Intelligent Methods in Engineering Sciences 3(2): 070-081, 2024

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

https://www.imiens.org

International Open Access

June, 2024

e-ISSN 2979-9236

Research Article https://doi.org/10.58190/imiens.2024.101

Deep Learning-Based Classification of Skin Lesion Dermoscopic Images for Melanoma Diagnosis

Humar KAHRAMANLI ORNEK [a], ***Elham Tahsin YASIN*** [b], ***Burak YILMAZ*** [b], ***Murat KOKLU*** [a,*]

[a] *Department of Computer Engineering, Selcuk University, Konya, Türkiye*
[b] *Graduate School of Natural and Applied Sciences, Selcuk University, Konya, Türkiye*

ARTICLE INFO

Article history:
Received 14 April 2024
Accepted 29 June 2024

Keywords:
Deep learning,
Lesion Classification,
Melanoma,
Skin Cancer,
Skin Lesion

ABSTRACT

This study examined the effectiveness of artificial intelligence and machine learning methods in accurately categorizing skin cancers. We used the ISIC 2019 Skin Lesion dataset, which consists of images divided into eight categories. To prepare the data, we applied preprocessing techniques and extracted features using the SqueezeNet deep learning model. The dataset was divided into training and test sets using cross validation. Four well-known machine learning algorithms, namely Artificial Neural Network (ANN), k-nearest neighbors (kNN), Random Forest (RF), and Logistic Regression (LR), were employed to perform classification tasks. Each algorithm was specifically designed to suit its processing methodology. The algorithms are assessed based on several important measures. The results indicate that the Artificial Neural Network (ANN) achieved the highest accuracy rate of 71.80%, whereas the k-nearest neighbors (kNN), Random Forest (RF), and Logistic Regression (LR) achieved accuracy rates of 69.70%, 67.00%, and 67.20%, respectively. The results emphasize the capacity of machine learning algorithms to augment clinical decision-making in the field of dermatology, with the goal of enhancing early detection and treatment effectiveness for individuals with skin cancer.

1. Introduction

Melanoma is a deadly form of skin cancer caused by the uncontrollable growth of cancerous skin cells that can metastasize to other parts of the body. Dermatologists suggest that various factors such as genetics and the environment can contribute to the development of melanoma [1]. Exposure to sun ultraviolet (UV) radiation is considered the primary cause of DNA damage in melanocytes, although tanning beds and lamps are also risk factors. To reduce the incidence of malignant melanoma cases, it is recommended to limit exposure to UV rays [2].

Skin lesions refer to various skin abnormalities that can range from harmless moles to severe conditions like melanoma, which is the deadliest form of skin cancer caused by the uncontrolled growth of cancerous skin cells that produce pigment. Early melanoma detection and diagnosis are critical for successful treatment and higher survival rates [3]. Deep learning algorithms have been widely used in the medical field, including for the classification of skin lesions, by analyzing dermoscopy images. Dermoscopy is a non-invasive imaging method that allows a detailed examination of skin lesions using a special microscope [4]. Deep learning models can be trained on large dermoscopy image datasets to classify skin lesions as benign or malignant based on their color, texture, and shape. Convolutional neural networks (CNNs) and other deep learning architectures have been used for skin lesion classification with promising results [5]. Some studies have focused on developing algorithms for skin lesion segmentation to improve the classification and diagnosis of lesions, with promising results achieved using deep learning-based segmentation methods such as FCNs and U-Net.

Skin infections can be caused by various factors such as fungi, bacteria, and allergies. These infections can alter the skin's texture and color and may lead to chronic skin disorders, some of which can develop into skin cancer. Early detection of skin disorders is crucial to minimize their growth and proliferation, but it can be time-consuming and costly. Ordinary people often lack medical education and may not be aware of the form and stage of a skin condition. Even dermatologists sometimes struggle to recognize skin ailments and may need to use expensive testing facilities. Laser and photonics technologies have made it possible to diagnose skin diseases more easily and

* **Corresponding Author:** mkoklu@selcuk.edu.tr

precisely, but the cost remains high. Therefore, an image processing-based approach is recommended to identify skin diseases. This technique uses digital images of the affected skin area and image recognition to classify the disease, providing a simpler and potentially more affordable solution [6, 7]. The last decision and identification is done by experts in this sector, the use of deep learning is for early identification and classification which is not the definite result.

The process of identifying a disease involves several criteria, including the recognition of disease requirements, taking an initial photo, using filters to remove noise from the image, segmenting the image to remove irrelevant details, extracting information parameters, and organizing the diseases using an appropriate classifier [6].

The skin is a crucial organ of the human body that comprises three layers: the epidermis, dermis, and hypodermis. It protects the body from environmental damage, regulates temperature, and provides sensation [8]. The stratum corneum is the outermost layer of the epidermis that is made up of keratinocytes producing keratin responsible for skin protection. The skin contains melanocytes that produce melanin to provide color and protect from harmful UV rays [9]. However, unusual growth of melanocytes can lead to the development of melanoma, a deadly form of skin cancer [10]. DNA comprises nucleotides, and the order of nitrogen bases in DNA sequences forms genes, deciding the formation, multiplication, division, and death of cells [11]. Oncogenes promote cell multiplication and division, while tumor suppressor genes inhibit cell growth. Skin cancers are of three types, including squamous cell carcinoma, basal cell carcinoma, and malignant melanoma, with MM being the most lethal [12]. Melanin presence in the dermis is the most critical designation of melanoma, causing visible changes in skin coloration. Biopsy and histology are necessary for the explicit diagnosis of skin cancer, while melanoma is characterized by four categories based on microscopic characterizations of lesions [13].

Hypothesis: Machine learning algorithms, specifically Artificial Neural Network (ANN), k-nearest neighbors (kNN), Random Forest (RF), and Logistic Regression (LR), applied to the ISIC 2019 Skin Lesion dataset, will demonstrate varying levels of accuracy in categorizing skin cancers. We hypothesize that the ANN will exhibit the highest classification accuracy among the tested algorithms, indicating its potential effectiveness in augmenting clinical decision-making for early detection and treatment of skin cancers.

2. Related Works

In the literature, there are numerous studies conducted on skin cancer research, using various methodologies for classification processes. Researchers have employed different machine learning algorithms and deep learning models to classify and analyze skin cancer data. Some of the notable studies on skin cancer research in the literature are highlighted below.

Tschandl et al., introduced a new dataset of dermatoscopic images, named HAM10000, which contains 10015 images covering a variety of diagnostic categories for pigmented lesions. The dataset was created using diverse acquisition and cleaning methods, and semi-automatic workflows that employed specially trained neural networks. Over 50% of the lesions were validated by pathology, and the remaining cases were confirmed through follow-up, expert consensus, or in-vivo confocal microscopy. The dataset has been made public and can be used as a training set for academic machine-learning research and is available through the ISIC archive [14].

Arif et al., introduced an automated system for skin lesion classification that utilizes deep learning and transfer learning to detect melanoma in its early stages. The system involves several digital images processing techniques, including noise removal and image enhancement, followed by image segmentation using clustering. The extracted features are selected using Harris Hawks optimization (HHO), and various classifiers are employed to predict the stages of skin cancer with higher values of sensitivity, precision, accuracy, and specificity. Highest accuracy of 97.3 was obtained by the hybrid classifiers of CNN and IHHO. The study shows that the proposed approach outperforms current methods in terms of classification rate [15].

Safdar et al., outlined a melanoma detection framework that uses deep learning and multiple skin lesion databases. It involves pre-processing of dermoscopic images and segmentation using a Fully Convolutional Network, followed by an ensemble of deep ResNet-50 and Inception-V3 for binary classification of benign or melanoma lesions. An accuracy of 93.4% was found. The proposed approach outperforms other melanoma diagnosis frameworks, and it can serve as a practical medical assistant for dermatologists [2].

Gavrilov and colleagues introduced a deep Artificial Neural Network (ANN) that utilizes the Inception-V3 model to diagnose various skin lesions in a timely manner. They trained the model using images from the ImageNet dataset and constructed an ensemble of five models to improve lesion detection accuracy. The top-performing models achieved an accuracy score of 91%, sensitivity of 85%, AUC ROC of 96%, and specificity of 92% when evaluated on the ImageNet dataset [16].

Waheed et al., aimed to present a machine learning approach for the efficient detection of melanoma from dermoscopic images. It extracts different types of color and texture features from the images and feeds them to the classifier for classification. The proposed method is tested on the PH2 dataset, and 96% result was achieved in terms of accuracy [17].

Adegun and Viriri The abstract presents a method for automated detection and segmentation of melanoma lesions using deep learning. The system uses a multi-stage and multi-scale approach, and a softmax classifier is used for pixel-wise classification of melanoma lesions. A new method, Lesion-classifier, is proposed for the classification of skin lesions into melanoma and non-melanoma categories. The proposed method achieves higher accuracy and dice coefficient compared to some state-of-the-art methods on two benchmark skin lesion datasets. The method achieves 95% accuracy and 92% dice coefficient on ISIC 2017 and 95% accuracy and 93% dice coefficient on PH2 [18].

The proposed study suggests a Gabor wavelet-based deep convolutional neural network for accurately detecting two common types of skin lesions: malignant melanoma and seborrheic keratosis. The proposed method decomposes input images into seven directional sub-bands and employs eight parallel CNNs for generating probabilistic predictions. The sub-band images and input image are used for decision fusion to classify the skin lesion. The study highlights the significance of early detection and treatment of skin cancer and claims that the proposed method outperforms the current literature's alternative methods for skin cancer detection. An accuracy of 91.0% was obtained from Gabor and ResNet-18 classifier [19].

Hasan et al., aimed to create a semantic segmentation network for accurate skin lesion segmentation, called Dermoscopic Skin Network (DSNet), which uses a depth-wise separable convolution to reduce the number of parameters and make the network lightweight. DSNet was tested against two other networks, U-Net and Fully Convolutional Network (FCN8s), on two publicly available datasets: ISIC-2017 and PH2. The results showed that DSNet outperformed the other methods with mean Intersection over Union (mIoU) scores of 77.5% and 87.0% on the two datasets, respectively. On the ISIC-2017 dataset, DSNet outperformed U-Net and FCN8s by 3.6% and 6.8%, respectively. The study suggests that DSNet could lead to better melanoma detection performance, and the trained model and source code are available for further research [20].

Hasan et al., focused on providing an automated and accurate method for segmenting skin lesions to aid in the early diagnosis of skin cancer. The study utilizes image processing techniques in the first stage to remove artifacts from the images, and in the second stage, a modified U-Net architecture with a 46-layer structure is proposed for successful lesion segmentation. Two different U-Net architectures (U-Net 32 and U-Net 46) were tested in the experiments conducted in this study. Skin cancer, particularly malignant melanoma, is a dangerous disease that can be effectively treated if detected early [21].

Shan et al., proposed a new method for automatic skin lesion segmentation in dermoscopy images. The method, called FC-DPN, is built on a combination of fully convolutional networks (FCN) and dual path networks (DPN) and aims to address challenges such as low contrast, artefacts, and diverse skin lesion characteristics. FC-DPN uses sub-DPN projection and processing blocks to acquire more representative and discriminative features for accurate segmentation. The proposed method outperforms other established segmentation algorithms and achieves high accuracy scores on two datasets, with an average Dice coefficient of 88.13% and a Jaccard index of 80.02% on the modified ISBI 2017 Skin Lesion Challenge test dataset and 90.26% and 83.51%, respectively, on the PH2 dataset [22].

Murgugan et al., presented a study to detect skin cancer using a computer-based method that utilizes image processing. The skin region is preprocessed by applying a median filter and a mean shift segmentation method to isolate the affected area from the healthy skin. Different features such as Moment Invariant, Gray Level Co-occurrence Matrix (GLCM), and Gray Level Run Length Matrix (GLRLM) are extracted from the image and classified using several methods including SVM, PNN, RF, and Combined SVM+RF. The results show that Combined SVM+RF method outperforms other classifiers with an accuracy of 89.31% for GLCM, 86.12% for GLRLM, and 84.63% for Moment Invariant features. This approach is an effective and accurate way to diagnose skin cancer, providing patients with timely treatment and saving them time [23].

Rodrigues et al., discussed the use of Transfer Learning and Deep Learning in an IoT system in their abstract as a means of assisting doctors in the diagnosis of common skin lesions like typical nevi and melanoma. The study used various Convolutional Neural Networks (CNNs) and classifiers to extract features and classify skin lesions from two datasets. The DenseNet201 extraction model and KNN classifier achieved high accuracy rates of 96.805% for the ISBI-ISIC dataset and 93.167% for the PH2 dataset. This method can provide reliable and efficient assistance to doctors in diagnosing skin lesions, particularly in areas with limited medical resources [24].

Nida et al., suggested a deep learning approach for automatically segmenting the melanoma region achieved exceptional performance measures on the ISIC-2016 benchmark dataset. The method demonstrated an average pixel level specificity of 0.9417, pixel level sensitivity of 0.9781, F1-score of 0.9589, and pixel level accuracy of 0.948. Furthermore, the average dice score for segmentation was 0.94, and the Jaccard coefficient had an average value of 0.93 on the entire testing images. These outcomes establish the superiority of the proposed method compared to state-of-the-art approaches. The deep region-based convolutional neural network used in this method was competent in computing deep features and enhancing the melanoma region's segmentation performance [25].

Monika et al., objectives to use machine learning and image processing techniques to detect and classify different types of skin cancer. The process involves using dermoscopic images as input, and applying techniques such as hair particle removal, image smoothing, noise filtering and edge preservation using various filters. Segmentation is performed using k-means clustering and features are extracted using various methods. The experimental analysis is conducted on a dataset called ISIC 2019 Challenge, and Multi-class Support Vector Machine (MSVM) is used for classification, with an accuracy of 96.25%. The goal is to help in the early detection of skin cancer, which can prevent its progression and save lives [26].

Ozkan and Koklu presented in the International Journal of Intelligent Systems and Applications in Engineering focuses on the classification of skin lesions using machine learning techniques to enhance diagnostic accuracy and support medical decision-making in dermatology. The research employs four classification methods - Artificial Neural Networks (ANN), Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Decision Trees (DT) - to analyze dermoscopic images of skin lesions. Through experimental studies and cross-validation methods, the study determines that ANN outperforms SVM, KNN, and DT, achieving an accuracy rate of 92.50%. The developed ANN classifier is identified as a valuable medical decision support system for dermatologists in diagnosing skin lesions. The study also highlights the potential for further advancements by incorporating different data processing techniques and hybrid classification algorithms, as well as integrating image processing methods for autonomous decision-making in medical applications. Overall, the research contributes to the field of dermatology by providing a robust framework for accurate skin lesion classification using machine learning algorithms [27].

Various studies have been conducted on skin cancer and still studies are going to be presented with different and new datasets. This study is also among them with one of the public datasets for skin lesion and preprocesses have taken place on the dataset. Feature extraction using deep learning will be applied. Later the output was classified by machine learning algorithms. The outcomes will be analyzed and evaluated.

3. Material and Methods

In this section, we will discuss the methods employed in the study to tackle the research objectives and address the classification task of the ISIC 2019 Skin Lesion dataset. The dataset was obtained from the public respository. After analysing the dataset some preprocessing methods were applied to the dataset. After the dataset prepared for the classification methods. Since the research data consists of images we planed to extract the features from the images. After the feature extraction which takes place with the help of SqueezeNet deep learning. The dataset seperated into train data and test data with cross validation method. The classification method carried out using four distinct machine learning techniques [28]. The methods were Artificial Neural Network (ANN), k Nearest Neighbors (kNN), Randon Forest (RF), and Logistic Regression (LR). Every machine learning technique possesses its own distinct processing methodology, which will be elaborated upon in the subsequent sections. After all the obtained results from the machine learning algorithms anaylsed and evaluated. The step by step research flow in presented in Figure 1.

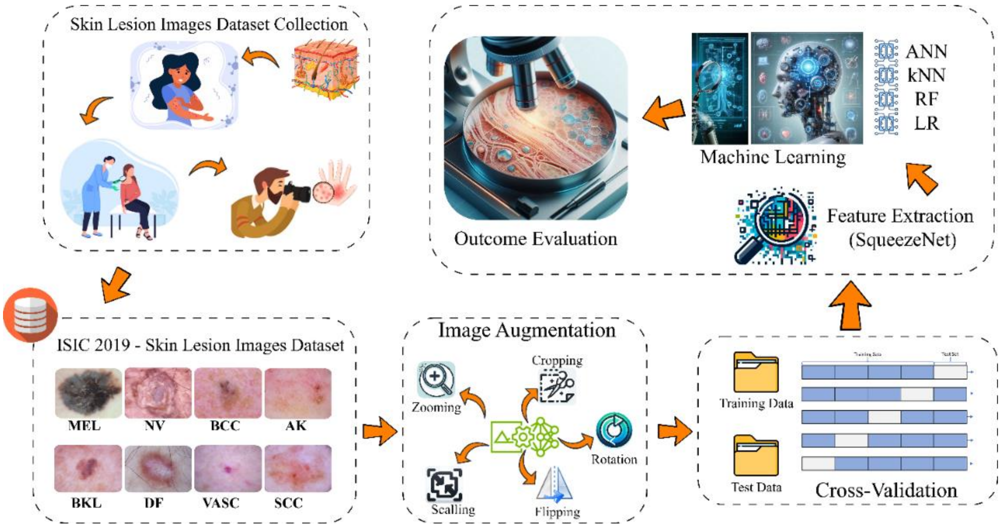
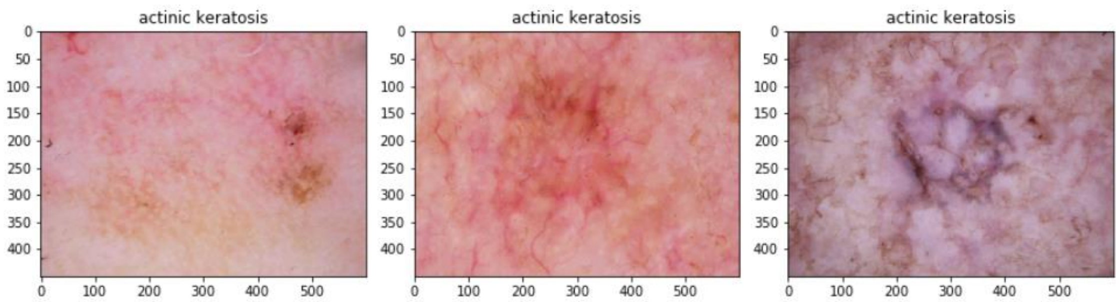


Figure 1. Skin Cancer Image Classification Flow Diagram

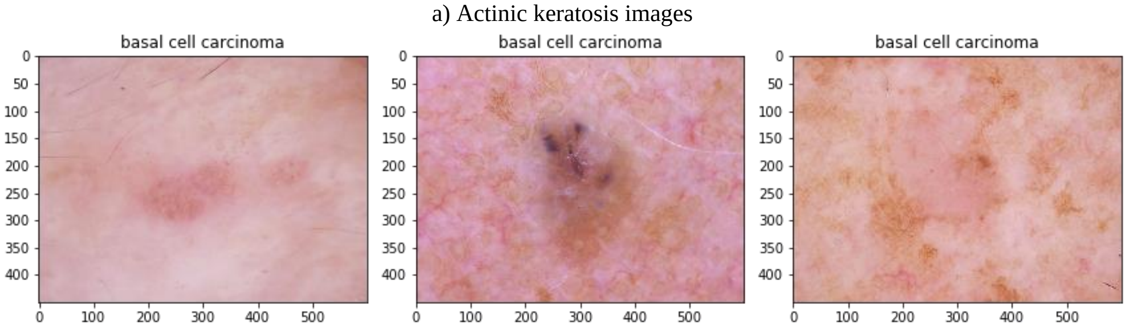
3.1. Skin Lesion Images for Melanoma

The dataset used in this study is the Skin Lesion Images for Melanoma Classification dataset, which consists of 25,331 dermoscopic images belonging to 9 different diagnostic categories. The dataset includes images from the ISIC 2019 challenge, as well as data from previous years (2018 and 2017). The images are labeled with the following categories: Melanoma (MEL), Basal cell carcinoma (BCC), Melanocytic nevus (NV), Actinic keratosis (AK), Dermatofibroma (DF), Benign keratosis (solar lentigo/seborrheic keratosis/lichen planus-like keratosis) (BKL), Squamous cell carcinoma (SCC), Vascular lesion (VASC), and None of the above. The Skin Lesion Images for Melanoma dataset consists of images categorized into nine classes. However, it should be noted that the last class, labeled as "None of the above," exhibited all-zero values in the accompanying Excel file. This indicates that there were no images present in this class, rendering it empty. Consequently, our study focused on analyzing the remaining eight classes within the dataset. The dataset sources include the BCN_20000 Dataset from the Department of Dermatology [29], Hospital Clínic de Barcelona, the HAM10000 Dataset [14] from the ViDIR Group, Department of Dermatology, Medical University of Vienna, and the MSK Dataset [30, 31]. The dataset is licensed under CC BY NC 4.0. This dataset provides a comprehensive collection of diverse skin lesion images for the classification of melanoma, offering valuable resources for research and development of accurate diagnostic models in the field of dermatology.

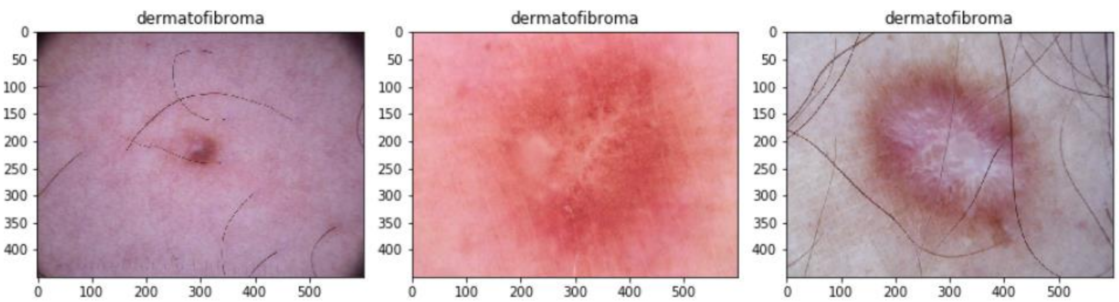
To visually showcase representative images from each class within the dataset, a visualization process is implemented. Initially, the labels of all unique classes present in the dataset are collected. Subsequently, a specified number of example images are selected and displayed for each class. The resulting visual representation of these examples is presented in Figure 2 (a-h).



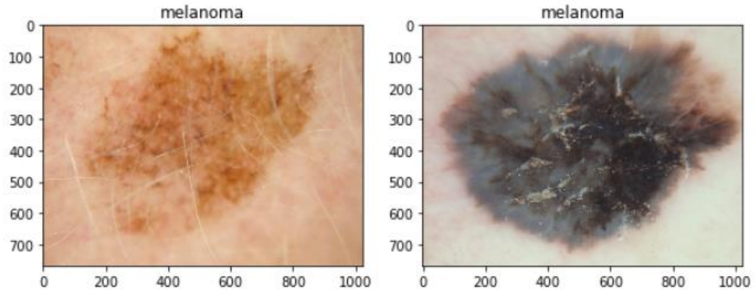
a) Actinic keratosis images



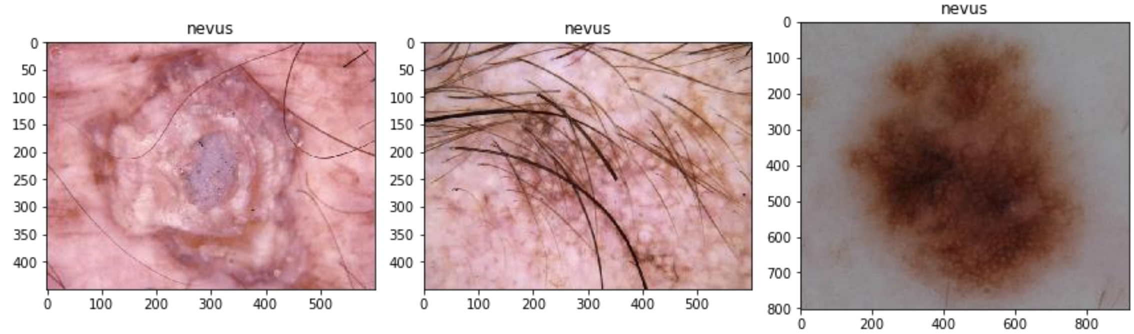
b) Basal cell carcinoma



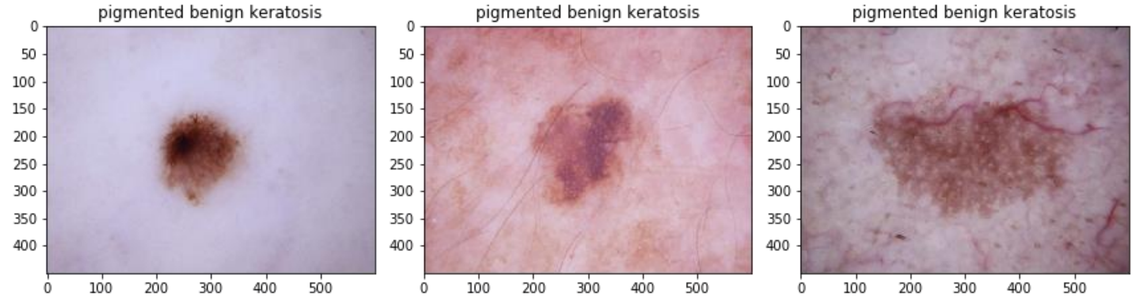
c) Dermatofibroma



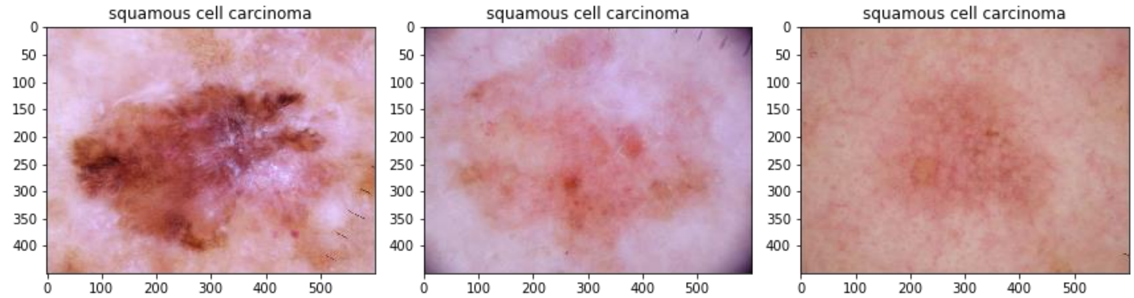
d) Melanoma



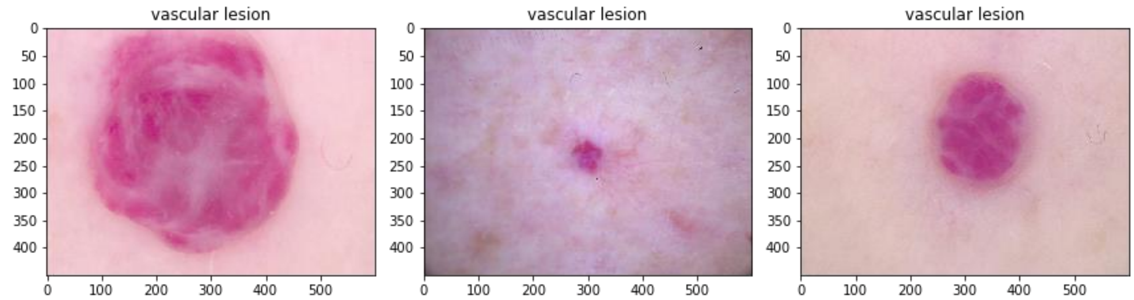
e) Melanocytic nevus



f) Benign keratosis



g) Squamous cell carcinoma



h) Vascular lesion

Figure 2**.** Skin Lesion Dataset Image Presentation for all 8 classes (a-h)

3.2. Cross validation

The ISIC 2019 Skin Lesion dataset was evaluated using 10-fold cross-validation. In this method, the data is divided into 10 subsets. Each subset is used as a validation set once while the remaining nine subsets form the training set. This process is repeated 10 times, ensuring that every data point is used for both training and validation. This approach helps us thoroughly assess the model's performance in classifying skin lesion images by providing a robust evaluation framework that maximizes data utilization for training and validation. A visual representation of the dataset distribution is shown in Figure 3.

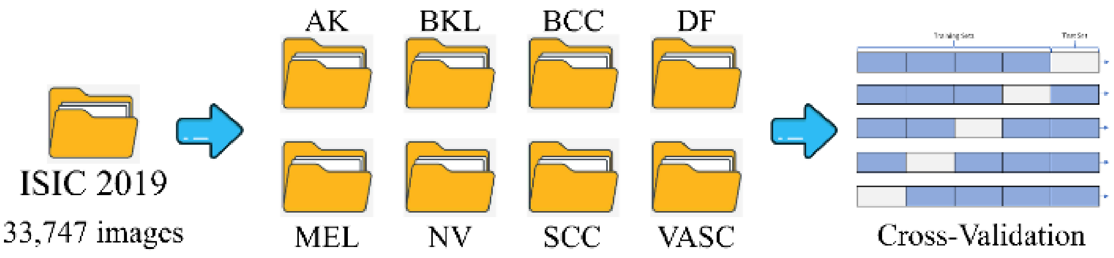


Figure 3. ISIC 2019 Skin Lesion dataset distribution

3.3. Image Preprocessing

The initial dataset acquired from the public repository had an imbalance, wherein certain classes contained a much higher number of photos compared to others. This difference might result in model bias, overfitting, reduced accuracy, and limited generalization [32].

Data augmentation is an efficient method for addressing these concerns. Data augmentation is a technique that helps balance the dataset by artificially boosting the size of the minority classes. This is achieved by methods such as rotation, flipping, scaling, and cropping. This improves the model's capacity to learn from all classes, therefore enhancing overall performance and generalization capabilities.

The initial dataset was augmented from 25,331 dermoscopic images to 33,747 images across 8 distinct categories. Augmentation was only used for classes that have a smaller quantity of images, specifically AK, DF, SCC, and VASC. Table 1 displays the number of images in each category prior to and following augmentation.

Table 1. Image Counts Per Class Before and After Augmentation

Classes	Image Numbers	
	Before Augmentation	After Augmentation
AK	867	3467
BCC	3323	3323
BKL	2624	2624
DF	239	2149
MEL	4522	4522
NV	12875	12875
SCC	628	2512
VASC	253	2275

3.4. Confusion Matrix

When working with deep learning, it's important to be able to evaluate how well your classification model is performing. Confusion matrix allows you to compare the predicted class labels with the actual ones in different situations. Specifically, when you're dealing with binary classification problems, the confusion matrix has four parts: true positives, false positives, true negatives, and false negatives. True positives means that instances were correctly classified as positive, while false positives happen when negative instances are mistakenly labeled as positive. True negatives are when negative instances are correctly classified, and false negatives happen when instances are incorrectly labeled as negative [33].

The confusion matrix is a useful tool for evaluating the performance of a model in distinguishing between distinct classes. Values that are high in the top-left corner of the data suggest accurate forecasts, but lower values or numbers that are not on the diagonal indicate errors. By analyzing the confusion matrix, one can identify the strengths and flaws of the model, which can then be used to guide modifications.

3.5. Feature extraction

The dataset comprises a vast number of images. To effectively apply machine learning techniques, we began by extracting features from each image. This crucial step involved using the SqueezeNet deep learning model **[34]**, which efficiently extracts 1,000 features from each image. By leveraging these detailed image embeddings, we were able to feed the extracted features into various machine learning algorithms, enhancing their ability to learn and make accurate predictions based on the rich, nuanced information captured by SqueezeNet. SqueezeNet's design as a lightweight deep learning model is one of the reasons it's a great choice for feature extraction [35]. SqueezeNet is perfect for deployment on devices with limited storage and processing capability, including mobile phones and embedded systems, because it produces a much smaller model size than conventional models like AlexNet and VGGNet. Its architecture allows for reduced computational complexity and fewer parameters by using 1x1 convolutions to limit the number of input channels to 3x3 filters (squeeze layers) and then expanding them with a combination of 1x1 and 3x3 convolutions (expand layers). Faster inference times and effective memory use are the outcomes, which are essential for real-time applications and resource-constrained contexts. SqueezeNet offers competitive performance and accuracy levels equivalent to bigger models despite its compact size, making it a well-balanced choice for applications that demand both efficiency and performance. It may also be applied to a wide range of tasks because of its flexibility in fine-tuning and transfer learning, which increases its application in a variety of real-world circumstances [33, 36, 37].

3.6. Classification Models

In this study, we applied four machine learning algorithms to the extracted features of the dataset: Artificial Neural Networks (ANN), k-Nearest Neighbors (kNN), Random Forest (RF), and Logistic Regression (LR) [38, 39]. The architecture for each algorithm is shown in Figure 4 [40-44].

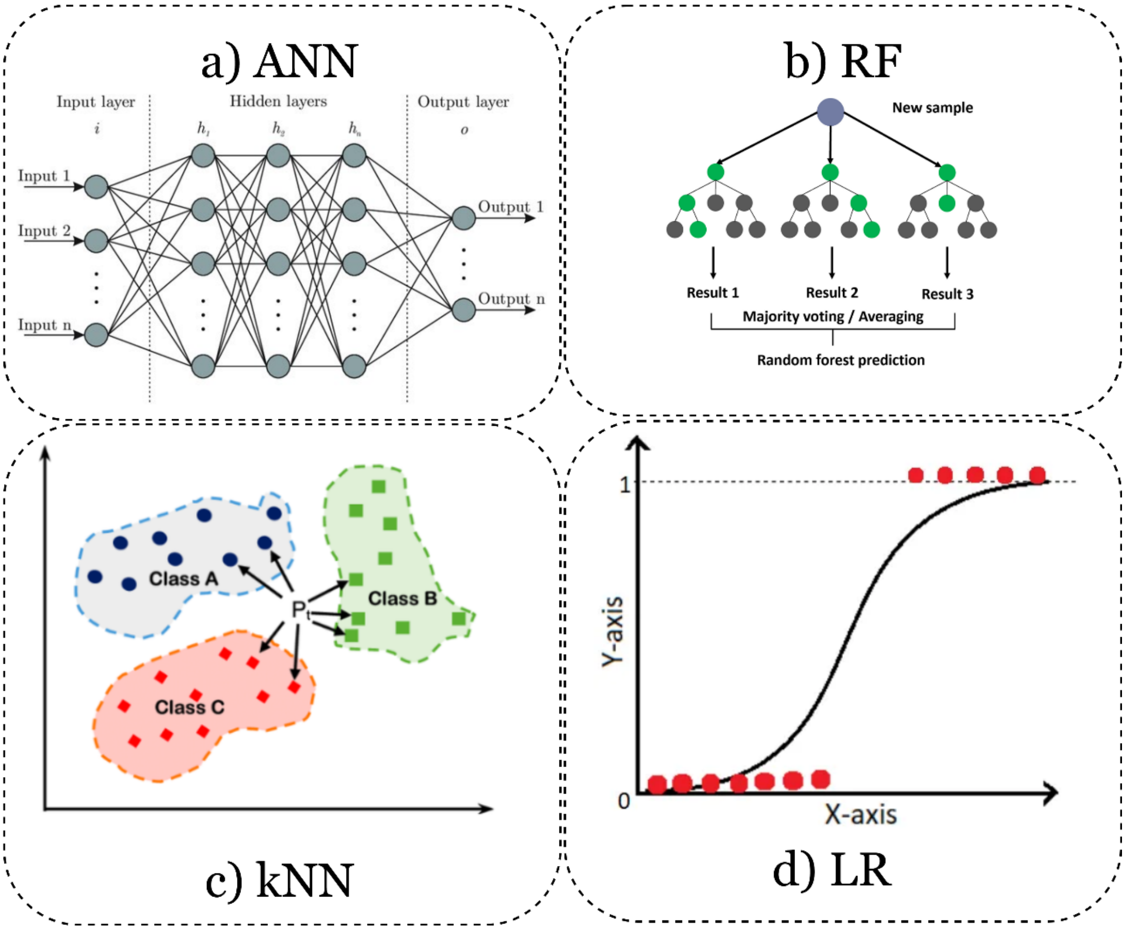


Figure 4. Architectures of Used Machine Learning Algorithms

Artificial Neural Networks (ANN): are computational models that draw inspiration from the structure and functioning of the human brain. Neural networks are composed of layers of interconnected nodes, or neurons, with each connection having a specific weight. The input features are propagated across the network, with each layer conducting weighted summation and activation functions, resulting in the output [45-47]. ANN employed with 100 neurons in the hidden layer, ReLU activation function, Adam solver, regularization set to 0.0001, and a maximum of 200 iterations.

The k-Nearest Neighbors (kNN): algorithm is a straightforward and instance-based learning method. The data point is classified by determining the majority class of its k nearest neighbors in the feature space, using distance metrics such as Euclidean distance [45, 48]. The k-Nearest Neighbors algorithm applied with 10 neighbors, Euclidean distance metric, and uniform weighting scheme.

Random Forest (RF): is an ensemble learning technique that builds several decision trees during the training process. Every tree provides a categorization, and the forest chooses the class with the highest number of votes. This approach mitigates the problem of overfitting and enhances the accuracy of the results [45, 48]. We employed the Random Forest (RF) algorithm with 10 trees in our analysis.

Logistic Regression (LR): is a statistical model specifically designed for binary classification tasks. The logistic function is employed to represent the likelihood of a class label given input features. The model determines the optimal parameters that maximize the likelihood of the observed data [45, 46]. We used the Logistic Regression

(LR) algorithm with Ridge (L2) regularization and a regularization strength parameter (C = 1).

3.7. Performance Metrics

These metrics offer valuable information about various facets of model performance, aiding in the assessment of classification accuracy, comprehensiveness, and resilience [49, 50]. A concise explanation of each performance metric:

Accuracy: Accuracy measures the proportion of correctly predicted instances (both true positives and true negatives) out of the total instances.

Precision: Precision measures the proportion of true positive predictions out of all positive predictions made by the model. Precision indicates how many of the predicted positive instances are actually positive.

Recall (Sensitivity): Recall measures the proportion of true positive instances that are correctly predicted by the model out of all actual positive instances. Recall indicates the model's ability to correctly identify positive instances.

F1 Score: The F1 score is the harmonic mean of precision and recall. It provides a single metric that balances both precision and recall. The F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

Matthews Correlation Coefficient (MCC): MCC measures the quality of binary classifications, considering all four confusion matrix values (true positives, true negatives, false positives, and false negatives). MCC ranges from -1 (complete disagreement between prediction and observation) to +1 (perfect prediction). A coefficient of 0 indicates a random prediction.

The calculation formula for each metric is shown in figure 5.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

Figure 5. Formulas for Performance Metrics

4. Experimental Results and Discussion

We employed machine learning, for the diagnosis and categorization of skin malignancies using artificial intelligence. The study obtained skin cancer photos from the comprehensive "ISIC-2019" collection, which consists of eight different types of lesions detected and classified by medical experts. The lesion types are distinguished by their corresponding codes: 'MEL' (Melanoma), 'NV' (Melanocytic nevi), 'BCC' (Basal cell carcinoma), 'AK' (Actinic keratoses), 'BKL' (Benign keratosis), 'DF' (Dermatofibroma), 'VASC' (Vascular lesions), and 'SCC' (Squamous cell carcinoma). The results are presented in Table 2.

Table 2. Performance Metrics Results Presentation

Algorithms	Performance Metrics				
	Accuracy	Precision	Recall	F1-Score	MCC
ANN	71.80%	0.717	0.718	0.718	0.645
kNN	69.70%	0.687	0.697	0.683	0.618
RF	67.00%	0.672	0.670	0.645	0.574
LR	67.20%	0.660	0.672	0.662	0.583

Table 2 provides an analysis of four machine learning algorithms: Artificial Neural Network (ANN), k-Nearest Neighbors (kNN), Random Forest (RF), and Logistic Regression (LR). These algorithms were evaluated using various performance criteria. The accuracy of predictions, which measure the total correctness, indicates that ANN has a leading accuracy of 71.80%, closely followed by kNN with 69.70%. The precision, which measures the accuracy of positive predictions, is higher for the ANN model at 0.717 compared to the LR model which has a little lower precision of 0.660. The recall metric, which measures the accuracy of properly identifying genuine positive cases, shows that ANN has the highest score of 0.718, while kNN does well with a score of 0.697. The F1-Score, which combines precision and recall into a single metric, demonstrates that the ANN achieves a value of 0.718, hence validating its balanced performance across many measures. The MCC, which assesses the accuracy of binary classifications, indicates that the ANN performs the best with a score of 0.645, while the LR model follows with a score of 0.583. In general, ANN exhibits strong performance in all criteria, indicating its appropriateness for the categorization task at hand.

The multi-class confusion matrix displays the classification performance of four algorithms. ANN, kNN, RF, and LR, on eight different classes of skin lesion images. AK, BCC, BKL, DF, MEL, NV, SCC, and VASC are several types of skin conditions. Every individual cell within the matrix denotes the frequency at which a specific class was forecasted by an algorithm in relation to the true class. The matrix facilitates the assessment of classification accuracy by identifying the patterns of accurate and inaccurate predictions for each method within the different classes. It offers a significant analysis of the strengths and limitations of algorithms, helping to evaluate and choose the most efficient model for jobs involving the classification of skin lesions. The confusion matrix for this study is shown in Table 3.

Table 3. Multi Class Confusion Matrix of Skin Lesion Images

ANN / kNN / RF / LR		PREDICTED							
		AK	BCC	BKL	DF	MEL	NV	SCC	VASC
ACTUAL	AK	2579	197	99	70	106	74	330	12
		2821	179	56	71	38	136	155	11
		2530	168	40	22	93	448	156	10
		2256	258	98	145	123	157	403	27
	BCC	198	2078	294	34	217	364	109	29
		421	1886	124	95	129	502	134	32
		346	1567	62	12	185	1006	122	23
		253	2024	147	43	172	488	163	33
	BKL	103	243	1127	19	422	647	53	10
		283	428	747	71	236	746	88	25
		210	299	466	2	329	1228	65	25
		151	347	828	38	346	853	40	21
	DF	82	43	31	1716	27	115	113	22
		52	45	12	1903	9	99	25	4
		82	50	6	1131	20	767	77	16
		184	58	20	1341	31	285	199	31
	MEL	126	222	405	33	2401	1241	77	17
		339	351	219	101	1691	1658	125	38
		229	196	112	5	1779	2100	76	25
		193	291	286	34	1936	1656	92	34
	NV	105	362	606	110	1166	10384	63	79
		268	599	261	259	626	10665	99	98
		104	254	100	25	390	11889	45	68
		117	440	321	178	713	10920	87	99
	SCC	299	116	58	96	72	51	1813	7
		318	158	31	130	52	155	1659	9
		376	104	20	31	121	540	1310	10
		529	166	49	215	94	126	1309	24
	VASC	9	24	6	15	19	59	13	2130
		11	9	7	10	3	68	7	2160
		14	25	3	10	17	268	6	1932
		17	22	6	38	28	74	15	2075

We acknowledge several limitations in our study and propose potential future research directions to address them. First, the reliance on the ISIC 2019 Skin Lesion dataset may limit the generalizability of our findings due to its lack of real-world diversity. The preprocessing techniques used were standard; employing more sophisticated methods could enhance feature extraction and classification accuracy.

5. Conclusion

This study employs an artificial intelligence-based approach using machine learning algorithms for accurate classification of skin cancers. By leveraging the ISIC 2019 Skin Lesion dataset, comprising images labeled across eight different classes, we applied preprocessing methods and feature extraction using the SqueezeNet deep learning architecture. The dataset was divided into training and test sets using cross-validation. Classification was performed using four distinct machine learning algorithms: ANN, kNN, RF, and LR. Every algorithm was customized to its particular processing methodology, as outlined in this research. Results from these algorithms were analyzed and evaluated to assess their effectiveness in classifying skin lesions. The findings highlight the potential of deep learning models in supporting clinical decision-making in dermatology, aiming to improve early diagnosis and treatment outcomes for skin cancer patients.

Data Availability

The data is available on the public website of Kaggle through this link: https://www.kaggle.com/datasets/andrewmvd/isic-2019

Conflicts of Interest

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] G. B. Loganathan *et al.*, "Melanoma classification using enhanced fuzzy clustering and DCNN on dermoscopy images," *NeuroQuantology,* vol. 20, no. 12, p. 196, 2022, doi: 10.14704/NQ.2022.20.12.NQ77017.

[2] K. Safdar, S. Akbar, and S. Gull, "An automated deep learning based ensemble approach for malignant melanoma detection using dermoscopy images," in *2021 International Conference on Frontiers of Information Technology (FIT)*, 2021: IEEE, pp. 206-211.

[3] K. Feng, L. Ren, G. Wang, H. Wang, and Y. Li, "SLT-Net: A codec network for skin lesion segmentation," *Computers in Biology and Medicine,* vol. 148, p. 105942, 2022.

[4] X. Wang, X. Jiang, H. Ding, Y. Zhao, and J. Liu, "Knowledge-aware deep framework for collaborative skin lesion segmentation and melanoma recognition," *Pattern Recognition,* vol. 120, p. 108075, 2021.

[5] A. Mahbod, P. Tschandl, G. Langs, R. Ecker, and I. Ellinger, "The effects of skin lesion segmentation on the performance of dermatoscopic image classification," *Computer Methods and Programs in Biomedicine,* vol. 197, p. 105725, 2020.

[6] P. Choudhary, J. Singhai, and J. Yadav, "Skin lesion detection based on deep neural networks," *Chemometrics and Intelligent Laboratory Systems,* vol. 230, p. 104659, 2022.

[7] N. S. A. ALEnezi, "A method of skin disease detection using image processing and machine learning," *Procedia Computer Science,* vol. 163, pp. 85-92, 2019.

[8] S. Pathan, K. G. Prabhu, and P. Siddalingaswamy, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review," *Biomedical Signal Processing and Control,* vol. 39, pp. 237-262, 2018.

[9] A. Uong and L. I. Zon, "Melanocytes in development and cancer," *Journal of cellular physiology,* vol. 222, no. 1, pp. 38-41, 2010.

[10] Y. S. Taspinar, I. Cinar, R. KURSUN, and M. KOKLU, "Monkeypox Skin Lesion Detection with Deep Learning Models and Development of Its Mobile Application," *Public health,* vol. 500, p. 5.

[11] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012," *JAMA dermatology,* vol. 151, no. 10, pp. 1081-1086, 2015.

[12] J. Feng, N. G. Isern, S. D. Burton, and J. Z. Hu, "Studies of secondary melanoma on C57BL/6J mouse liver using 1H NMR metabolomics," *Metabolites,* vol. 3, no. 4, pp. 1011-1035, 2013.

[13] H. Li, Y. Pan, J. Zhao, and L. Zhang, "Skin disease diagnosis with deep learning: a review," *Neurocomputing,* vol. 464, pp. 364-393, 2021.

[14] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data,* vol. 5, no. 1, pp. 1-9, 2018, doi: 10.1038/sdata.2018.161.

[15] M. Arif, F. M. Philip, F. Ajesh, D. Izdrui, M. D. Craciun, and O. Geman, "Automated detection of nonmelanoma skin cancer based on deep convolutional neural network," *Journal of Healthcare Engineering,* vol. 2022, 2022.

[16] D. Gavrilov, A. V. Melerzanov, N. Shchelkunov, and E. Zakirov, "Use of neural network-based deep learning techniques for the diagnostics of skin diseases," *Biomedical Engineering,* vol. 52, pp. 348-352, 2019.

[17] Z. Waheed, A. Waheed, M. Zafar, and F. Riaz, "An efficient machine learning approach for the detection of melanoma using dermoscopic images," in *2017 International conference on communication, computing and digital systems (C-CODE)*, 2017: IEEE, pp. 316-319.

[18] A. A. Adegun and S. Viriri, "Deep learning-based system for automatic melanoma detection," *IEEE Access,* vol. 8, pp. 7160-7172, 2019.

[19] S. Serte and H. Demirel, "Gabor wavelet-based deep learning for skin lesion classification," *Computers in biology and medicine,* vol. 113, p. 103423, 2019.

[20] M. K. Hasan, L. Dahal, P. N. Samarakoon, F. I. Tushar, and R. Martí, "DSNet: Automatic dermoscopic skin lesion segmentation," *Computers in biology and medicine,* vol. 120, p. 103738, 2020.

[21] S. N. Hasan, M. Gezer, R. A. Azeez, and S. Gulsecen, "Skin lesion segmentation by using deep learning techniques," in *2019 Medical Technologies Congress (TIPTEKNO)*, 2019: IEEE, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895078.

[22] P. Shan, Y. Wang, C. Fu, W. Song, and J. Chen, "Automatic skin lesion segmentation based on FC-DPN," *Computers in Biology and Medicine,* vol. 123, p. 103762, 2020.

[23] A. Murugan, S. A. H. Nair, A. A. P. Preethi, and K. S. Kumar, "Diagnosis of skin cancer using machine learning techniques," *Microprocessors and Microsystems,* vol. 81, p. 103727, 2021.

[24] D. d. A. Rodrigues, R. F. Ivo, S. C. Satapathy, S. Wang, J. Hemanth, and P. P. Reboucas Filho, "A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system," *Pattern Recognition Letters,* vol. 136, pp. 8-15, 2020.

[25] N. Nida, A. Irtaza, A. Javed, M. H. Yousaf, and M. T. Mahmood, "Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering," *International journal of medical informatics,* vol. 124, pp. 37-48, 2019.

[26] M. K. Monika, N. A. Vignesh, C. U. Kumari, M. Kumar, and E. L. Lydia, "Skin cancer detection and classification using machine learning," *Materials Today: Proceedings,* vol. 33, pp. 4266-4270, 2020.

[27] I. A. Ozkan and M. Koklu, "Skin lesion classification using machine learning algorithms," *International Journal of Intelligent Systems and Applications in Engineering,* vol. 5, no. 4, pp. 285-289, 2017, doi: 10.18201/ijisae.2017534420.

[28] Y. S. Taspinar, I. Cinar, and M. Koklu, "Classification by a stacking model using CNN features for COVID-19 infection diagnosis," *Journal of X-ray science and technology,* vol. 30, no. 1, pp. 73-88, 2022, doi: 10.3233/XST-211031.

[29] M. Combalia *et al. BCN_20000*, doi: 10.48550/arXiv.1908.022.

[30] N. C. Codella *et al.*, "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)*, 2018: IEEE, pp. 168-172, doi: 10.48550/arXiv.1710.05006.

[31] N. Codella *et al.*, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1902.03368,* 2019, doi: 10.48550/arXiv.1902.03368.

[32] E. Avuclu, A. Elen, and H. Kahramanli Ornek, "Making inferences about settlements from satellite images using glowworm swarm optimization," *Journal of Electrical Engineering & Technology*, vol. 15, pp. 2345-2360, 2020, doi: 10.1007/s42835-020-00509-3.

[33] R. Kursun, I. Cinar, Y. S. Taspinar, and M. Koklu, "Flower recognition system with optimized features for deep features," in *2022 11th Mediterranean Conference on Embedded Computing (MECO)*, 2022: IEEE, pp. 1-4, doi: 10.1109/MECO55406.2022.9797103.

[34] M. Kizilgul *et al.*, "Real-time detection of acromegaly from facial images with artificial intelligence," *European journal of endocrinology*, vol. 188, no. 1, pp. 158-165, 2023, doi: 10.1093/ejendo/lvad005.

[35] I. Cinar, "Disaster Detection Using Machine Learning Methods with Deep Features," *disasters*, vol. 19, p. 21, 2023.

[36] R. Butuner, I. Cinar, Y. S. Taspinar, R. Kursun, M. H. Calp, and M. Koklu, "Classification of deep image features of lentil varieties with machine learning techniques," *European Food Research and Technology*, vol. 249, no. 5, pp. 1303-1316, 2023, doi: 10.1007/s00217-023-04214-z.

[37] Y. Unal, Y. S. Taspinar, I. Cinar, R. Kursun, and M. Koklu, "Application of pre-trained deep convolutional neural networks for coffee beans species detection," *Food Analytical Methods*, vol. 15, no. 12, pp. 3232-3243, 2022, doi: 10.1007/s12161-022-02362-8.

[38] I. Cinar, Y. S. Taspinar, and M. Koklu, "Development of early stage diabetes prediction model based on stacking approach," *Tehnički glasnik*, vol. 17, no. 2, pp. 153-159, 2023, doi: 10.31803/tg-20211119133806.

[39] Y. S. Taspinar, "Diabetic Rethinopathy phase identification with deep features," in *2022 11th Mediterranean conference on embedded computing (MECO)*, 2022: IEEE, pp. 1-4, doi: 10.1109/MECO55406.2022.9797136.

[40] F. Bre, J. M. Gimenez, and V. D. Fachinotti, "Prediction of wind pressure coefficients on building surfaces using artificial neural networks," *Energy and Buildings*, vol. 158, pp. 1429-1441, 2018, doi: 10.1016/j.enbuild.2017.11.045.

[41] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, 2003: Springer, pp. 986-996.

[42] R. Gandhi, "Nearest Neighbours—Introduction to Machine Learning Algorithms," *Towards Data Science*, 2018.

[43] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197-227, 2016.

[44] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[45] E. T. Yasin, I. A. Ozkan, and M. Koklu, "Detection of fish freshness using artificial intelligence methods," *European Food Research and Technology*, vol. 249, no. 8, pp. 1979-1990, 2023, doi: 10.1007/s00217-023-04271-4.

[46] K. Erdem, M. B. Yildiz, E. T. Yasin, and M. Koklu, "A Detailed Analysis of Detecting Heart Diseases Using Artificial Intelligence Methods," *Intelligent Methods In Engineering Sciences*, vol. 2, no. 4, pp. 115-124, 2023, doi: 10.58190/imiens.2023.71.

[47] A. Feyzioglu and Y. S. Taspinar, "Malicious UAVs classification using various CNN architectures features and machine learning algorithms," *International Journal of 3D Printing Technologies and Digital Industry*, vol. 7, no. 2, pp. 277-285, 2023, doi: 10.46519/ij3dptdi.1268605.

[48] B. Gencturk *et al.*, "Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models," *European food research and technology*, vol. 250, no. 1, pp. 97-110, 2024, doi: 10.1007/s00217-023-04369-9.

[49] E. Abdullayeva and H. K. Ornek, "Diagnosing Epilepsy from EEG Using Machine Learning and Welch Spectral Analysis," *Traitement du Signal*, vol. 41, no. 2, 2024, doi: 10.18280/ts.410237.

[50] S. B. Khojasteh and H. K. Ornek, "The Robust EEG Based Emotion Recognition using Deep Neural Network," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 9, no. 4, pp. 191-197, 2021, doi: 10.18201/ijisae.2021473639.