DOI: 10.15849/IJASCA.241130.07

nt. J. Advance Soft Compu. Appl, Vol. 16, No. 3, November2024
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Deep Learning and Statistical Operations Based features extraction for Skin Cancer Detection

Amal Abu Shehab[1], and Saleh Al-Sharaeh[2]

King Abdullah II School of Information Technology
The University of Jordan
Amman, Jordan
e-mail: AML8180635@ju.edu.jo

King Abdullah II School of Information Technology
The University of Jordan
Amman, Jordan
e-mail: ssharaeh@ju.edu.jo

Abstract

Skin cancer is considered as one of the most serious types of cancer that leads to death worldwide. The number of deaths that caused by skin cancer can be reduced if it is diagnosed at early stages. Skin cancer is usually diagnosed using visual inspection, but it is less accurate. Using deep learning-based methods have been proposed to assist the doctors to diagnose the skin cancers at early accurately. The investigation was achieved on 3600 images collected from kaggle. Two deep learning algorithms used in the study: Vgg-19 and Alexnet to extract the layers 6 and 7. Then, the two layers will be merged to generate the statistical operations like: median, lowest, highest and joined of two layers. The following classifiers were used in the investigation: K-Nearest Neighbor, Random Forest, Naïve Bayes and Decision Tree. However, the study considered the following measures: accuracy, precision, recall, and F-Measure. Three training datasets sizes will be used to investigate the influence on classification accuracy. The results of all datasets were slightly similar. This approves that the extracted features and the statistical operations have an influence on the classification accuracy. The results show that Alexnet performs high accuracy and consumes less time that required for training the model compared to vgg-19. The results of classifiers showed that Random Forest scored high classification accuracy (85.6) compared to other ML classifiers

Keywords: *Skin cancer detection; Deep learning; Convolutional neural networks (CNN); VGG19; Alexnet; Skin tumor detection.*

1 Introduction

One of the most common forms of cancer in the modern world is thought to be skin cancer [1]. It makes sense that skin cancer is the most prevalent kind of cancer in people, given that the skin is the largest organ in the body [2]. Benign and malignant skin cancer

Received 4 August 2024; Accepted 10November 2024

are the two main classifications for the disease. Skin cancer is a rare, lethal, and serious condition [3]

According toAmerican Cancer Society report that issued in 2022, it says that the melanoma cases of skin cancer make up only 1% of the total cases, but it caused a higher average of death [4]. While in the statistical report issued by [5], it states that - in the United States, the diagnosed people with skin cancer every day is more than 9,500. More than two people die from the disease every hour. Also, they stated that – in worldwide, more than 5,400 people die from non-melanoma skin cancer every month.

Skin cancer develops in human cells known as melanocytes. It happens when healthy melanocytes are growing out of control, resulting in a cancerous tumor [6]. Whereas melanoma skin cancer appears in areas exposed directly to sunlight, such as the faces, or hands, etc. It can be cured if it was diagnosed early; otherwise, it will be spread out to other areas of body and cause health condition problem and death [7]. Therefore, diagnosis and treatment at early stages is considered the critical factor to avoid the dangerous cases [8]. The traditional method used for detection is the biopsy for skin cancer. In this method, the doctors take a sample of a suspected skin for medical check purposes to decide if it is affected or not [8]. This traditional method is painful, slow for the patients, and very effort and time-consuming for the doctors. Using computer-based technology can provide an easy method for humans with less expensive, and speedy detecting skin cancer [9].

Deep learning has achieved notable results compared with other approaches of machine learning. It is now considered the most advanced subfield of machine learning that deals with artificial neural network algorithms. These algorithms are built based on the function and structure of the human brain. Whereas the DL is implemented in a wide scope of research e.g. voice recognition [10], human pattern recognition [11], and in skin cancer detection [12].

Despite it is possible to identify skin cancer manually by specialists, but this takes a long time and effort. Although algorithms based on automatic detection can aid in identification processes, there is currently no golden process for skin cancer image detection. There are multi researches have been conducted to reach to high classification accuracy. This study is one of them that conducted on skin cancer to find desirable results using different methods compared with previous studies [13].

This research is prepared to detect skin cancer images if it is benign or malignant. The proposed model provides a new technique based on extracting features (i.e. fc6 and fc7) from MRI images using the models: VGG-19 and Alexnet. Based on the extracted features: fc6 and fc7, new datasets will be generated and give the title: Statistical Operations, which includes: median, highest, lowest, and joined between features fc6 and fc7. The total datasets we will have in this study is six datasets: (i.e. fc6, fc7, median, highest, lowest, and joined between fc6-fc7 features). In order to classify the datasets, the following Machine Learning (ML) algorithms are used: KNN, Naïve bayes, Decision Tree, and Random Forest. This study will provide the literature field with comparison results for the extracted features from VGG-19 and Alexnet using the aforementioned classifiers. This considered as a contribution for this study. Also, to provide an insight into how to use fully connected layers (fc6 and fc7) to build new datasets for statistical operations to be used for skin cancer detection. Then, to find their influence on the classification accuracy; this can be considered as a contribution for this study.

This study aims at examining the use of deep learning models on classifying skin cancer images if it is benign and malignant. This aim can be performed by taking into account the following set of objectives: 1) To evaluate the possibility of classifying skin cancer images using ML based on deep learning techniques. 2)To evaluate the performance of different classifiers (i.e. KNN, Naïve bayes, Random Forest, and Decision Tree) on the reliable database of skin cancer. 3)To assess the influence of using the statistical operations on the classification accuracy. 4) Provide the literature work with a comparison of the results of ML based on using two deep learning models (VGG-19, and Alexnet) with considering different training datasets sizes. The results will be evaluated using Accuracy, Precision, Recall and F- Measure.

Since no large dataset is being used in real-world datasets, as well as the pre-trained CNN models required a huge number of images for ideal training to provide enhanced performance. Based on Vgg19 and Alexne, it was suggested to use different scenarios for training dataset sizes to investigate their influence on classification accuracy. This is to increase the variability, which can help the model generalize better for different number of image in the dataset, and this is especially significant in medical imaging that have limitation of sizes and variability.

The remaining part of the paper is organized as follows: Section two demonstrated the related works that have been done in skin cancer. The proposed methodology is investigated in Section three. Section four illustrates the results in details with some performance measures. While section five draws the conclusion.

2 Related Work

Recently, CNN is the most powerful deep learning technique used for image classification in the medical field [4]. Also, it has shown amazing findings compared with other techniques like finding in clinical epidermal 2 cell [6], cervical cell image classification and skin cancer classification [14] [15] [16] [17].

There are number of researchers who used CNN technique, such as: the researchers in [14] designed a model using CNN and a different of learning frameworks with minimal training data. Another researcher has in [18] used a pre-trained CNN technique and received the advantage that represented by the finding .

During the last few decades, several models have been suggested by researchers for the classification of skin cancer problems. These were suggested for feature extraction, and then were categorized in the International Skin Imaging Challenge (ISIC) in 2016. This is to classify the skin cancer images into two categories: benign and malignant [19]. In the subsequent year (2017), a larger dataset was collected and released relative to the 2016 version for segmentation, identification, and classification skin cancer images: known as [ISIC 2017] [19]. A new automated method was designed for melanoma recognition in the year 2017, then was added to [ISBI-2016] [20]. The affected skin was then separated to identify melanoma and non-melanoma lesions.

Recently, a new study conducted to examine the effectiveness of features extracted by eight contemporary CNN models using four popular datasets including: PH2, ISIC 2016, ISIC 2017 and HAM10000. The CNN models used in the study are: AlexNet, VGG-16, VGG-19, Inception V3, MobileNet, ResNet-50, DenseNet-121 and EfficientNet B0. Their results indicated that the DenseNet-121 with multi-layer perceptron have achieved a higher accuracy of 98.33%, 80.47%, 81.16% and 81% on the following datasets: PH2, ISIC 2016, ISIC 2017 and HAM10000 when compared to different CNNs models and state-of-the-art methods [21].

In the same field, Faghihi et al. [22] investigated the CNN technique on skin cancer by applying transfer learning models based on VGG16 and VGG19 architectures. They evaluated the performance of the models using standard metrics such as accuracy, precision, recall, and F1 score. The results showed the effectiveness of the transfer learning method and showed an improvement in classification accuracy with an increase of 3% (from 94.2% to 98.18%) when compared to different methods [22].

The paper by Ankush Singh et al. (2023) suggested some advanced computational methods to enhance the accuracy and efficiency of detection for skin cancer. They used convolutional neural networks (CNNs): Alexnet and VGG-16 to automatically extract features from skin images. The study emphasizes the significance of leveraging deep learning algorithms for their ability to discern intricate patterns indicative of skin cancer. The results showcase the efficacy of their approach and get an accuracy of 88.48% in the case of AlexNet and 90.41% in the case of VGG 16 respectively. On ISIC 2020, VGG16, VGG19, and SVM exhibit accuracies of 90%, 92%, and 92%, respectively [23].

The paper by Md. Mahbubur Rahman et al. (2023) proposed a new hybrid system for melanoma skin cancer detection. The aim of this research is to improve investigation of skin cancer from dermoscopy images by mitigating multiplicative speckle noise using an anisotropic diffusion filtering technique. The fast-bounding box (FBB) technique is used to partition the areas affected by skin cancer. There are two different feature extractors used: the VGG19-based Convolutional Neural Network (CNN) and the Hybrid Feature Extractor (HFE). The HFE creates a single feature vector by combining the Histogram-Oriented Gradient (HOG), Local Binary Pattern (LBP), and Speed Up Robust Feature (SURF) extraction techniques. The CNN approach also pulls features from training and test datasets. The classification model is built by fusing the two-feature vectors that are produced. Using academic torrents and ISIC 2017 datasets, the suggested technique is tested and shows impressive results with 99.85% accuracy, 91.65% sensitivity, and 95.70% specificity [24].

Using different of CNNs variants and public common datasets named as ISIC 2019 and 2020, the researchers in [25] rigorously preprocessed the data using advanced Generative of AI techniques like: GANs and ESRGAN, they were used for augmentation. The results demonstrated the potential of using generative AI to augment data volume and confirmed the effectiveness of transfer learning models based on convolutional neural networks in improving the accuracy of skin cancer classification. A thorough evaluation and comparison of various generative techniques' efficacy is conducted. Several transfer learning models are used in the CNN-based technique, such as VGG16, VGG19, SVM, and a hybrid model that joined VGG19 and SVM. The hybrid VGG19+SVM model performs exceptionally well with a 96% accuracy on ISIC 2019, whereas VGG16 and VGG19 obtain promising accuracies of 92% and 93%, respectively while VGG16, VGG19, and SVM show accuracies of 90%, 92%, and 92% on ISIC 2020, respectively,[25].

In addition, the researchers in [36] conducted a study to evaluate number of deep learning models for the diagnosis of skin cancer, which includes: DenseNet-201, MobileNet-V2, ResNet-50V2 [54], ResNet-152V2, X-ception, VGG-16, VGG-19, and GoogleNet. In their study, they trained all models on dataset contains 7164 images from the dataset named: ISIC 2019. Their results showed that GoogleNetachieved an accuracy of 76.09%, which is considered a quite low accuracy, therefore, their proposal model was classified only for the binary classification case. This means that they achieved less in the model: Vgg19 and ALexnet that we used as well in our study; this proves that we can achieve higher in our work

The researchers in [37]evaluated deep learning-based approach for skin cancer detectionusing: (GoogLeNet, AlexNet, ResNet, and VGGNet). The classification was designed based on three categories (melanoma, nevus, and seborrheic keratosis). They suggested to use an ensemble of deep CNNs, where built their model to classify the data based on vote outputs of multiple CNNs. The results show that the ensemble approach outperforms individual CNNs, where the accuracy they achieved was 79.9.

Table 1 bellow display some other researchers on skin cancer detection that contains number of CNNs model including the model that we used in our model. Our work differs from their works that published in the most recent in 2021 and 2023.

Table 1: other researchers on skin cancer detection

Article	*Year*	*Field*
[38]	2019	Reviewed DL models for detecting the skin cancer
[39]	2021	Reviewed DL based decision support in skin cancer detection
[40]	2021	Reviewed DL techniques for the analysis of skin cancer and melanoma cancer diagnosis
[41]	2021	Reviewed DL algorithms used for detecting the skin cancer
[42]	2023	Reviewed for classification of skin cancer and segmentation based on using (GANs)

3 Methodologies

This section presents the two subsections. 1) The description of dataset and Data Augmentation. 2) The experimental Process.

3.1 The Description of Dataset- and Data Augmentation

It is common knowledge that pre-trained CNN models will perform better when trained with a wider set of images but this is rarely available. In such cases, data augmentation techniques are deployed to expand the training set beyond the original size and composition providing the model to be more inclusive and able to fit to more variations of data. This even becomes more important in the scopes of medical imaging. This technique of data augmentation is standard practice to introduce some changes in the original images to create more images of brain tumors. In particular, augmentation influencing training images to increase their number and allow building a model which generalizes better to varying conditions. A total of 50%, 70%, and 80% of tumor images were multiplied, thus increasing the size of every one of them. However, no augmentation was applied to the images used in testing and validation from the dataset.

In this research, Figure 1 outlines the data augmentation techniques applied in this research including random horizontal reflection.

The skin cancer dataset is downloaded from [33]. The dataset contains a balanced dataset of images of benign skin moles and malignant skin moles. It consists of two folders with 1800 pictures (224x224) of the two types of moles.

3.2 The Experimental Process

The study is designed on three different scenarios of datasets sizes. Which includes 80% and 20%, 70% and 30%, and 50% and 50% for training and test datasets respectively. In each scenario – the five steps of the proposed model are implemented (see figure 2). Thus, the proposed model will be implemented by the number of three different scenarios. Then, the results were presented and evaluated for each scenario. Next, are the steps for the proposed model.

Step 1) in this step, we followed set of preprocessing for preparing the images for next steps, as follows:

1. Collect the dataset from the internet. One class (i.e. folder) contains two subfolders for each of benign and malignant.
2. The images were augmented randomly to increase the dataset size. This to produce a diversity for the training images.
3. The images were converted from grayscale to RGB format. This to be the images are compatible with the CNN models.
4. The input images were resized into (224,224) to be all the images are in the same dimension.
5. Images are shuffled randomly before different scenarios of (80%, 70%, and 50%) to prevent any bias in the selection. Then, the prepared images now are entered into each model of this study (VGG-19, AlexNet).

Step 2) The feature vectors will be extracted (that includes fc6 and fc7-as applied in the study [32]) from images automatically using MATLAB for Pre-trained VGG-19 (i.e. step 2.1), and Alexnet (i.e. step 2.2) to build datasets. Any dataset used in the study (e.g. median, fc6) has 4096 columns that represents feature vectors and number of rows that represents the number of images. Then to be used in step 4.

The layers (fc6 and fc7) were selected in our study because of the ability of these layers to capture high-level semantic information relevant to a wide range of computer vision tasks, especially when utilizing pre-trained models for transfer learning [28] [29].

(a) Original Skin Image (b) Random X Reflection Image

Figure 1. Data augmentation for Random X Reflection

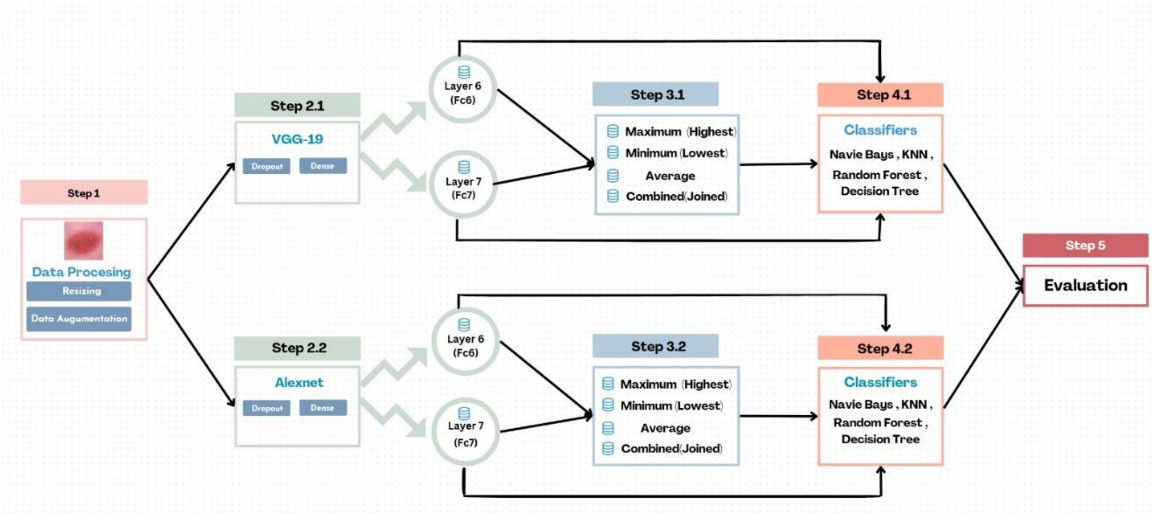

Figure 2. The proposed Model

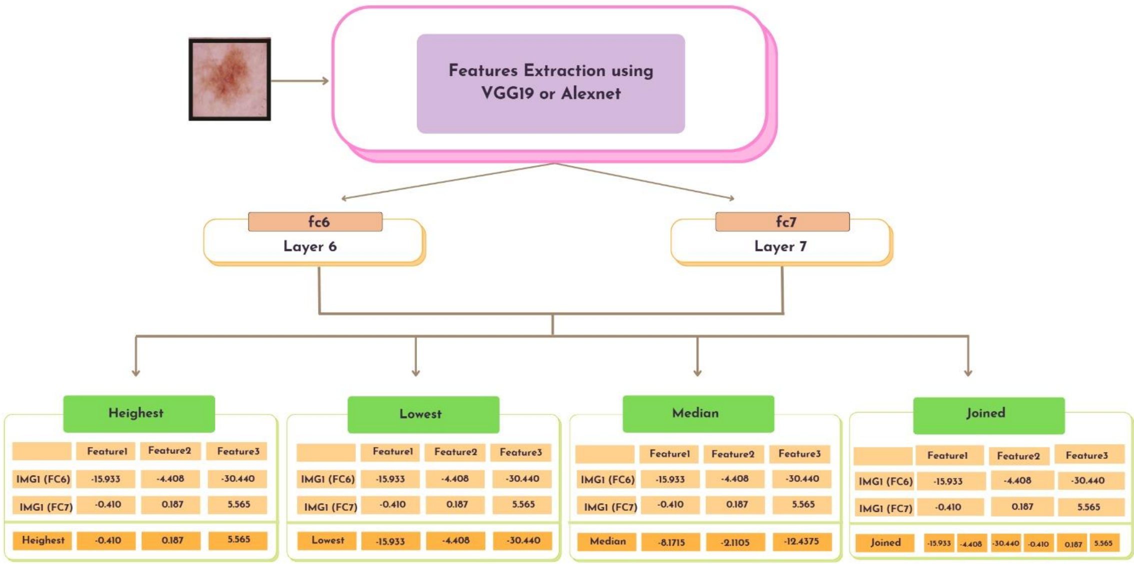

Figure 3. shows how these statistical operations are design

Step 3) The statistical operations. New datasets (i.e., lowest, highest, median, and joined them together) will be created from the two layers (i.e., fc6 and fc7). It includes step 3.1 for datasets extracted from vgg-19, and step 3.2 for datasets extracted from Alexnet. We suggest using the statical operations to investigate their influence on classification accuracy; this explains their role and importance in the study. These datasets will be classified in step 4. Figure 3 shows how these statistical operations are design. Next, the explanation for these statistical operations.

- Highest (fc6 and fc7): To calculate the large value of the two values in the two datasets (fc6 - fc7).
- Lowest (fc6 and fc7): To calculate the less value of the two values in the two datasets (fc6 - fc7).
- Median (fc6 and fc7): To calculate the median of the two values in the two datasets (fc6 - fc7).

- Joined them together (fc6 and fc7): The dataset fc6 (4096) has been joined next to the dataset fc7 (4096), then the joined dataset will contain 8192 features.

Step 4) The results will be achieved in step 4.1 (i.e. vgg-19) and step 4.2 (i.e. Alexnet) using set of ML classifiers on the datasets gained from second step and third step.

Step 5) The results will be assessed using various of performance measures that include Accuracy, precision, Recall and finally F- Measure.

The values of these measures take a range of [0-1], where 0 represents the worst performance and 1 represents the best performance. They are explained as follows:

Precision: This metric, which provides the precise percentage of positive examples compared to the entire number, aids in evaluating the accuracy of the classifiers' output, as shown in equation (1):

$$Precision = \frac{Tp}{TP + FP} \qquad (1)$$

Recall: This metric, which represents the proportion of positive examples displayed on the total number, is employed in the finalization test of the classifier's results, as shown in equation (2).

$$Recall = \frac{Tp}{TP + FN} \qquad (2)$$

F-Measure: The F-Measure is located in mid-precision and recalls an important focal point in the middle that aims to increase system reliability by accuracy and recall and determine which produces more efficient results, as shown in equation (3).

$$F\text{-}Measure = \frac{2*TP \quad 2*Tp}{2*Tp+FP+FN2*FP+FN} \qquad (3)$$

Accuracy: The most commonly used metric for assessing the classification process's accuracy is accuracy, which is typically calculated as the proportion of correctly categorized examples to all examples.

$$Accuracy = \frac{TP+TN}{Tp+FP+TN+FN} \qquad (4)$$

Where TP, TN, FP, and FN indicate to true positive class, true negative class, false positive, and false negative class, respectively.

4 Experimental Results and Discussions

The outcomes will be assessed using the measures. 1) Accuracy, 2) Recall, 3) F-measure, 4) Precision, 5) Required Time for training the model of each classifier per seconds. The following are classifiers considered in this study. 1) KNN, 2) Random Forest, 3) Naïve Bayes, 4) Decision Tree.

The following sub-sections presents the results of each scenario.

4.1 First Scenario

This scenario is built based on the split ratio oftraining-testing of 80%20%. The aim for this scenario is to evaluate the effect of training dataset size on the classification accuracy. Next, the results and the discussion for both original fc6 and original fc7 feature vector datasets separately. Then, the results for statistical operations: Median, highest, lowest, and joined of fc6 and fc7.

4.1.1 Results of both original fc6 and original fc7 dataset.

Table (1) shows that Random Forest (RF) has scored the highest accuracy results in bothVGG-19 and Alexnet for fc6 and fc7.

In the dataset FC6, RF has scored accuracy of 83.4 in Vgg-19 and 84.2 in Alexnet. While in FC7: RF has scored accuracy of 82.2 in Vgg-19 and 83.6 in Alexnet. The second-best accuracy for fc6 is KNN that perform 81.4 in VGG-19 and 79.4 in Alexnet. While the second-best accuracy for fc7 is Naive Bayes that perform 79.4 in VGG-19 and 80.9 in Alexnet. This study matches with the study [30] for having high classification accuracy for (RF) using Vgg-19 and Alexnet. Their study conducted based on one scenario. The results show that Alexnet performs high accuracy and consumes less time that required for training the model compared to vgg-19 [31]

In required time that needed for training the model, Decision Tree spend the longest time that required for training the model, while KNNs spend less time compared to other classifiers because KNN does not have a training model. It matches the test example directly with other examples in the training set. This explains why KNN is slow in testing when there are large number of examples in the training set. [32].

4.1.2 Results of median feature vector (fc6⨁fc7) datasets

Table (2) shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for median features vector. RF has scored accuracy of 84.2 in VGG-19 and 83.2 in Alexnet. The second-best accuracy in VGG-19 is KNN that perform 81.4 while the second-best accuracy in Alexnet is Decision Tree that performs 80.3. Decision Tree spend the longest time that required for training the model. This approves the results in section 1 of first scenario in high accuracy for RF classifier.

4.1.3 Results of highest feature vector (fc6⨁fc7) datasets

Table (2) shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for highest features vectors. RF has scored accuracy of 81.5 in VGG-19 and 83.1 in Alexnet. The second-best accuracy in VGG-19 is KNN that performs 79.1 while the second-best accuracy in Alexnet is Naive Bayes that performs 83.2. Decision Tree spends the longest time that required for training the model. The results of this section approves the results in above sections 1 and 2 of first scenario in high accuracy for RF classifier.

4.1.4 Results of lowest feature vector (fc6⨁fc7) datasets

Table (2) shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for lowest features vectors. RF has scored accuracy of 82.9 in VGG-19 and83.6 in alexnet. The second-best accuracy in VGG-19 and Alexnet is KNN that perform 81.6 in VGG-19 and 79.5 in Alexnet. Decision Tree spends the longest time that required for training the model, while Naive Bayes spend less time.

4.1.5 Results of joined features of fc6 and fc7.

Table (3) shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for joined features vectors. RF has scored accuracy of 82.6 in VGG-19 and 83.9 in Alexnet. The second-best accuracy in VGG-19 is KNN that perform

81.33 while the second-best accuracy in Alexnet is Naive Bayes that performs 81.1. Decision Tree spend the longest time that required for training the model.

In Summary for the results in first scenario, all results of all datasets were slightly similar. This approves that the extracted features and the statistical operations have an influence on the classification accuracy. The results show that Alexnet performs high accuracy and consumes less time that required for training the model compared with vgg-19 [31].

4.2 Second Scenario

This Scenario is built based on the split ratio of training-testing of 70%30%. The aim for thisscenario is to evaluate the effect of training dataset size on the classification accuracy. Next, the results and the discussion for both original fc6 and original fc7 feature vector datasets separately. Then, the results for statistical operations: Median, highest, lowest, and Joined of fc6 and fc7.

4.2.1 Results of both original fc6 and original fc7 dataset.

Table 2 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for fc6 and fc7.

For FC6, RF has scored accuracy of 84.32 in Vgg-19 and 84.2 in Alexnet. While for FC7: RF has scored accuracy of 83.9 in Vgg-19 and 83.6 in Alexnet. The second-best accuracy for fc6 is KNN that perform 81.4 in VGG-19 and 79.4 in Alexnet. While The second-best accuracy for fc7 is Naive Bayes that perform 79.4 in VGG-19 and 80.9 in Alexnet. This study matches with the study [Gairola et al., 2022b] for having high classification accuracy for (RF) using Vgg-19 and Alexnet. Their study conducted based on one scenario.

In required time for training the model, Decision Tree spend the longest time that required to train the model, while KNNs spend less time that required to train the model compared with other classifiers because KNN does not have a training model. It matches the test example directly with other examples in the training set. This explains why KNN is slow in testing when there are large number of examples in the training set [32].

4.2.2 Results of median feature vector (fc6⨁fc7) datasets

Table 3 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for median features vectors. RF has scored accuracy of 84.3 in VGG-19 and 83.0 in Alexnet. The second-best accuracy in VGG-19 is KNN that perform 80.8 while the second-best accuracy in Alexnet is Decision Tree that performs 81.7. Decision Tree spend the longest time that required for training the model. This approves the results in section 1 of first and second scenarios in high accuracy for RF classifier.

4.2.3 Results of highest feature vector (fc6⨁fc7) datasets

Table 3 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for Highest features vectors. RF has scored accuracy of 82.7 in VGG-19 and 84.5 in Alexnet. The second-best accuracy is KNN for VGG-19 and Alexnet is that perform 79.7 and 81.5 respectively. Decision Tree spend the longest time that required for training the model. The results of this section approves the results in above sections 1 and 2 of first and second scenarios in high accuracy for RF classifier.

4.2.4 Results of Lowest feature vector (fc6⨁fc7) datasets

Table 3 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for Lowest features vectors. RF has scored accuracy of 83.1 in VGG-19 and 83.5 in Alexnet. The second-best accuracy in VGG-19 and Alexnet is KNN that perform 79.9 in VGG-19 and 81.3 in Alexnet. Decision Tree spend the longest time that required for training the model, while Naive Bayes spend less time.

4.2.5 Results of joined features of fc6 and fc7.

Table 4 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for joined features vector. RF has scored accuracy of 83.1 in VGG-19 and 83.9 in Alexnet. The second-best accuracy in VGG-19 and Alexnet is KNN that perform 80.7 and 81.1 respectively. Decision Tree spend the longest time that required for training the model.

In Summary for the results in second scenario, all results of all datasets were slightly similar. This approves that the extracted features and the statistical operations have an influence on the classification accuracy. The results show that Alexnet performs high accuracy and consumes less time that required for training the model compared to vgg-19 [31].

Table 2: Result of fc6 and fc7 feature vector

	Algorithms	Accuracy	Precision	Recall	F-measure	Time (s)	Accuracy	Precision	Recall	F-measure	Time (s)	Accuracy	Precision	Recall	F-measure	Time (s)
		FC6 (First Scenario) 80-20					***FC6 (Second Scenario) 70-30***					***FC6 (Third Scenario) 50-50***				
Vgg19	**KNN**	81.84	0.819	0.818	0.818	0	81.29	0.813	0.813	0.812	0	81.79	0.819	0.818	0.817	0
	NB	73.73	0.739	0.737	0.734	0.94	75.93	0.760	0.759	0.757	0.28	75.84	0.759	0.758	0.757	1.85
	RF	83.39	0.836	0.834	0.834	2.86	84.32	0.845	0.843	0.844	0.93	84.22	0.843	0.842	0.842	5.75
	DT	76.95	0.769	0.770	0.769	10.45	74.72	0.747	0.747	0.747	2.15	77.30	0.773	0.773	0.773	18.15
Alexnet	**KNN**	79.36	0.796	0.794	0.792	0.01	81.79	0.819	0.818	0.817	0	83.85	0.840	0.839	0.838	0
	NB	78.60	0.787	0.786	0.786	0.19	76.33	0.764	0.763	0.764	0.9	76.94	0.770	0.769	0.770	0.51
	RF	84.21	0.845	0.842	0.842	0.87	84.22	0.844	0.842	0.843	3.08	85.61	0.859	0.85.6	0.856	1.51
	DT	78.30	0.783	0.783	0.783	1.56	79.97	0.800	0.800	0.800	5.89	77.85	0.778	0.779	0.778	3.92
		FC7 (first Scenario) 80-20					***FC7 (Second Scenario) 70-30***					***FC7 (Third Scenario) 50-50***				
Vgg19	**KNN**	78.90	0.791	0.789	0.787	0	79.87	0.799	0.799	0.798	0	80.76	0.808	0.808	0.807	0.01
	NB	79.36	0.793	0.794	0.793	0.24	77.14	0.771	0.771	0.771	0.28	78.51	0.785	0.785	0.785	1.77
	RF	82.24	0.823	0.822	0.823	0.69	83.92	0.840	0.839	0.839	0.94	84.16	0.843	0.842	0.842	5.47
	DT	74.35	0.743	0.744	0.743	1.39	76.5	0.765	0.765	0.765	2.42	77.06	0.770	0.771	0.771	17.6
Alexnet	**KNN**	78.45	0.786	0.785	0.783	0.01	80.88	0.812	0.809	0.807	0	82.58	0.827	0.826	0.825	0
	NB	80.88	0.813	0.809	0.809	0.35	79.97	0.806	0.800	0.800	0.98	81.00	0.814	0.810	0.810	0.51
	RF	83.61	0.837	0.836	0.836	0.79	83.61	0.839	0.836	0.837	3.13	85.25	0.855	0.853	0.853	1.68
	DT	76.32	0.763	0.763	0.763	1.42	77.95	0.779	0.780	0.779	6.55	77.66	0.777	0.777	0.777	4.3

Table 3.Results of median, Highest, and Lowest feature vector (fc6⊕fc7)

	Algorithms	Accuracy	Precision	Recall	F-measure	Time (s)	Accuracy	Precision	Recall	F-measure	Time (s)	Accuracy	Precision	Recall	F-measure	Time (s)
		Median (First Scenario) 80-20					***Median (Second Scenario) 70-30***					***Median (Third Scenario) 50-50***				
Vgg19	**KNN**	81.38	0.814	0.814	0.813	0	80.78	0.808	0.808	0.807	0	81.43	0.814	0.814	0.814	0
	NB	73.38	0.735	0.734	0.731	1.49	79.47	0.796	0.795	0.795	0.83	75.72	0.758	0.757	0.755	0.89
	RF	84.19	0.844	0.842	0.842	3.47	84.32	0.845	0.843	0.844	0.99	84.28	0.845	0.843	0.843	5.99
	DT	75.96	0.759	0.760	0.759	14.53	77.14	0.771	0.771	0.771	2.24	76.75	0.768	0.768	0.768	18.24
Alexnet	**KNN**	79.05	0.792	0.791	0.789	0	81.69	0.818	0.817	0.816	0	84.58	0.848	0.846	0.845	0.01
	NB	78.14	0.782	0.781	0.782	0.26	76.23	0.763	0.762	0.762	1.09	77.06	0.771	0.771	0.771	0.57
	RF	83.15	0.833	0.832	0.832	0.81	84.0	0.840	0.837	0.838	3.35	85.55	0.858	0.856	0.856	1.72
	DT	80.27	0.803	0.803	0.803	1.69	78.15	0.782	0.782	0.782	7.8	77.97	0.780	0.780	0.780	4.13
		Highest (first Scenario) 80-20					***Highest (Second Scenario) 70-30***					***Highest (Third Scenario) 50-50***				
Vgg19	**KNN**	79.05	0.791	0.791	0.790	0	79.67	0.797	0.797	0.796	0	80.64	0.808	0.806	0.805	0
	NB	78.75	0.787	0.788	0.787	0.21	78.05	0.780	0.781	0.780	0.28	78.64	0.786	0.786	0.786	1.9
	RF	81.48	0.815	0.815	0.815	0.93	82.70	0.828	0.827	0.827	0.96	83.61	0.838	0.836	0.836	5.54
	DT	74.96	0.749	0.750	0.749	1.28	75.32	0.753	0.753	0.753	1.94	79.55	0.796	0.796	0.796	15.59
Alexnet	**KNN**	79.21	0.795	0.792	0.790	0	81.49	0.819	0.815	0.813	0	83.19	0.835	0.832	0.830	0
	NB	81.63	0.820	0.816	0.817	0.2	81.19	0.817	0.812	0.812	0.91	82.34	0.826	0.823	0.824	0.52
	RF	83.15	0.833	0.832	0.832	0.79	84.52	0.849	0.845	0.846	3.31	85.49	0.858	0.855	0.855	1.63
	DT	74.35	0.744	0.744	0.744	1.42	76.64	0.767	0.766	0.767	7.62	78.88	0.789	0.789	0.789	4.16
		Lowest (first Scenario) 80-20					***Lowest (Second Scenario) 70-30***					***Lowest (Third Scenario) 50-50***				
Vgg19	**KNN**	81.63	0.818	0.816	0.815	0	79.97	0.800	0.800	0.799	0	81.25	0.813	0.813	0.812	0
	NB	73.44	0.741	0.734	0.728	0.18	73.20	0.737	0.732	0.727	0.28	74.27	0.748	0.743	0.738	1.96
	RF	82.85	0.829	0.829	0.829	0.67	83.11	0.833	0.831	0.831	0.99	83.79	0.839	0.838	0.838	5.7
	DT	74.50	0.745	0.745	0.745	1.28	74.41	0.744	0.744	0.743	2.26	76.15	0.762	0.762	0.762	17.99
Alexnet	**KNN**	79.51	0.797	0.795	0.794	0	81.29	0.814	0.813	0.812	0	83.98	0.842	0.840	0.839	0
	NB	76.63	0.766	0.766	0.766	0.21	75.6	0.756	0.756	0.756	0.94	76.33	0.763	0.763	0.763	0.53
	RF	83.61	0.839	0.836	0.836	0.76	83.51	0.838	0.835	0.836	3.08	85.19	0.855	0.852	0.852	1.59
	DT	76.78	0.768	0.768	0.768	1.43	78.26	0.782	0.783	0.783	6.86	77.73	0.777	0.777	0.777	3.88

Table 4: Results of Joined feature vector (fc6⊕fc7)

	Algorithms	Accuracy	Precision	Recall	F-measure	Time (s)	Accuracy	Precision	Recall	F-measure	Time (s)	Accuracy	Precision	Recall	F-measure	Time (s)
		Joined (First Scenario) 80-20					***Joined (Second Scenario) 70-30***					***Joined (Third Scenario) 50-50***				
Vgg19	**KNN**	81.33	0.814	0.813	0.813	0	80.78	0.808	0.808	0.807	0	81.55	0.817	0.816	0.814	0.02
	NB	77.99	0.780	0.780	0.779	0.43	77.04	0.771	0.770	0.769	0.62	77.79	0.778	0.778	0.777	3.97
	RF	82.54	0.826	0.825	0.826	1.13	83.11	0.832	0.831	0.831	1.47	83.73	0.839	0.837	0.838	9.22
	DT	73.29	0.732	0.733	0.732	3.09	72.90	0.728	0.729	0.728	4.13	78.27	0.782	0.783	0.783	37.82
Alexnet	**KNN**	80.88	0.810	0.809	0.807	0	81.19	0.814	0.812	0.810	0.02	83.79	0.840	0.838	0.837	0
	NB	81.18	0.814	0.812	0.812	0.42	79.37	0.796	0.794	0.794	1.94	80.15	0.803	0.802	0.802	1.08
	RF	83.91	0.842	0.839	0.839	1.11	83.92	0.841	0.839	0.839	5.56	85.55	0.858	0.856	0.856	2.35
	DT	79.8179	0.798	0.798	0.798	2.42	78.5642	0.786	0.786	0.786	14.1	77.12	0.772	0.771	0.771	9.24

Table 5: Comparative Analysis of Related Studies

	Method	Dataset (name)	Result
[26]	CNN based on Random Forest.	**Skin Cancer: Malignant vs. Benign**	Accuracy is 85%
[30]	DL algorithms: (Alex-Net, Googlenet, Vgg16, Vgg19, ResNet18, ResNet50, ResNet101, DenseNet121 and DenseNet161.0) Classifiers: (Decision Tree, KNN, Logistic Regression, Naive Bayes, Random Forest, and SVM)	kin Lesion Analysis: Towards Melanoma Detection	The highest accuracy is for Alex-net using Random Forest (82.3%)
[27]	CNN (VGG-16) and Random forest	ISIC 2020 Challenge Dataset	Accuracy is 96%
Proposed model	Deep learning algorithms with CNN models (Alex-Net, and VGG-19) Classifiers (Decision Tree, KNN, Naive Bayes and Random Forest.	[33]	The Highest Accuracy is for Joined, Median, and (fc6+fc7) datasets in Alexnet Using Random Forest (85.5%)

4.3 Third Scenario

This Scenario is built based on the split ratio of training-testing of 50%50%. The aim for this scenario is to evaluate the effect of training dataset size on the classification accuracy. Next, the results and the discussion for both original fc6 and original fc7 feature vector datasets separately. Then, the results for statistical operations: Median, highest, lowest, and Joined of fc6 and fc7.

4.3.1 Results of both original fc6 and original fc7 datasets.

Table 2 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for fc6 and fc7.

For FC6, RF has scored accuracy of 84.2 in Vgg-19 and 85.6 in Alexnet. While for FC7: it has scored accuracy of 84.1 in Vgg-19 and 85.3 in Alexnet. The second-best accuracy for fc6 is KNN that perform 81.8 in VGG-19 and 83.9 in Alexnet. While The second-best accuracy for fc7 is KNN that perform 80.8 in VGG-19 and 82.6 in Alexnet.

In the required time that needed for training the model, Decision Tree spend the longest time that required for training the model, while KNNs spend less time compared to other classifiers because KNN does not have a training model. It matches the test example directly with other examples in the training set. This explains why KNN is slow in testing when there are large number of examples in the training set [32]. This study matches with the study [30] for having high classification accuracy for (RF) using Vgg-19 and Alexnet. Their study conducted based on one scenario.

4.3.2 Results of median feature vector (fc6⨁fc7) datasets

Table 3 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for median features vectors. RF has scored accuracy of 84.3 in VGG-19 and 85.6 in alexnet. The second-best accuracy in VGG-19 and alexnet is KNN that perform 81.4 and 84.6 respectively. Decision Tree spend the longest time that required for training the model and KNN is the shortest time. This approves the results in section 1 of third scenario in high accuracy for RF classifier.

4.3.3 Results of highest feature vector (fc6⨁fc7) datasets

Table 3 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for Highest features vectors. RF has scored accuracy of 83.6 in VGG-19 and85.5 in Alexnet. The second-best accuracy is KNN for VGG-19 and Alexnet is that perform 80.6 and 83.2 respectively. Decision Tree spend the longest time that required for training the model. The results of this section approves the results in above sections 1 and 2 of third scenario in high accuracy for RF classifier.

4.3.4 Results of Lowest feature vector (fc6⨁fc7) datasets

Table 3 shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for Lowest features vectors. RF has scored accuracy of 83.8 in VGG-19 and 85.2 in Alexnet. The second-best accuracy in VGG-19 and Alexnet is KNN that perform 81.3 in VGG-19 and 84.0 in Alexnet. Decision Tree spend the longest time that required for training the model, while KNN spend less time.

4.3.5 Results of joined features of fc6 and fc7.

Table (3) shows that Random Forest has scored the highest accuracy results in both VGG-19 and Alexnet for joined features vectors. It has scored accuracy of 83.7 in VGG-19 and 85.6 in Alexnet. The second-best accuracy in VGG-19 and Alexnet is KNN that perform 81.6 and 83.8 respectively. Decision Tree spend the longest time that required for training the model.

In Summary for the results in third scenario, all results of all datasets were slightly similar. This approves that the extracted features and the statistical operations have an influence on the classification accuracy. The results show that Alexnet performs high accuracy and consumes less required time for training the model compared to vgg-19 [31].

The Results in the three scenarios were slightly similar this approves that the study can provide high accuracy even the size of the training dataset was minimal.

As comparison between the results of the proposed approach with three similar approaches for skin cancer detection is illustrated in Table 5.

It has approved that the output of this proposal that can be considered one of the interesting study compared to the previous researches, for several reasons:

1. Some of previous studies were achieved on small dataset skin cancer detection compared to this proposed model.
2. Number of previous studies were achieved on large training dataset size in their skin cancer studies compared to this stud. This study is conducted on different scenarios of training datasetssizes. Few number of images usually leads to low accuracy compared to the large examples, but in constant it was not in this proposed model.
3. The three scenarios used in this study have proved that the system can achieve high classification accuracy even the dataset sizes is small.

6 Conclusion

The purpose of this study is to investigate how deep learning models may be used to classify skin cancer images if they are benign or malignant. The features were extracted using: VGG-19 and Alexnet. These features (fc6 and fc7) were used to generate statistical operations (e.g. highest, lowest, joined, etc). Then, all features and the generated ones were classified based on ML algorithms. The experimental study consists of 36 experiments: 2 pre-trained approaches (Vgg-19 and Alexnet) * 6 datasets * 3 scenarios datasets. The evaluation was made based on the results of the classification accuracy.

Based on the investigation that illustrated in Tables above (from table 1 to 3); it is important to make the following claims, Random Forest (85.6) is the optimal classifier for deep features extraction of skin cancer detection across all datasets in all scenarios – No matter the size of the datasets. While Decision Tree has almost scored less accuracy. In addition, KNN required less time that required for training the model compared to other algorithms. This can be explained by the fact that KNN does not involve training a model; the test instance is directly matched with the training instances, which accounts for the reduced testing time. Also, KNN (84.5) is estimated as the second-best of classifiers after Random Forest in the most Scenarios, this for all operations i.e. (FC-layer-6, FC-layer-7, Median, Highest, Lowest, and Joined). In general, there were no distinct results based on features between Layer 6 and Layer 7 (see Table (1)), or any other dataset generated from them. Deep learning (VGG-19 and Alexnet) provides features that can enhance detection systems such as the skin cancer detection system. The order of machine learning algorithms affects accuracy from highest to lowest, as follows: Random Forest was first, followed by KNN, then Decision Tree, and finally Naïve Bayes.

The results of all datasets were slightly similar. This approves that the extracted features and the statistical operations have an influence on the classification accuracy. The results show that Alexnet performs high accuracy and consumes less time that required for training the model- compared to Vgg-19.

Future work will address several issues, the most important of which are:

- Attempt actually to get better accuracy results and then also try to train it faster.

- Trying to improve the collected database by obtaining clearer and less noise images in addition to increasing the number of images.
- Higher number of Deep Learning Networks will be investigated such as NasNet-Mobile (S. Addagarla et al. 2020), GoogleNet(Szegedy et al. 2014), etc.
- Focus on the techniques of Dimensionality of the deep features reduction such as PCA, DCT, wavelet transform, etc.

References

[1]. The Skin Cancer Foundation. (2023, January). Skin cancer facts & statistics. Retrieved from https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/

[2]. American Cancer Society. (n.d.). Skin cancer facts & common skin cancer types. Retrieved from https://www.cancer.org/cancer/types/skin-cancer.html

[3]. Akter, M., Hossain, M. S., Ahmed, T. U., & Andersson, K. (2020). Mosquito classification using convolutional neural network with data augmentation. In *3rd International Conference on Intelligent Computing & Optimization (ICO 2020)*.

[4]. American Cancer Society. (2022, June 7). Key statistics for melanoma skin cancer. Retrieved from https://www.cancer.org/content/dam/CRC/PDF/Public/8823.00.pdf

[5]. The Skin Cancer Foundation. (2024, February). Skin cancer facts & statistics. Retrieved from https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/

[6]. Kasinathan, G., Jayakumar, S., Gandomi, A. H., Ramachandran, M., Fong, S. J., & Patan, R. (2019). Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier. *Expert Systems with Applications*, *134*, 112–119.

[7]. Gao, Z., Wang, L., Zhou, L., & Zhang, J. (2016). Hep-2 cell image classification with deep convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics*, *21*(2), 416–428.

[8]. Wang, P., Wang, L., Li, Y., Song, Q., Lv, S., & Hu, X. (2019). Automatic cell nuclei segmentation and classification of cervical pap smear images. *Biomedical Signal Processing and Control*, *48*, 93–103.

[9]. Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health*, *18*(10), 5479. https://doi.org/10.3390/ijerph18105479

[10]. Sreena, S., & Lijiya, A. (2019). Skin lesion analysis towards melanoma detection. In *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)* (Vol. 1, pp. 32–36). IEEE.

[11]. Rashid, H., Tanveer, M. A., & Aqeel Khan, H. (2019). Skin lesion classification using GAN-based data augmentation. In *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 916–919). Berlin, Germany.

[12]. Bisla, D., Choromanska, A., Stein, J. A., Polsky, D., & Berman, R. (2019). Towards automated melanoma detection with deep learning: Data purification and augmentation. *arXiv*. Retrieved from https://arxiv.org/abs/1902.03280

[13]. Rutjes, A., Reitsma, J., Coomarasamy, A., Khan, K., & Bossuyt, P. (2007). Evaluation of diagnostic tests when there is no gold standard: A review of methods. *Health Technology Assessment*, *11*(50). https://doi.org/10.3310/hta11500

[14]. Hosny, K. M., Kassem, M. A., & Foaud, M. M. (2019). Classification of skin lesions using transfer learning and augmentation with AlexNet. *PLOS ONE*, *14*(5), e0217293.

[15]. He, X., Yu, Z., Wang, T., Lei, B., & Shi, Y. (2018). Dense deconvolution net: Multi-path fusion and dense deconvolution for high resolution skin lesion segmentation. *Technology and Health Care*, *26*(S1), 307–316.

[16]. Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of Biomedical Informatics*, *86*, 25–32.
[17]. Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., Schilling, B., Haferkamp, S., Schadendorf, D., & Holland-Letz, T. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer, 113*, 47–54.
[18]. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118.
[19]. Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., & others. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 168–172). IEEE.
[20]. Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P.-A. (2016). Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Transactions on Medical Imaging*, *36*(4), 994–1004.
[21]. Himanshu, K., & others. (2023). A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. *Biomedical Signal Processing and Control, 79*(2), 104186. https://doi.org/10.1016/j.bspc.2022.104186
[22]. Faghihi, A., & others. (2023). Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-17735-2
[23]. Singh, A., Singh, N. P., & Rathee, P. (2023). Skin cancer detection using deep learning approach. In T. Senjyu, C. So-In, & A. Joshi (Eds.), *Smart Trends in Computing and Communications. SmartCom 2023* (pp. 1–16). Springer. https://doi.org/10.1007/978-981-99-0838-7_45
[24]. Rahman, M. M., & others. (2023). Proposing a hybrid technique of feature fusion and convolutional neural network for melanoma skin cancer detection. *Journal of Pathology Informatics*, *14*, 100341. https://doi.org/10.1016/j.jpi.2023.100341
[25]. Saeed, M., Naseer, A., Masood, H., Rehman, S. U., & Gruhn, V. (2023). The power of generative AI to augment for enhanced skin cancer classification: A deep learning approach. *IEEE Access*, *11*, 130330–130344. https://doi.org/10.1109/ACCESS.2023.3332628
[26]. Khasanah, N., Komarudin, R., Afni, N., Maulana, Y. I., & Salim, A. (2021). Skin cancer classification using random forest algorithm. *SISFOTENIKA*.
[27]. V, N., S, S., B, S., & Kumar, C. R. (2023). Early detection of melanoma using convolutional neural network and random forest algorithm. *Journal of Medical Imaging and Health Informatics, 13*(3), 673–679.
[28]. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556
[29]. Kaushik, A. (2020, February 26). Understanding the VGG19 architecture. *OpenGenus IQ: Computing Expertise & Legacy*. Retrieved from https://iq.opengenus.org/vgg19-architecture/
[30]. Gairola, A. K., Kumar, V., & Sahoo, A. K. (2022). Exploring the strengths of pre-trained CNN models with machine learning techniques for skin cancer diagnosis. In *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)* (pp. 1–6). https://doi.org/10.1109/MysuruCon55714.2022.9974465
[31]. Abbasi, A., Yu, D., & Shrestha, P. (2023). A hybrid deep learning approach for skin cancer detection using transfer learning and ensemble methods. *Pattern Recognition Letters, 174*, 56–63.

[32]. Saha, S., Khatun, F., & Kim, Y. (2022). Skin cancer detection using hybrid deep convolutional neural network model. *Journal of Electrical Engineering & Technology, 17*(6), 1845–1854. https://doi.org/10.1007/s42835-022-00988-4

[33]. Fancon, C. (2019). Skin cancer: Malignant vs. benign. Kaggle. https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign

[34]. Addagarla, S., Chakravarthi, G., & Anitha, P. (2020). Real-time multi-scale facial mask detection and classification using deep transfer learning techniques. *International Journal of Advanced Trends in Computer Science and Engineering, 9*, 4402–4408.

[35]. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–9). https://doi.org/10.1109/CVPR.2015.7298594

[36]. Aljohani, K., & Turki, T. (2022). Automatic classification of melanoma skin cancer with deep convolutional neural networks. *AI, 3*, 512–525. https://doi.org/10.3390/ai3020029

[37]. Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of Biomedical Informatics, 86*, 25–32. https://doi.org/10.1016/j.jbi.2018.08.006

[38]. Pacheco, A. G., & Krohling, R. A. (2019). Recent advances in deep learning applied to skin cancer detection. *arXiv*. https://doi.org/10.31219/osf.io/epvdu

[39]. Lucieri, A., Dengel, A., & Ahmed, S. (2021). Deep learning-based decision support for medicine—A case study on skin cancer diagnosis. *arXiv*. https://doi.org/10.31219/osf.io/ey8fq

[40]. Adegun, A., & Viriri, S. (2021). Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. *Artificial Intelligence Review, 54*(1), 811–841. https://doi.org/10.1007/s10462-020-09865-y

[41]. Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health, 18*(10), 5479. https://doi.org/10.3390/ijerph18105479

[42]. Gilani, S. Q., & Marques, O. (2023). Skin lesion analysis using generative adversarial networks: A review. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-0

Notes on contributors

Prof. Saleh Al-Sharaeh is an accomplished academic and industry expert with a strong background in Computer Engineering and Artificial Intelligence. He earned his BS degree in Computer Engineering from Jordan University of Science and Technology in 1989. Subsequently, he pursued higher studies and received his MS degree in AI from Tennessee State University, USA, in 1992, followed by his Ph.D. in Computer Engineering from the University of Alabama in Huntsville, USA, in 1996, specializing in Parallel and Distributed Computing and Networking, particularly focusing on challenging global problems Optimization and Visualization.

Amal Abu Shehab is a researcher specializing in the analysis of medical images based on deep learning. She primarily focuses on developing deep learning models for enhancing diagnostic accuracy in medical imaging. She has a strong interest in using set of techniques like data augmentation and transfer learning to improve model performance.