



Article Computer-Aided Diagnosis for Early Signs of Skin Diseases Using Multi Types Feature Fusion Based on a Hybrid Deep Learning Model

Saleh Naif Almuayqil¹, Sameh Abd El-Ghany^{1,2,*} and Mohammed Elmogy³

- ¹ Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakaka 72388, Al-Jouf, Saudi Arabia
- ² Department of Information Systems, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt
- ³ Department of Information Technology, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt
- * Correspondence: saabdelwahab@ju.edu.sa; Tel.: +966-503524918

Abstract: According to medical reports and statistics, skin diseases have millions of victims worldwide. These diseases might affect the health and life of patients and increase the costs of healthcare services. Delays in diagnosing such diseases make it difficult to overcome the consequences of these types of disease. Usually, diagnosis is performed using dermoscopic images, where specialists utilize certain measures to produce the results. This approach to diagnosis faces multiple disadvantages, such as overlapping infectious and inflammatory skin diseases and high levels of visual diversity, obstructing accurate diagnosis. Therefore, this article uses medical image analysis and artificial intelligence to present an automatic diagnosis system of different skin lesion categories using dermoscopic images. The addressed diseases are actinic keratoses (solar keratoses), benign keratosis (BKL), melanocytic nevi (NV), basal cell carcinoma (BCC), dermatofibroma (DF), melanoma (MEL), and vascular skin lesions (VASC). The proposed system consists of four main steps: (i) preprocessing the input raw image data and metadata; (ii) feature extraction using six pre-trained deep learning models (i.e., VGG19, InceptionV3, ResNet50, DenseNet201, and Xception); (iii) features concatenation; and (iv) classification/diagnosis using machine learning techniques. The evaluation results showed an average accuracy, sensitivity, specificity, precision, and disc similarity coefficient (DSC) of around 99.94%, 91.48%, 98.82%, 97.01%, and 94.00%, respectively.

Keywords: skin disorders' diagnosis; deep learning techniques; computer-aided diagnosis system; multi-label classification

1. Introduction

A key issue in healthcare is to discover a disease at its early stage and determine its type to commence with proper treatment. Today, millions of people worldwide are impacted by skin diseases, placing a burden on the health of the individual and the economy of governments, particularly if not treated in its early stages [1]. The American Cancer Society reported that 7180 people died of melanoma (melanoma cancer) in 2021. At the same time, estimates released in the 2022 annual report of the American Cancer Society reported that the number of new cases in the skin disease group (melanoma) was estimated to be approximately 99,780, with the death rate expected to include 7650 people [2].

In contrast, others diseases can cause significant impairment and deformation, due to symptoms such as itching or pain. Moreover, damage to the skin caused by such diseases can also harm an individual's self-confidence and wellbeing [3]. In general, most people believe that some skin sicknesses do not cause significant problems. In contrast, most individuals try to deal with these skin diseases using their own personal strategies. If



Citation: Almuayqil, S.N.; Abd El-Ghany, S.; Elmogy, M. Computer-Aided Diagnosis for Early Signs of Skin Diseases Using Multi Types Feature Fusion Based on a Hybrid Deep Learning Model. *Electronics* 2022, *11*, 4009. https:// doi.org/10.3390/electronics11234009

Academic Editors: George A. Papakostas and Jose Santamaria

Received: 15 October 2022 Accepted: 24 November 2022 Published: 2 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). available medicines are not effective for a specific skin disease, they will simply aggravate the disease further. Additionally, the individual might not even be aware of the seriousness of their skin condition [4].

The conventional approach to diagnosing skin disease is focused on examining dermoscopic images. Diagnoses through dermoscopic images involve dermatologists using multiple dermoscopic equipment, such as the pigment network, dots/globules, and regression of colors. However, this approach has many disadvantages, as it requires high-level dermoscopic instruments, and requires dermatologists to undergo specific training to use dermoscopic equipment, which is time-consuming and requires more effort [5,6]. Moreover, due to the similar patterns and overlapping characteristics of infectious and inflammatory skin diseases, there is often high visual diversity, as well as irregularities in the shape and texture of skin lesions. It has been difficult for dermatologists, especially those who are inexperienced, to identify these slight differences with the naked eye.

With the recent advances in artificial intelligence (AI), especially in the medical field, including medical image processing, AI has become a promising method by which to develop methods for medical image analysis. As a promising AI approach, deep learning networks have proven successful in image analysis, with unique characteristics, and can automatically learn image representation. Consequently, intelligent-solutions-based CAD systems are urgently required to help dermatologists to accurately and quickly diagnose skin diseases, reducing the burden on healthcare systems and the waiting times for medical dermoscopic screenings [7].

Deep learning, particularly convolutional neural networks (CNNs), has outperformed other conventional techniques in human disease diagnostics [8]. As a result, this paper proposes a computer-aided diagnosis (CAD) system for detecting skin disease based on multi-modality data-fusion techniques. This paper is oriented towards the diagnosis of dermoscopic images in addition to analyzing metadata. The proposed CAD system uses AI techniques to fuse and classify metadata and dermoscopic images for accurate disease diagnosis. In this paper, multi-CNN models are used as a backbone for feature extraction from dermoscopic images, which are classified using different machine learning algorithms to obtain a reliable, accurate, and fast skin-disease diagnosis. Therefore, the proposed CAD system relies on the reliable fusion of multiple extracted features to practically help physicians diagnose skin diseases in real-time. The proposed model can result in more accurate diagnoses in metadata and dermoscopic images. The main contributions of this paper can be summarized in the following points:

- A multi-type feature fusion approach based on image and meta-data features is developed for multi-type skin lesion detection with heterogeneous ensemble classifiers. We utilized six convolutional deep learning models, including VGG19, ResNet50, InceptionV3, InceptionResnet, Xception, and DenseNet201, which were pre-trained on the ImageNet dataset. We used transfer learning to train a public skin lesion dataset containing more than 10,000 dermoscopic images.
- 2. The proposed approach automatically extracts features from the processed images, avoiding complex manual feature extraction processes.
- 3. After extracting features, the patient's metadata are fused with the extracted features to diagnose different skin lesions using machine learning classifiers.
- 4. The experimental results show that the proposed approach achieved promising results for the diagnosis of different skin diseases.

The remainder of this paper is structured as follows. Section 2 discusses current work for skin lesion classification and identification. Section 3 describes the proposed CAD system, including the process of dermoscopic image preprocessing, feature extraction, patient metadata concatenation, and skin lesion classification. Section 4 describes the results and provides a comparison with other systems. Finally, Section 5 presents the conclusions and future work.

2. Related Work

Diseases of the skin are a significant concern for the human body. However, some disadvantages can influence their diagnosis, such as high sensitivity, and diagnosis is laborious in nature, time-consuming, and involves complex manual operation. Moreover, they are frequently misclassified due to the degree of similarities across many skin lesions [9].

Therefore, many recent studies on skin disease diagnosis are presented. For example, Shanthi et al. [1] proposed a CNN approach to classifying four types of skin disease: keratosis, acne, eczema herpeticum, and urticaria. The utilized CNN model was AlexNet, which has accuracy values of 85.7%, 92.3%, 93.3%, and 92.8, respectively. Wei et al. [3] proposed a framework that can be used to recognize three types of skin disease. Their model utilized a median filter to remove noise and irrelevant background information in the preprocessing stage. Then, the skin images were segmented using a gray-level co-occurrence matrix (GLCM), while texture and color features were utilized for the feature extraction of different skin disease images. Finally, the support vector machines (SVM) technique was used to classify three skin diseases. The classification accuracy was 90%, 85%, and 95% for dermatitis, herpes, and psoriasis, respectively.

Bajwa et al. [5] proposed a CAD system based on a deep-learning network for skin disease classification. The deep learning models utilized in this work were DenseNet-161, ResNet-152, NASNet, and SE-ResNeXt-101. Their models were trained and tested using the DermNet and ISIC datasets. It achieved an average accuracy of 92.4% for the detection of skin diseases using the DermNet dataset, while the average accuracy was 93% for the ISIC dataset. Wu et al. [10] proposed a system called AIDDA to diagnose inflammatory skin diseases through dermatology analysis. The proposed AIDDA was able to diagnose three diseases: Pso, Ecz, AD, and healthy skin (HC). The authors investigated whether the proposed AIDDA was able to predict inflammation by employing CNN, an EfficientNet-b4 model. Their network was trained and tested on 4740 images. It achieved an average accuracy (ACC) for skin disease diagnoses equal to 95.80% \pm 0.09%, a sensitivity (SEN) equal to 94.40%, and a specificity equal to 97.20%.

Khan et al. [11] proposed a model called DarkeNet19, a pre-trained deep neural network model that used a hybrid optimization technique (EKWO) to pick the most discriminating feature information for classification using the softmax layer. The experimental approach used three datasets, including ISBI2018, HAM10000, and ISBI2019, to obtain 97.1%, 95.8%, and 85.35% accuracy, respectively. Alsaade et al. [12] proposed a hybrid feature extraction technique to process skin lesions. The CNN technique was used in the second system, AlexNet and ResNet50, to efficiently classify skin disorders. According to the results, the approach was tested on HP2 and ISIC 2018 datasets. The ANN model had the best accuracy for PH2 (97.50%) and ISIC 2018 (98.35%) compared to the CNN model.

Ali et al. [13] introduced a deep CNN (DCNN) model for the classification of malignant and benign skin lesions. The DCNN was tested on the HAM10000 dataset and achieved an accuracy of 91.93%. Ameri [14] proposed a model based on the AlexNet model as the pre-trained model for detecting skin cancer using images of skin lesions. The model was evaluated using the HAM10000 database. The accuracy was 84%, the sensitivity was 81%, and the specificity was 88%. Manne et al. [15] introduced a diagnosis model based on ResNet150. The datasets HAM10000 and PH2 were used for model evaluation. HAM1000 had a 98.16% accuracy, whereas PH2 had a 96% accuracy.

Rajput et al. [16] adapted the activation function in the AlexNet model to identify skin cancer diseases in the HAM10000 dataset. The accuracy, recall, and F-score scores increased to 98.20%. Raza et al. [17] proposed an ensemble model for skin lesion classification. Xception, Inceptionv3, InceptionResNet-V2, DenseNet121, and DenseNet201 are stacked ensemble methods that use the principles of transfer learning and fine-tuning. The findings indicated that the suggested model surpasses state-of-the-art procedures, with 97.93 % accuracy. Gouda et al. [18] presented a new method called ESRGAN for preprocessing the ISIC 2018 images, then classified images using Resnet50, InceptionV3, and InceptionResnet

deep learning models. They achieved an 83.2% accuracy rate using the CNN, Resnet50 (83.7%), InceptionV3 (85.8%), and InceptionResnet (84%) models.

Ur Rehman et al. [19] improved pre-trained MobileNetV2 and DenseNet201 deep learning models to more successfully detect skin cancer by adding more convolution layers. The update for both models comprises three stacked convolutional layers. Both benign and malignant classes may be detected using the suggested strategy. Compared to the existing literature methodologies, the suggested modified DenseNet201 model achieves 95.50% accuracy and state-of-the-art performance. Furthermore, the Modified DenseNet201 model has a sensitivity and specificity of 93.96% and 97.03%, a respectively.

Aldhyani et al. [20] proposed a lightweight model to precisely categorize skin lesions. To obtain the best results, dynamic-sized kernels are employed in layers, resulting in extremely few trainable parameters. Furthermore, the suggested model uses both ReLU and leakyReLU activation functions. The model correctly categorized all of the HAM10000 dataset's classes. The model's total accuracy was 97.85%.

Using the HAM10000 dataset, Kousis et al. [21] trained and tested 11 CNN architectures on seven skin lesion classifications. They used data augmentation (during training), transfer learning, and fine-tuning to address the imbalance problem and the significant similarity between photos of particular skin lesions. DenseNet169 outperformed the other 11 CNN architecture variants, achieving an accuracy of 92.25%, a recall (sensitivity) of 93.59%, and an F1-score of 93.27%.

Hasan et al. [22] proposed a dermoscopic Expert (Dermo-Expert), an automated dermoscopic SLC framework. A preprocessing and a hybrid convolutional neural network are combined (hybrid-CNN). Three different feature extractor modules are used in the proposed hybrid CNN to provide better-depth feature maps of the lesion. These single and fused feature maps are then ensembled to predict a lesion class after being categorized using several fully connected layers. Lesion segmentation, augmentation (geometry- and intensity-based), and class rebalancing (penalizing the majority class's loss and merging extra pictures with the minority classes) are all used in the proposed preprocessing stage in their framework. The ISIC-2016, ISIC-2017, and ISIC-2018 datasets are used to test Dermo-Expert, and the DermoExpert acquired an area under the receiver operating characteristic curve (AUC) of 0.96, 0.95, and 0.97, respectively.

The limitations of the current studies can be summarized in the following points. First, most existing studies use only raw images to diagnose skin disease, which cannot efficiently and accurately detect disease. Second, most studies have low accuracy regarding the importance of incorporating multi-features for skin disease diagnosis. To overcome the limitations of these studies, we propose a CAD system that detects various skin diseases. The proposed system extracts the main significant features from two different data modalities: demographic data and dermoscopic images.

3. Materials and Methods

3.1. Dataset Description

The Human Against Machine with the 10,000 Training Pictures (HAM1000) dataset contains 10,015 skin lesion images, divided into seven classes of skin lesions. A complete description of each lesion exists in [23]. All images have the associated demographic information of the patient. Demographic data (gender and age) and the anatomical location of the potentially diagnostically relevant lesion are also inclued. We used the hold-out technique to divide the dataset into training, validation, and testing sets. We randomly used 70% of the dataset as a training set to train all tested models. We used 6% of the dataset as a validation and 24% as a testing set. Therefore, the models were tested on new data that were unseen during training. We summarize different classes of the dataset in Table 1. Figure 1 shows skin lesion images of different class samples from the used dataset.

Class	Training Set	Testing Set	Validation Set	Total
Actinic Keratoses (Solar Keratoses) (akeic)	222	92	13	327
Basal cell carcinoma (BCC)	339	150	25	514
Benign keratosis (BKL)	770	294	49	1113
Dermatofibroma (DF)	68	40	7	115
Melanoma (MEL)	4340	2027	338	6705
Melanocytic nevi (NV)	680	353	66	1099
Vascular skin lesions (VASC)	90	49	3	142
Total	6509	3005	501	10,015

Fable 1 Class distribution in the HAM10000 datas



Figure 1. Samples for different skin disorders from the HAM10000 dataset.

3.2. Framework Architecture and Model Training

In this study, we employed a combination of pre-trained deep learning models and machine learning classifiers to autonomously identify skin lesions to further enhance the generalization capability and accuracy of the deep models. Three machine-learning classification methods were used to separate skin lesions by extracting and storing the bottleneck characteristics from six previously trained models. Figure 2 shows the pipeline of the proposed ensemble feature fusion approach for skin disease diagnosis.



Figure 2. The pipeline of the proposed ensemble feature fusion approach for skin disease diagnosis.

3.3. Stages of the Proposed Ensemble Approach

This paper presents an autonomous ensemble approach to diagnosing skin lesions that fuses the deep features of dermoscopic images with clinical information. The main stages of the proposed approaches include input images, metadata preprocessing, pre-trained deep learning models, and extracted features. Multi-feature fusion is implemented using different state-of-the-art classifiers.

Stage 1 (image pre-processing): In this work, we aim to reduce time-consuming preprocessing steps and improve the generalizability of the CNN design. As a result, we only used two popular preprocessing techniques—image resizing and image normalization processes—when we trained the deep learning model. Image scaling was used because the image dataset may be of various sizes, with significant variations in the image size and intensity. The image normalization process is inevitable because some of the images in the collection of skin lesions may come from various acquisition sources. Each image's pixel intensity might differ significantly due to undesired artifacts such as fluctuations in

picture quality or size, pixel-level noise, bright text, symbols, etc. Data preprocessing is an unavoidable step in this study.

Furthermore, skin images may exhibit variations in image contrast. To circumvent this problem, the contrasts in the training pictures were normalized throughout the training process. As a result, the pictures were normalized by dividing each pixel value by 255. As a result, we set all image intensity values to range between [-1 and 1].

Stage 2 (feature extraction): Six CNN models (Xception, ResNet50, DenseNet201, InceptionV3, VGG19, and InceptionResnet) were utilized to extract features. Compared to retraining the model after fine-tuning, the extracted features were a low-dimensional vector, significantly reducing the model's training time.

The training step of deep learning approaches involves the extraction of deep features, and the testing phase involves evaluating their performance when diagnosing new images. The fact that deep learning models include several layers for feature extraction is their key strength. The primary characteristic of deep learning models includes, for instance, the extraction of geometric features by the first convolutional layer, the detection of edges by the second layer, the extraction of color features by the third layer, the extraction of texture features by the fourth layer, and so forth. The feature extraction in CNN models comprises a convolutional layer and pooling layer pairs that are piled on each other. As the name indicates, the convolution layer uses the convolution technique to transform data. This may be viewed as a set of digital filters—the pooling layer functions as a threshold and dimension reduction layer. The number of features extracted from the CNN models is 2048, and the vector length is N 2048, where N signifies the training image's number.

Stage 3 (metadata pre-processing): This removes missing data from the clinical information. Since most of the demographic features are categorical variables that are represented as 'strings' or 'categories' and are finite in number, these categories' features are converted to the categorical data format, using one-hot encoding. For each level of a categorical feature, we created a new variable. Each category was mapped with a binary variable containing either 0 or 1. For example, sex type was converted into two new features (male and female categories). Here, 0 represents the absence, and 1 represents the presence of that category. Moreover, the numeric demographic features, such as age, were normalized.

Stage 4 (feature concatenation): This stage is responsible for fusing the image features and metadata into a single feature vector. First, we fed the preprocessed skin disease images into the CNN models. The CNN models extract deep features through convolutional, pooling, and auxiliary layers. This produced 2048 deep features and stored them in feature vectors with a size of 6509×64 . The number of demographic features is 6509×5 . The concatenated feature vector is 6509×65 . After converting the categorical features of demographic features into one-hot encoding, the length of the concatenated vector will be 6509×85 features

Stage 5 (skin lesion classification): This feeds the generated concatenated features to different machine learning classifiers. Finally, all skin lesion images were classified into seven classes.

3.4. Transfer Learning (TR)

In machine learning applications involving image categorization, TR is employed. This entails reusing knowledge from a trained CNN model that performed well in the original domain [24]. The weights for this trained model were derived from sizable, labeled datasets such as ImageNet. After the model is trained, the generated weights may be applied to a particular dataset in a new domain. TR may be utilized as a model for the retrieved features. We removed the final fully connected layer and treated the remaining CNN for the feature extraction process.

The pre-trained CNN models were trained on the ImageNet dataset, which includes 1000 classes. The first group of layers in each CNN model generates low-level features for the trained classes, which can be used in any application. The final fully connected

layers are used to classify the 1000 classes as they were pre-trained on the ImageNet dataset. For this purpose, these deep architectures' last fully connected layers were removed and replaced by a fully connected layer to ensure they were adaptable for our case study. Softmax was used as a classification layer. Thus, these deep architectures were adapted for the detection of multiple skin disorders.

As a result, TR was utilized to assort images of skin lesions. Therefore, our research reused pre-trained models trained on the ImageNet dataset for skin lesion classification. In this work, we deleted the last fully connected layer and replaced it with a new layer based on the number of skin disease classes. The new fine-tuned model was then trained using transfer learning. We trained a model using 70% skin photos. We set the learning rate to 1×10^{-4} , the batch size to 64, the dropout factor to 0.5, and the weight decay value to 1×10^{-5} for training. Once we learned this new model (GAP), we reduced features from the global average pooling layer. The number of extracted features on this layer was 2048, and the vector length was N 2048, where N signifies the training image's number. We also employed early stopping to improve the trained neural network's generalization performance. In this paper, we tested the performance of six state-of-the-art networks: InceptionV3 [25], ResNet50 [26], VGG19 [27], Xception [28], DenseNet201 [29] and Inception-ResNet [30]. We chose these DL models because they were the most suitable models for diagnosing skin disorders with high performance. We removed each model's top layer and froze the former convolutional layer. Then, dense layers were added at the bottom. A dropout layer was added to the dense layer. Then, the L1-norm was added to avoid model overfitting. Finally, the loss was defined as a categorical cross-entropy. Adam was chosen as an optimizer for the utilized models. In the following, the architecture of these models is briefly outlined.

3.5. Convolution Deep Learning Model for Feature Extraction Process

The VGG19 model consists of 19 layers. It contains 16 convolutional layers and 3 fully connected layers trained on the ImageNet. It uses multiple 3×3 convolutional filters and a stride of 1, followed by multiple non-linear layers. Five MaxPool layers and 1 SoftMax layer are applied to reduce the feature size and achieve high accuracy in image classification. The last three thick layers in block 6 have dimensions of 4096, 4096, and 1000, respectively. VGG categorizes the supplied images into 1000 separate groups. This research has seven output classes, and the dimension of fc8 is set to seven.

The InceptionV3 model is utilized to enhance computing resources by emphasizing the importance of memory management and the model's computing power. Therefore, it speeds up calculations and reduces the number of parameters that are used. The model's architecture has 48 layers with skipped connections trained on millions of images, including 1000 classes. It is repeated with max-pooling to reduce the feature dimensions.

The ResNet50 model is a common CNN model at present. It uses a residual structure that supplies more effective training and a more straightforward gradient flow. The network can accept input images with a height and width and channel of $224 \times 224 \times 3$. The ResNet50 model is divided into four stages. Every ResNet design uses 77 and 33 kernel sizes for initial convolution and max-pooling, respectively. Following then, Stage 1 of the network begins with three Residual blocks, each comprising three layers. The kernels used to conduct the convolution operation in all three stages of the stage 1 block are 64, 64, and 128.

The denseNet201 model utilizes dense connections, which replace the direct connections of the hidden layers. Therefore, it allows for network features to be re-utilized. It also maximizes information transmission between layers in the model. The network takes 224 224 input and runs this via a convolution and max pool layer. Then, four dense blocks are interspersed with three transition blocks, and 14×14 feature maps are output.

The Xception model assumes that cross-channel and spatial correlation can be separated. The same parameter numbers are utilized, with a better performance. The depthwise separable convolutions used in the Xception architecture, which expands the Inception architecture, replace the normal Inception modules. Instead of dividing the input data into several compressed pieces, it maps the spatial correlations for each output channel individually, before performing a 1×1 depthwise convolution to capture cross-channel correlation.

The Inception-ResNet model combines the inception structure with residual connections. It has multiple-sized convolution filters that are trained on millions of images to avoid the degradation problem. All layers developed before the FC layer contain the fundamental components of the Inception-ResnetV2 architecture. The kernel size of the Conv (convolutional layer), Pool (pooling layer), or FC is the patch size. The stride is the distance between two procedures assigned to number 2 in the experiment. Softmax is a network classification function, and Filter contact is a module that connects many Conv. The Inception-ResNetV2 model incorporates three main inception modules: Inception-ResNet-A, Inception-ResNet-B, and Inception-ResNet-C. These modules are in charge of lowering the number of parameters in tiny Conv layers (for example, 17, 71) and creating discriminating features. Each module has its own Conv and pool layers.

Table 2 lists the values of the parameters of the tested DL models. We believe that adding a trainable parameter via Hyper-Parameter Optimization (HPO) algorithms might increase the model's computational complexity. Therefore, the model parameters are the hyperparameters obtained using the trial-and-error methodology. Then, we list the values for the best hyperparameter results. Table 3 lists all used hyperparameters that achieved the highest performance for each tested model.

Model	Number of Parameters	Number of Layers	Input Image Size	Kernel Size
Xception	30,375,912	71	71×71	3×3
Resnet50	27,857,668	50	71 imes 71	7 imes 7
DenseNet201	22,331,392	121	71×71	5×5
InceptionV3	21,817,127	48	75 imes 75	3×3
VGG19	20,027,975	19	75 imes 75	3×3
InceptionResNetV2	54,347,495	164	71×71	3×3

Table 3. Values of the hyperparameters used in the CNN model architectures.

Hyper-Parameters	Value
Number of epoch	100
Batch size	64
Pooling	Global average pooling
Optimizer	Adam
Initial learning rate	$1 imes 10^{-4}$
Dropout	0.5
Patience	10
Loss function	Categorical-cross entrop

4. Results

4.1. Implementation Details

Our implementation was completed on Kaggle. This offers full access to Keras library. It also offers free access to NVidia K80 GPUs in kernels with 11.86 GB RAM and Intel(R) Xeon(R) CPU 2.30 GHz with 16 cores. We uploaded the source code of all models on Kaggle, and this is available to anyone. The source code will be available on request from the corresponding author after manuscript publication.

4.2. Evaluation Metrics

The proposed model was evaluated using five various performance metrics. The evaluation metrics are precision (PRE), disc similarity coefficient (DSC), accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC).

$$PRE = \frac{TP}{TP + FP}$$
(1)

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}$$
(2)

$$ACC = \frac{TP + TN}{(TP + FP + TN + FN)} \times 100$$
(3)

$$SEN = Recall = \frac{TP}{TP + TN}$$
(4)

$$SPE = \frac{TN}{TN + FP}$$
(5)

In the above equations, TP presents the number of positive instances that are labeled correctly. FP presents the number of positive instances that are mislabeled. TN presents the number of negative instances that are correctly labeled. FN presents the number of negative instances that are mislabeled. PRE represents the samples' proportion, classified as positive or actually positive. DSC can be introduced as a harmonic average of precision and recall.

4.3. Results

4.3.1. Classification Results on HAM10000 Dataset without Metadata

Six pre-trained models were used in the first experiment to classify skin lesion images. Table 4 presents the results of a detailed comparison of six different pre-trained models using different evaluation metrics.

Model	SEN (%)	SPE (%)	PRE (%)	DSC (%)	ACC (%)	Training Time (s)	Testing Time (s)
Xception	92.29	98.68	93.69	92.96	96.83	328.1691	12.1131
Resnet50	93.23	99.02	95.49	94.44	96.93	353.5990	11.6892
DenseNet201	92.76	98.88	97.45	94.74	97.16	594.0691	16.3945
InceptionV3	91.52	98.43	93.44	92.46	96.62	376.2660	12.8734
VGG19	92.96	99.09	91.31	92.11	96.97	411.8653	11.1327
InceptionResnet	92.79	98.87	94.87	93.28	96.65	746.5525	26.3085

Table 4. The performance evaluation of the CNN classifiers without metadata.

The overall average ACC for DenseNet201 was 97.16%. The model's impact was steadier when compared to various other models. Moreover, the DenseNet201 model exhibits high accuracy (average PRE is 97.45%) and performs well in the DSC rate (94.74%), which is also worth noting. Three machine-learning classification techniques were utilized to classify the extracted features from the previously trained CNN models to distinguish skin lesions. We employed a combination of deep learning classifiers that had already been pre-trained, as well as machine learning classifiers, to autonomously identify skin lesions to further enhance the generalization capability and accuracy of the deep models. The confusion matrix of each deep learning classifier is displayed in Figure 3.



Figure 3. The confusion matrices for deep learning classifiers.

Figure 3 illustrates the confusion matrix for the six CNN models for the skin disease classes. The support of the class akeic is 92, the support of class akeic class is 150 images, BKL class is 294 images, the DF class is 40 images, MEL class is 2027 images, NV is 353 images, and the support of VASC class is 49 images.

For class VASC, the best model is DenseNet201, as it correctly classified 87 images out of 92, while only five images were misclassified. The best model for the akeic class is DenseNet201, which correctly classified 145 images, while only five were misclassified. For class BKL, the best modes are DenseNet201 and ResNet50, which can classify 278 images correctly. The VGG 19 is the best model that classified the DF class, as it correctly classified 37 out of 40 images. Again, the DenseNet201 is the best model to classify the MEL class, as it can correctly classify 2013 out of 2027 images. For class NV, the best model is VGG19, as it can correctly classify 320 out of 353 images. Finally, models VGG 19, Xception, and Resnet50 can best classify the VASC class.

Additionally, as shown in Figure 4, we calculated the receiver operator characteristic (ROC) curve for True Positive Rate (TPR) vs. False Positive Rate (FPR) for the classifier that solely uses deep image features. We represented the values of area Under the Curve (AUC) as class-wise boxplots for the HAM10000 dataset for each skin lesion.



Figure 4. The ROC curves for the deep learning classifiers.

Figure 4 shows that the DensNet201 is the best model to obtain the ROC curve for all skin disease classes, as it obtains an AUC that is equal to 1 for all classes. The worst model is the Xception model, which obtains an AUC of less than 1 for all skin disease classes.

4.3.2. Classification Result of Combined Classifier

The obtained results for the three machine learning classifiers and deep feature extractors are summarized in Tables 5–10. These three machine-learning classifiers are SVM, random forest (RF), and logistic regression (LR). These tables show that the assessment metrics were enhanced compared to the conventional CNN models. It is important to note that all the pre-trained models using various classifiers performed quite well. Every hybrid technique improved model performance when including clinical meta-data as an additional feature, as shown in Tables 5–10. Compared to simply using image features, employing metadata results in a more significant performance gain for Sensitivity, Precision, and F-score across all hybrid models. Despite the overall improvement, ACC and SPE exhibited modest improvements compared to the other metrics in all studies. The bold values in each table indicate the highest value in all metrics.

Table 5. The performance evaluation for VGG19 combined with different classifiers.

Model	SEN(%)	SPE(%)	PRE(%)	DSC(%)	ACC(%)	Test Time (s)
VGG19 + RF	93.39	99.72	97.80	95.39	99.73	9.5132
VGG19 + LR	97.90	99.95	98.79	98.12	99.92	9.4659
VGG19 + SVM	90.74	98.87	94.81	92.57	99.19	10.6649

Table 6. The performance evaluation for InceptionV3 combined with different classifiers.

Model	SEN(%)	SPE(%)	PRE(%)	DSC(%)	ACC(%)	Test Time (s)
InceptionV3 + RF	92.72	98.73	95.61	94.08	99.01	9.8076
InceptionV3 + LR	99.49	99.83	99.17	99.24	99.90	9.7577
InceptionV3 + SVM	91.14	98.37	95.60	93.26	98.84	11.0787

Table 7. The performance evaluation for ResNet50 combined with different classifiers.

Model	SEN(%)	SPE(%)	PRE(%)	DSC(%)	ACC(%)	Test Time (s)
Resent50 + RF	93.15	99.59	97.18	94.68	99.44	10.6397
Resent50 + LR	96.83	99.77	98.01	97.16	99.72	10.5896
Resent50 + SVM	90.45	98.96	95.47	92.31	98.99	11.8022

Table 8. The performance evaluation for DenseNet201 combined with different classifiers.

Model	SEN(%)	SPE(%)	PRE(%)	DSC(%)	ACC(%)	Test Time (s)
DenseNet201 + RF	92.50	99.68	97.66	94.58	99.49	9.8076
DenseNet201 + LR	95.08	99.81	98.23	96.99	99.71	9.7577
DenseNet201 + SVM	91.55	99.22	96.49	93.37	99.21	11.0787

Table 9. The performance evaluation for Xception combined with different classifiers.

Model	SEN(%)	SPE(%)	PRE(%)	DSC(%)	ACC(%)	Test Time (s)
Xception + RF	94.09	99.41	97.62	95.54	99.93	10.1391
Xception + LR	98.16	99.96	99.24	98.78	99.36	10.0877
Xception + SVM	91.48	98.82	97.01	93.90	99.09	11.3043

Table 10. The performance evaluation for InceptionResnetV2 combined with different classifiers.

Model	SEN(%)	SPE(%)	PRE(%)	DSC(%)	ACC(%)	Test Time (s)
InceptionResnetV2 + RF	93.25	99.51	97.43	95.26	99.34	19.9461
InceptionResnetV2 + LR	98.84	99.97	99.59	99.16	99.94	21.1139
InceptionResnetV2 + SVM	92.21	98.88	96.85	94.05	98.99	21.5670

Table 5 shows that the use of VGG19 as a feature-extractor and LR as a classifier achieved the best results in all performance metrics (SEN, SPE, PRE, DSC, and ACC). Moreover, the testing time also achieved the best value using the same combination.

Table 6 shows that the combination of InceptionV3 as a feature-extractor and LR as a machine learning classifier also achieved the best results for all metrics. The testing time also achieved the smallest value using the same combination.

Table 7 shows that the combination of Resent50 as a feature extractor and LR as a machine learning classifier also achieved the best results for all metrics. The testing time also achieved the smallest value using the same combination.

Table 8 shows that the combination of DenseNet201 as a feature-extractor and LR as a machine learning classifier also achieved the best results for all metrics. The testing time also achieved the smallest value using the same combination.

Table 9 shows that the combination of Xception as a feature-extractor and LR as a machine learning classifier also achieved the best results for all metrics and the testing time achieved the smallest value using the same combination.

Table 10 shows that the combination of InceptionResnetV2 as feature-extractor and LR as machine learning classifier achieved the best results for all metrics and the testing time is also achieved the smallest value using the combination of InceptionResnetV2 + RF.

The best SEN and DSC metrics were obtained using hybrid InceptionV3 + LR. The results reached 99.49% and 99.24%, respectively, as shown in Table 6. The best SPE, PRE, and ACC results were 99.97%, 99.59%, and 99.94%, respectively. These metrics are shown in Table 10 using hybrid InceptionResnet + LR.

4.3.3. Comparative Study with the State-of-the-Art Systems

Table 11 shows a comparative analysis of the interoperability of pre-trained models using state-of-the-art methods. In general, our hybrid methods outperform other state-of-the-art methods. The worst results, as in [14], used the AlexNet model for classification tasks, as illustrated in related work. In general, the results demonstrate the superiority of the proposed methodology for hybrid data fusion and machine learning with deep feature integration, achieving a remarkably higher average accuracy results of 99.44% in the case of the DenseNet121, model for feature extraction, and logistic regression (LR), for the classification task.

Work	Year	Dataset	Method	ACC	SEN	SPE
Prposed Syatem	2022	HAM10000	DenseNet201 + logistic regression (LR)	99.94%	98.84%	99.97%
Bajwa et al. [5]	2020	DermNet and ISIC datasets		92.4% for DermNet, 93%, for ISIC		
Ameri [14]	2020	HAM10000		84%	81%	88%
Manne et al. [15]	2020	HAM10000 and PH2		98.16%, And 96%		
Khan et al. [11]	2021	HAM10000, ISBI2018, and ISBI2019		95.8%, 97.1%, and 85.35%,		
Alsaade et al. [12]	2021	HP2 and ISIC 2018		(97.50%), (98.35%)		
Ali et al. [13]	2021	HAM10000		91.93%		
Rajput et al. [16]	2022	HAM10000		98.20%	98.20%	98.20%
Raza et al. [17]	2022	Figshare		97.93%	97.83%	97.50%
Gouda et al. [18]	2022	ISIC 2018		85.8%		

Table 11. The comparison of results of our proposed methodology with some related work.

5. Conclusions

Skin diseases are widespread around the world, affecting the health of patients, and the cost of healthcare services that are provided through the government. Therefore, due to the positive effect of the diagnosis of skin diseases, this article proposes a computer-aided diagnosis system. The proposed system relies on deep learning and machine learning algorithms and utilizes dermoscopic images. According to the obtained results, the proposed system achieved promising results, which may encourage the authors to evaluate the proposed system using other types of skin diseases and other disease categories. The limitation of this work is its lack of dimensionality reduction methods to select the best features among all extracted features. In future work, we will test other deep-learning techniques to improve the classification accuracy of skin disorders. In addition, we will test our proposed system by using other benchmark datasets with different skin disorders.

Author Contributions: Conceptualization, S.N.A., S.A.E.-G. and M.E.; Project Administration, S.N.A.; Supervision, S.N.A.; Methodology, S.A.E.-G.; Formal Analysis, M.E.; Software, S.A.E.-G.; Writing—Review and Editing, S.N.A., S.A.E.-G. and M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deanship of Scientific Research at Jouf University grant number (DSR2020-04-2616).

Data Availability Statement: Publicly available datasets were analyzed in this study. These dataset can be found at https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000, last access 30 September 2022.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at Jouf University for funding this work through research grant no (DSR2020-04-2616).

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

- Shanthi, T.; Sabeenian, R.S.; Anand, R. Automatic Diagnosis of Skin Diseases Using Convolution Neural Network. *Microprocess. Microsyst.* 2020, 76, 103074. [CrossRef]
- 2. American Cancer Society. Key Statistics for Melanoma Skin Cancer. 2022. Available online: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html (accessed on 27 October 2022).
- Wei, L.-S.; Gan, Q.; Ji, T. Skin Disease Recognition Method Based on Image Color and Texture Features. *Comput. Math. Methods* Med. 2018, 2018, 8145713. [CrossRef] [PubMed]
- 4. Amarathunga, A.A.L.C.; Ellawala, E.P.W.C.; Abeysekara, G.N.; Amalraj, C.R.J. Expert System for Diagnosis Of Skin Diseases. *Int. J. Sci. Technol. Res.* 2015, *4*, 174–178.
- Bajwa, M.N.; Muta, K.; Malik, M.I.; Siddiqui, S.A.; Braun, S.A.; Homey, B.; Dengel, A.; Ahmed, S. Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks. *Appl. Sci.* 2020, 10, 2488. [CrossRef]
- Monisha, M.; Suresh, A.; Rashmi, M.R. Artificial Intelligence Based Skin Classification Using GMM. J. Med. Syst. 2019, 43, 3. [CrossRef]
- Kassem, M.A.; Hosny, K.M.; Damaševičius, R.; Eltoukhy, M.M. Machine Learning and Deep Learning Methods for Skin Lesion Classification and Diagnosis: A Systematic Review. *Diagnostics* 2021, 11, 1390. [CrossRef]
- 8. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 444. [CrossRef]
- Hosny, K.M.; Kassem, M.A.; Fouad, M.M. Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet. J. Digit Imaging 2020, 33, 1325–1334. [CrossRef]
- 10. Wu, H.; Yin, H.; Chen, H.; Sun, M.; Liu, X.; Yu, Y.; Tang, Y.; Long, H.; Zhang, B.; Zhang, J.; et al. A Deep Learning, Image-Based Approach for Automated Diagnosis for Inflammatory Skin Diseases. *Ann. Transl. Med.* **2020**, *8*, 581. [CrossRef]
- 11. Khan, M.A.; Akram, T.; Sharif, M.; Kadry, S.; Nam, Y. Computer decision support system for skin cancer localization and classification. *Comput. Mater. Contin.* **2021**, *68*, 1041–1064.
- Alsaade, F.W.; Aldhyani, T.H.H.; Al-Adhaileh, M.H. Developing a Recognition System for Diagnosing Melanoma Skin Lesions Using Artificial Intelligence Algorithms. *Comput. Math. Methods Med.* 2021, 2021, 9998379. [CrossRef]
- 13. Ali, S.; Miah, S.; Haque, J.; Rahman, M.; Islam, K. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Mach. Learn. Appl.* **2021**, *5*, 100036. [CrossRef]
- 14. Ameri, A. A deep learning approach to skin cancer detection in dermoscopy images. *J. Biomed. Phys. Eng.* **2020**, *10*, 801–806. [CrossRef]

- 15. Manne; Ravi; Kantheti, S.; Kantheti, S. Classification of Skin cancer using deep learning, Convolutional Neural Networks-Opportunities and vulnerabilities-A systematic Review. *Int. J. Mod. Trends Sci. Technol.* **2020**, *6*, 2455–3778.
- Rajput, G.; Agrawal, S.; Raut, G.; Vishvakarma, S.K. An accurate and noninvasive skin cancer screening based on imaging technique. *Int. J. Imaging Syst. Technol.* 2022, 32, 354–368. [CrossRef]
- 17. Raza, R.; Zulfiqar, F.; Tariq, S.; Anwar, G.B.; Sargano, A.B.; Habib, Z. Melanoma Classification from Dermoscopy Images Using Ensemble of Convolutional Neural Networks. *Mathematics* **2022**, *10*, 26. [CrossRef]
- Gouda, W.; Sama, N.U.; Al-Waakid, G.; Humayun, M.; Jhanjhi, N.Z. Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning. *Healthcare* 2022, 10, 1183. [CrossRef]
- Rehman, M.Z.U.; Ahmed, F.; Alsuhibany, S.A.; Jamal, S.S.; Ali, M.Z.; Ahmad, J. Classification of Skin Cancer Lesions Using Explainable Deep Learning. *Sensors* 2022, 22, 6915. [CrossRef]
- Aldhyani, T.H.H.; Verma, A.; Al-Adhaileh, M.H.; Koundal, D. Multi-Class Skin Lesion Classification Using a Lightweight Dynamic Kernel Deep-Learning-Based Convolutional Neural Network. *Diagnostics* 2022, 12, 2048. [CrossRef]
- Kousis, I.; Perikos, I.; Hatzilygeroudis, I.; Virvou, M. Deep Learning Methods for Accurate Skin Cancer Recognition and Mobile Application. *Electronics* 2022, 11, 1294. [CrossRef]
- 22. Hasan, K.; Elahi, T.E.; Alam, A.; Jawad, T.; Martí, R. DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation. *Inform. Med. Unlocked* **2022**, *28*, 100819. [CrossRef]
- Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Sci. Data* 2018, *5*, 180161. [CrossRef] [PubMed]
- 24. Baker, A.; Nermeen; Zengeler, N.; Handmann, U. A Transfer Learning Evaluation of Deep Neural Networks for Image Classification. *Mach. Learn. Knowl. Extr.* 2022, *4*, 22–41. [CrossRef]
- Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
- 27. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
- Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
- 29. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. Available online: https://arxiv.org/abs/1608.06993 (accessed on 14 October 2022).
- Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.