refers to the standard deviation of the dependent residuals, which is used to measure the deviation of residual prediction lines from regression lines to determine which distribution fits best. RMSE is commonly used in high-precision data classification such as medical image classification, climatology, forecasting, and regression analysis to verify experimental results.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (f_i - o_i)^2}{n}} \quad (9)$$

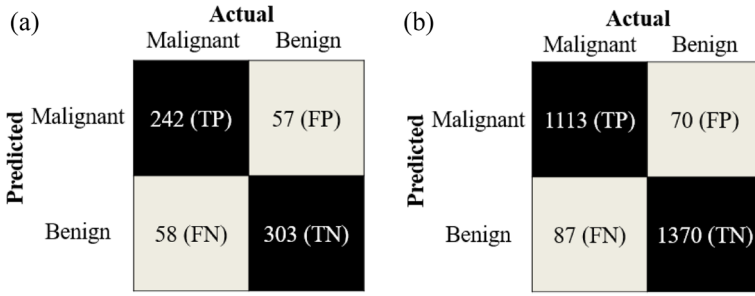
Where  $f_i$  and  $o_i$  are the expected and observed results of residuals. The difference between  $f_i$  and  $o_i$  represents the line of best fit.

### 3 Experimental results

#### 3.1 Performance analysis

The CNN ResNet50 architecture was developed from scratch to improve the skin cancer classification accuracy based on pre-trained images. Tensorboard, a web app extension of Tensorflow, was used for the development of the classification model, as well as for training and testing the model using predefined skin lesion datasets. Also, it allows to calculate the performance metrics such as accuracy and loss. Performance parameters were used to evaluate the proposed classification model, with the images described primarily by weight deviations and bias distribution over the trained epochs. Every slice in the image represents the distribution of the parameter at a specific epoch. The presented model was developed and tested using 50 epochs for optimal response.

The ResNet50 model was trained using 2640 images of both benign (1440) and malignant (1200) lesions from Google Colab. Skin lesion images were divided into two sets, with 80% of the data used for training and 20% for testing Kellgren-Lawrence grades. Knee joint features in image classification were extracted from fully connected MaxPool2D and conv2D layers. The input size was transformed and fed into the network, which was trained for 50 epochs. The trained data were split into three groups and cross-validated individually to increase accuracy. The performance of the pre-trained classification model was evaluated using test set images for loss and accuracy. A total of 660 images of both benign (360) and malignant (300) lesions were used to test the accuracy of the classification models. The classification performance of the proposed ResNet50 malignant classifier model is presented in Fig. 6a and b for the test model and training model, respectively. The confusion matrixes were used to compare the predicted values generated by the machine learning model with the actual target values, enabling evaluation of the performance of the model and identification of areas for improvement. Figure 7a and b present the receiver operating characteristic (ROC) curves of the proposed ResNet50 model during testing and training,

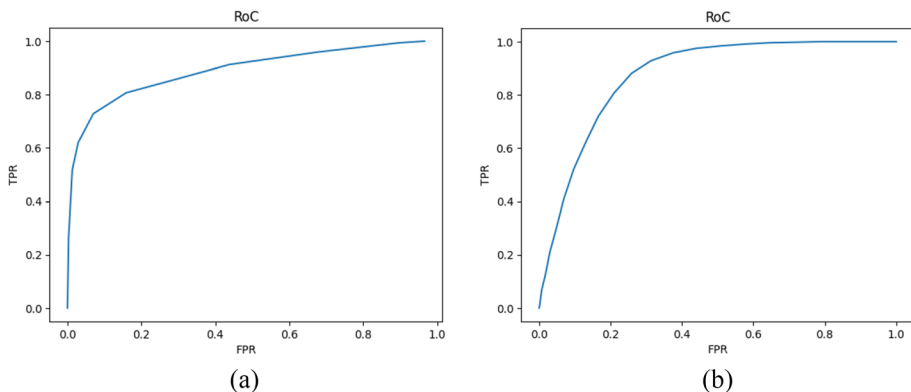


**Fig. 6** Confusion matrices on (a) test model (b) training model

respectively. These curves are a valuable tool for visualizing the trade-off between true positive rate (TPR) and false positive rate (FPR), allowing the identification of a threshold that minimizes the cost of misclassification. The ROC curves for the proposed ResNet50 binary classifier model indicate the outstanding diagnostic ability of the classifier.

To analyze the performance metrics, models based on DT, SVM, KNN, ANN, CNN, and CV-CNN were developed, trained, and tested on the same datasets used for the proposed ResNet50-based classification model. The experimental observations on precision, recall, accuracy, loss, and F1-score are recorded in Table 3 for performance analysis. In comparison, ResNet50 achieved a 48.3% improvement in accuracy compared to the lowest accuracy model and 3.21 times reduced loss compared to the highest loss model in skin cancer classification. The precision of CNN with ResNet50 is 30% higher than CNN, 22.69% higher than CV-CNN, and 12.03% higher than SqueezeNet. The accuracy and loss functions of the CNN model are plotted in Fig. 8a and b, respectively, while Fig. 8c and d show the accuracy and loss functions of ResNet50.

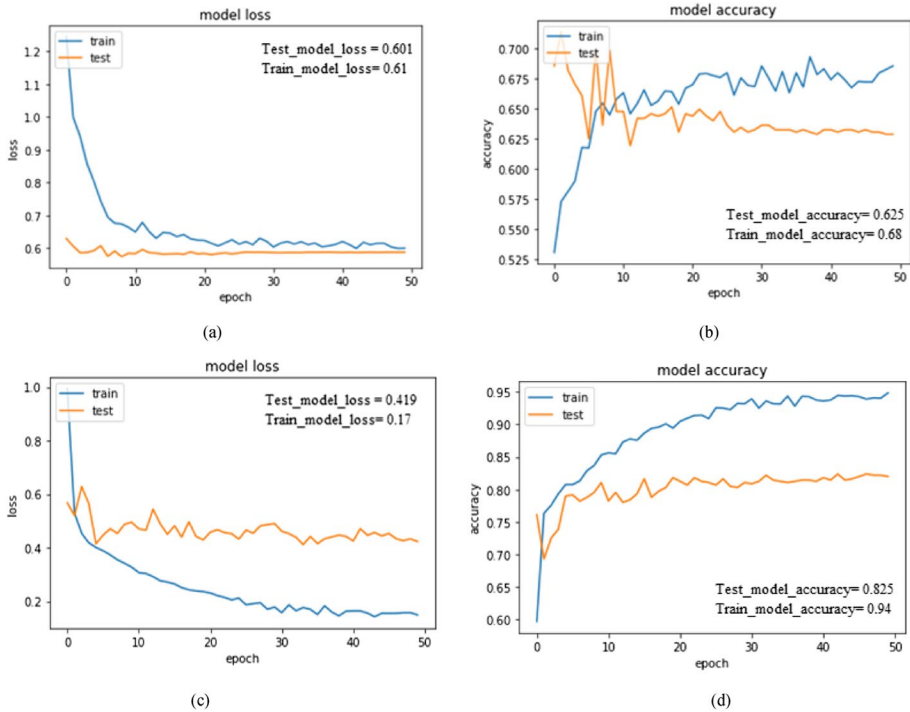
The analysis of variance (ANOVA) test was used to calculate the descriptive statistics for the accuracy and loss functions. The results were recorded in Table 4. ANOVA was used to perform statistical analysis to identify the unique features of each diagnosis system by examining the statistical difference between the test outcomes to determine the best fit. Figure 9a and b show the bar charts comparing the mean accuracy and loss, respectively,



**Fig. 7** ROC curve (a) test model (b) training model

**Table 3** Performance analysis (in %)

MODELS	Test Model					Trained Model				
	Precision	Recall	Accuracy	Loss	F1-Score	Precision	Recall	Accuracy	Loss	F1-Score
DT	44.8	44.2	43.4	71.9	44.5	52.8	52.0	51.8	53.6	52.4
SVM	47.8	47.5	47.1	68.3	47.6	58.7	57.1	56.4	51.2	57.9
KNN	43.9	42.7	41.3	72.8	43.3	49.3	49.7	48.6	54.6	49.5
ANN	57.8	57.4	56.4	65.1	57.6	62.1	60.8	60.3	48.2	61.6
CNN	63.7	63.2	62.5	60.1	63.4	66.7	65.2	68.0	42.7	66.9
CV-CNN	67.6	66.2	68.8	59.3	66.9	74.6	74.3	72.5	35.9	74.0
SQUEEZENET	80.2	79.9	80.4	42.5	80.1	84.9	84.3	85.7	36.8	84.6
RESNET50	81.1	80.6	82.5	41.9	80.8	94.1	92.7	94.0	17.0	93.4



**Fig. 8** Evolution plot during the training and testing (a) loss of CNN (b) accuracy of CNN, (c) loss of ResNet50, and (d) accuracy of ResNet50

among the different skin disease diagnosis models. These bar chart comparisons indicate the statistical differences between the means of the various diagnosis models. Figure 9c and d show the standard deviation (error bars) for accuracy and loss, respectively. These error bars help to characterize the data and assess the prediction fitness for the loss and accuracy models.

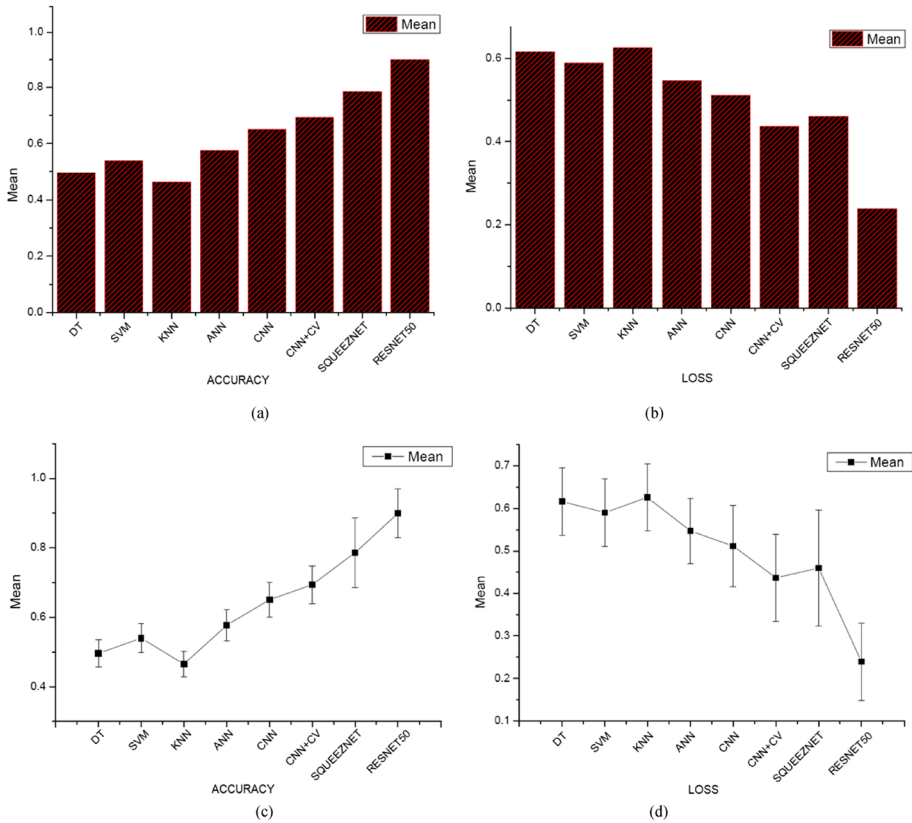
Table 4 presents the ANOVA observations from the descriptive statistical analysis in terms of mean, standard deviation, and standard error of the mean. In comparison, the ResNet50 architecture provided a 48.3% improvement in mean accuracy and 2.54 times less mean loss function compared to the lowest mean of the KNN model. The standard deviation in descriptive statistics measures the degree of clustering around the mean, while the standard error of the mean measures the dispersion of the data in relation to the mean. A large value of the standard deviation and standard error in the table indicates that the data samples are widely spread around the mean in the accuracy and loss functions of the ResNet50 model.

In Table 5, the overall ANOVA test results for the compared skin cancer diagnosis models are presented. The F-value in the one-way ANOVA helps to identify significant variance between the means of different populations. The F-value was calculated with a significance level of 0.05, by dividing the numerator degrees of freedom defined between the model count of 7 and the denominator degrees of freedom defined within the model count of 392. The statistical ANOVA test table shows that the F-value is much larger than the critical F-values for both accuracy and loss functions, indicating the rejection of the null



**Table 4** The descriptive statistics on accuracy and loss functions

Algorithm	Epochs	Accuracy (%)		Loss (%)					
		Accuracy		Loss					
		Mean	Standard Deviation	SE of Mean	Standard Deviation				
DT	50	51.8	0.496	0.038	0.0055	53.6	0.6162	0.0788	0.01115
SVM	50	56.4	0.540	0.042	0.0059	51.2	0.5900	0.0794	0.01123
KNN	50	48.6	0.465	0.036	0.0051	54.6	0.6259	0.0781	0.01104
ANN	50	60.3	0.577	0.045	0.0064	48.2	0.5467	0.0772	0.01092
CNN	50	68.0	0.651	0.051	0.0071	42.7	0.5115	0.0959	0.01356
CV-CNN	50	72.5	0.694	0.054	0.0076	35.9	0.4368	0.1018	0.01439
SQUEZENET	50	85.7	0.786	0.101	0.0142	36.8	0.4599	0.1373	0.01941
RESNET50	50	94.0	0.900	0.070	0.0099	17.0	0.2393	0.0916	0.01286



**Fig. 9** Comparative analysis of learning models (a) mean value plot of accuracy, (b) mean value plot of loss, (c) standard deviation error bars in accuracy, and (d) standard deviation error bars in loss

hypothesis. The large F-value of the proposed CNN-based classification models provides statistical evidence of significant performance improvement in skin cancer classification compared to the performance of the other seven models. The fit statistics show that the proposed classification algorithm has the lowest Root Mean Squared Error (MSE), indicating that it is the best fit in training linear regression models.

Table 6 presents a comparison of the performance of the proposed ResNet50 model with existing literature. The optimized ResNet50 model exhibited the greatest improvement in F1 score compared to other skin cancer classification models. A high F1 score indicates that the proposed CNN based on the ResNet50 architecture can classify skin disease images with a higher statistical accuracy than other algorithms.

### 3.2 Results from web-application

The final version of the web application was developed to display web-based melanoma identification, as shown in Fig. 10. To make it easier for end-users, a colorful user interface was created in the form of a web application to display information

**Table 5** Descriptive statistics

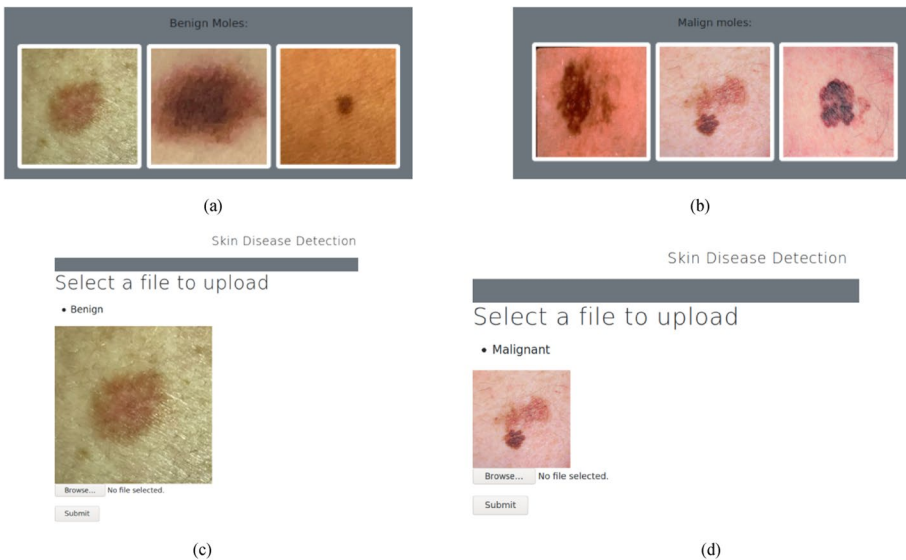
Source	DF	Accuracy			Loss				
		Sum of Squares	Mean Square	F Value	Critical F Value	Sum of Squares	Mean Square	F Value	Critical F Value
Between Models	7	7.853	1.1218	331.11	1.9384	5.662	0.8088	90.800	1.9384
Within Models	392	1.328	0.0034	–	–	3.492	0.0089	–	–
Corrected Total	399	9.181	–	–	–	9.153	–	–	–
Fit Statistics		R-Square	Coeff Var	Root MSE	Data Mean	R-Square	Coeff Var	Root MSE	Data Mean
		0.85534	0.09117	0.05821	0.63846	0.61853	0.18753	0.09438	0.50327

\*At the 0.05 level, the population means are significantly different. \*Obtained F-value is likely to occur by chance with a  $p < 0.05$  which rejects the null hypothesis

**Table 6** Comparative analysis

Techniques	Identification	Accuracy	Precision	Recall	F1-Score
Proposed	Melanoma	94.0	94.1	92.7	93.4
DilatedVGG16 [10]	Melanoma	90.1	87.0	87	87
SVM and CNN [1]	Melanoma	91.8	–	–	–
DilatInceptV3 [10]	Melanoma	90.95	89.0	89	89
SqueezeNet [16]	Melanoma	92.18	80.77	95.1	–
AlexNET CNN [15]	Acne	85.7	–	–	–
	Keratosi	93.3	–	–	–
	Eczema Herpeticum	93.3	–	–	–
	Urticaria	92.8	–	–	–
KELM [18]	Melanoma	90.67	–	90.20	–
FCN [19]	Melanoma	92.7	–	–	–
CNN-HRFB [19]	Melanoma	93.8	–	–	–

about the risk and prediction of skin cancer. The application informs people about the classification outcome as benign or malignant and provides a safety level or risk factor. Figure 10a and b show sample images used in the web-based skin cancer prediction process. Figure 10 c and d show the predicted results of a sample image uploaded to the web application. The technology-based skin disease prediction system allows for early detection of skin diseases and proper medication according to the severity of the disease. In cases of malignancy, seeking medical treatment with a doctor is essential, as it is a severe form of skin cancer.



**Fig. 10** **a** Sample benign images used for evaluation, **b** sample malignant images used for evaluation, **c** test result of a benign mole on web application, **d** test result of a malignant mole on web application

## 4 Conclusion

A novel web-based skin disease diagnostic model using CNNs has been proposed. A customized CNN model was developed from scratch to classify Malignant Melanoma Cancer. During testing, the developed CNN model achieved a maximum accuracy of 62.5%, while it achieved an accuracy of 68% during training. To improve the performance of skin cancer classification, CV-CNN, SqueezeNet, and ResNet50 were also developed, trained, and tested with the same set of skin images. The classification models were trained and validated using split validation, where 80% of images (2640) were used for training, and 20% of images (660) were used for testing and validation. In the performance analysis, ResNet50 achieved a maximum accuracy of 94% in the malignant melanoma image classification. The accuracy achieved by ResNet50 is 27.7% higher than that of CNN, 22.9% higher than that of CV-CNN, and 8.8% higher than that of SqueezeNet. Achieving a testing accuracy of 94% is a significant achievement in medical image classification. Furthermore, the large F-value in the statistical analysis provides evidence of significant performance improvement in skin cancer classification compared to the performance of the other seven models. The F1-score achieved by the proposed ResNet50 model is higher than that of other algorithms. The proposed ResNet50 model is the most suitable for predicting malignant melanoma skin cancer with the best fit and improved statistical rate of classification.

**Funding** No funding was received to assist with the preparation of this manuscript.

**Data availability** Not Applicable.

## Declarations

**Competing interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Adegun A, Viriri S (2020) FCN-based DenseNet framework for automated detection and classification of skin lesions in Dermoscopy images. *IEEE Access* 8:150377–150396
2. Adegun A, Viriri S (2021) Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. *Artif Intell Rev* 54:811s implemented–841
3. Ashraf R et al (2020) Region-of-interest based transfer learning assisted framework for skin Cancer detection. *IEEE Access* 8:147858–147871
4. Bakkouri I, Afdel K (2019) Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images. *Multimed Tools Appl* 79:20483–20518. <https://doi.org/10.1007/s11042-019-07988-1>
5. Bakkouri I, Afdel K (2022) MLCA2F: multi-level context attentional feature fusion for COVID-19 lesion segmentation from CT scans. *SIViP* <https://doi.org/10.1007/s11760-022-02325-w>
6. Ballerini L, Fisher RB, Aldridge B, Rees J (2013) A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. *Lecture notes in computational vision and biomechanics*, vol 6. Springer, Dordrecht. [https://doi.org/10.1007/978-94-007-5389-1\\_4](https://doi.org/10.1007/978-94-007-5389-1_4)
7. Chao X (2022) An improved transformer network for skin cancer classification. *Comput Biol Med* 149(2022):105939
8. Chaturvedi SS, Tembhurne JV, Diwan T (2020) A multi-class skin Cancer classification using deep convolutional neural networks. *Multimed Tools Appl* 79:28477–28498. <https://doi.org/10.1007/s11042-020-09388-2>

9. Dhivyaa CR, Sangeetha K, Balamurugan M et al. Skin lesion classification using decision trees and random forest algorithms. *J Ambient Intell Humaniz Comput* (2020). <https://doi.org/10.1007/s12652-020-02675-8>
10. Dorj UO, Lee KK, Choi JY et al (2018) The skin cancer classification using deep convolutional neural network. *Multimed Tools Appl* 77:9909–9924
11. Hatem MQ (2022) Skin lesion classification system using a K-nearest neighbor algorithm. *Hatem Visual Compu Indust Biomed Art* 5:7. <https://doi.org/10.1186/s42492-022-00103-6>
12. Jaisakthi S M, Mirunalini P, Aravindan, C et al. (2022) Classification of skin cancer from dermoscopic images using deep neural network architectures. *Multimed Tools Appl* <https://doi.org/10.1007/s11042-022-13847-3>
13. Kadampur MA, Al Riyae S (2020) Skin cancer detection: applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Inf Med Unlocked* 18:1–6
14. Manu G, Thomas K, Shaofeng Y, Saeed H (2020), artificial intelligence-based image classification methods for diagnosis of skin cancer: challenges and opportunities. *Comput Biol Med* 127 (2020) 104065
15. Monika, M, Vignesh, N, Kumari, U, Kumar, MNVSS, Lydia, L (2020) Skin cancer detection and classification using machine learning. *Mater Today Proceed* 33. <https://doi.org/10.1016/j.matpr.2020.07.366>.
16. Olusola OA, Robertas D, Sanjay M, Rytis M, Adebayo A (2021) Malignant skin melanoma detection using image augmentation malignant skin melanoma detection using image augmentation by over sampling in nonlinear lower-dimensional embedding manifold. *Turk J Electric Eng Comput* 29(8):2600–2614
17. Tumpa, PP, Kabir, MA (2021) An artificial neural network based detection and classification of melanoma skin cancer using hybrid texture features. *Sensors International* 2:100128. <https://doi.org/10.1016/j.sintl.2021.100128>
18. Raajan NR, Lakshmi VSR, Prabakaran N (2021) Non-invasive technique-based novel Corona(COVID-19) virus detection using CNN. *Natl Acad Sci Lett* 44:347–350
19. Shanthi T, Sabeenian RS, Anand R (2020) Automatic diagnosis of skin diseases using convolution neural network. *Microprocess Microsyst* 76:1–8
20. Tanaka T, Voigt MD (2018) Decision tree analysis to stratify risk of de novo non-melanoma skin cancer following liver transplantation. *J Cancer Res Clin Oncol* 144:607–615. <https://doi.org/10.1007/s00432-018-2589-5>
21. Walid B, Afifa K (2021, 2021) Skin lesion image retrieval using transfer learning-based approach for query-driven distance recommendation, *computers in biology and medicine*. 137:104825
22. Wei L, Ding K, Hu H (2020) Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. *IEEE Access* 8:99633–99647. <https://doi.org/10.1109/ACCESS.2020.2997710>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

M. Senthil Sivakumar<sup>1</sup>  · L. Megalan Leo<sup>2</sup> · T. Gurumekala<sup>3</sup> · V. Sindhu<sup>1</sup> · A. Saraswathi Priyadharshini<sup>4</sup>

✉ M. Senthil Sivakumar  
msenthilsivakumar@gmail.com

<sup>1</sup> Indian Institute of Information Technology Tiruchirappalli, Tiruchirappalli, Tamil Nadu, India

<sup>2</sup> Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

<sup>3</sup> MIT Campus, Anna University, Chennai, Tamil Nadu, India

<sup>4</sup> Vellore Institute of Technology, Vellore, Tamil Nadu, India