

Article

Classification of Skin Lesions Using Weighted Majority Voting Ensemble Deep Learning

Damilola A. Okuboyejo and Oludayo O. Olugbara * 

MICT SETA 4IR Center of Excellence, Durban University of Technology, Durban 4000, South Africa

* Correspondence: oludayo@dut.ac.za

Abstract: The conventional dermatology practice of performing noninvasive screening tests to detect skin diseases is a source of escapable diagnostic inaccuracies. Literature suggests that automated diagnosis is essential for improving diagnostic accuracies in medical fields such as dermatology, mammography, and colonography. Classification is an essential component of an assisted automation process that is rapidly gaining attention in the discipline of artificial intelligence for successful diagnosis, treatment, and recovery of patients. However, classifying skin lesions into multiple classes is challenging for most machine learning algorithms, especially for extremely imbalanced training datasets. This study proposes a novel ensemble deep learning algorithm based on the residual network with the next dimension and the dual path network with confidence preservation to improve the classification performance of skin lesions. The distributed computing paradigm was applied in the proposed algorithm to speed up the inference process by a factor of 0.25 for a faster classification of skin lesions. The algorithm was experimentally compared with 16 deep learning and 12 ensemble deep learning algorithms to establish its discriminating prowess. The experimental comparison was based on dermoscopic images congregated from the publicly available international skin imaging collaboration databases. We propitiously recorded up to 82.52% average sensitivity, 99.00% average specificity, 98.54% average balanced accuracy, and 92.84% multiclass accuracy without prior segmentation of skin lesions to outstrip numerous state-of-the-art deep learning algorithms investigated.



Citation: Okuboyejo, D.A.; Olugbara, O.O. Classification of Skin Lesions Using Weighted Majority Voting Ensemble Deep Learning. *Algorithms* **2022**, *15*, 443. <https://doi.org/10.3390/a15120443>

Academic Editor: Maryam Ravan

Received: 3 October 2022

Accepted: 20 November 2022

Published: 24 November 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/>).

Keywords: deep learning; ensemble learning; learning algorithm; lesion classification; skin disease; skin lesion

1. Introduction

Malignant skin lesions such as melanoma manifest a serious health risk with increasing incidence. The ability of melanoma to rapidly grow and widely proliferate to other parts of the human body is a unique characteristic that makes it one of the deadliest skin cancer diseases [1]. It is a fatal malignant skin tumor resulting from the improper synthesis of melanin because of unpredictable disorders in the melanocytic cell [2]. The good news though is that fatality of skin lesions can be prevented to a large degree if the disease is detected early before proliferation. The high mortality rate associated with melanoma makes it necessary to detect the disease at its early stages to improve health treatment as quickly as possible [3]. Statistics have indicated that over 9000 deaths are recorded yearly from melanoma infection, which warrants the necessity to provide a sustainable and easily reproducible method of early detection [4–7]. It has been judged to be the fifth most common skin cancer occurring among males, seventh in females, and second among young adults ranging from 15–29 years of age [8–10]. Moreover, given that melanoma accounts for 75–79% of skin cancer-related deaths, early detection has been proposed by medical practitioners as one of the vivacious keys to preventing untimely death associated with the disease [8,11]. This is one of the important reasons that manifold researchers are interested

in obtaining accurate automated systems that can assist in the early detection and diagnosis of the disease [12].

The diagnostic procedure for skin lesions often relies heavily on dermatologists who use visual pattern recognition with the aid of dermoscopic devices to identify lesions through rules such as asymmetry, border irregularity, color variation, the diameter of the lesion, a 7-point checklist, and Menzies score. However, the associated complexity, high subjectivity, and a great deal of experience in interpreting dermoscopic evaluation results present a serious concern [8,13,14]. Recent reports have noted the poor prognosis from expert dermatologists that can sometimes lead to expensive and unnecessary excision [15]. These challenges necessitate a reliable automated system to assist with the second opinion [4]. There are numerous methods in the literature to automatically determine whether a skin lesion is malignant or not, and most studies have reported promising results. The application of machine learning methods is widely accepted in recent times as having the potential capacity to act as a second opinion for the automatic classification of skin lesions [16]. Machine learning (ML) applies artificial intelligence (AI) methods to provide computing systems with the capability to learn the experience of performing non-trivial tasks. The frequently used ML methods include decision trees (DTs), support vector machines (SVMs), artificial neural networks (ANNs), and multiple regression analysis (MRA). Specifically, ANN is a widely known ML method that tends to simulate biological neural networks using collections of connected nodes to model biological neurons in the human brain.

There are copious applications of ANNs for classifying skin lesions with encouraging results [17–19]. Deep neural networks (DNNs) have recently evolved as variations of ANNs with consideration for more dense hidden layers to improve classification performance. The process of training DNNs for learning to perform a non-trivial assignment is termed deep learning (DL). Frequently used DNNs include convolution neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs). The CNN applications in computer vision have been widely lauded with increasing success, because of their ability to solve signal translation problems by convolving each input signal with a kernel detector to train and learn spatial relations among image pixels [20,21]. However, the classification of skin diseases is a difficult task because of the strong similarities in common symptoms for which AI methods were recommended to improve the accuracy of dermatology diagnosis [16]. There are several claims in the literature on the outstanding results produced by CNN methods on skin lesion images acquired from a clinical procedure, dermoscopy, mammography, colonography, or histopathology when compared to experienced dermatologists [22–28]. Moreover, the increasing number of publicly available image datasets has provided a huge opportunity for better computer vision possibilities [29]. The application of DL methods is one possibility to assist in predicting the correlated relationships among malignant skin tumors and many diseases in patients [30].

The purpose of the present study was to improve the overall performance of a multi-class skin lesion detection process using an ensemble deep learning that utilizes the strength of different heterogeneous models with the capability of parallel processing. The rest of the paper is structured concisely as follows. Section 2 summarizes the related studies. Section 3 describes the materials and methods of the study, including the introduced ensemble deep learning algorithms. Section 4 explains the experimental results of evaluating the introduced algorithms against 28 state-of-the-art algorithms. The paper is briefly concluded in Section 5.

2. Related Studies

The improved LeNet method that uses adaptive piecewise linear activation as an alternative to the conventional activation function was proposed to improve the classification of melanoma skin lesions [31]. The CNN framework termed Dermo-Doctor was recently developed to detect skin lesions [32]. The framework performs lesion segmentation, uses two encoders, and fuses feature maps from the encoders to obtain the desired classification

result. This past work was further expanded in a proposal called DermoExpert using a hybrid of CNNs, class rebalancing, and transfer learning [33]. The classification of skin lesions network (CSLNet) was suggested based on a specialized deep convolutional neural network (DCNN) [34]. The backpropagation algorithm was used to train the DCNN model to evolve the network weights. A leaky rectified linear unit (LeakyReLU) was used as an activation function in the convolution layers, while SoftMax was leveraged as an activation function in the output layer. The L2 normalization scheme was used to regularize the kernel in the dense layer to prevent the possibility of over-fitting, often associated with a learning method.

Most classification problems can be generally categorized as unary, binary, or multiclass [35]. Multiclass classification is the processing task of assigning a data sample to exactly one class among three or more classes. The task is particularly challenging for an imbalanced dataset of medical images that present multiple undesirable artifacts such as vignettes, ruler markings, specular reflections, shadows, and hair shafts. Moreover, a skewed dataset distribution can make conventional machine learning algorithms ineffective, especially when predicting examples of a minority class. A multiclass problem is sometimes resolved directly or by further reducing the problem into a series of two-class subproblems. Direct resolution can be achieved with the application of SoftMax or Sigmoid function. The use of either one-vs-one (OvO) and one-vs-rest/one-vs-all (OvR/OvA) schemes can be used to further break a multiclass problem into multiple binary subproblems [36–38]. Multiple sub-datasets are used for training both OvO and OvA and the prediction of input data is passed through each generated model where a model with the highest probability gives the final prediction. The OvO is a heuristic algorithm that splits a multiclass problem into multiple binary subproblems. The source dataset is divided into one sub-dataset per each class versus every other class to generate $n(n - 1)/2$ binary classification models where n is the number of classes. However, this could lead to countless sub-datasets that might be difficult to scale, especially for large source datasets. The OvA riffs a multiclass problem into several binary subproblems by dividing the source dataset into one sub-dataset for each class versus the rest of the classes to generate n binary classification models. The OvA is less verbose in comparison to the OvO for the required sub-datasets to be created, but it requires the creation of a model for each class. This requirement might be undesirable for large source datasets or multiclass problems with a considerable number of classes.

The introduction of deep learning has made a direct resolution of multiclass problems relatively exciting. The present study has followed the ensemble deep learning approach of a direct multiclass resolution using SoftMax to avoid the innate concerns associated with OvO and OvA. Multiclass classification can be a single-label prediction outcome or multi-label in nature. Single-label classification indicates that the outcome of a prediction generates exactly one class. A single SoftMax layer is often used as the last layer for neural networks that require a single-label prediction outcome. However, multi-label classification indicates that the outcome of a prediction generates two or more classes. Multiple sigmoid layers are typically used as the last layer for neural networks that require multiple-label prediction outcomes. The classification problems in the domain of skin lesion diagnosis are naturally multiclass with a single-label outcome. This is because a skin lesion can seldom be of two disease types concomitantly, except when dealing with different semantics such as a skin lesion that is considered both dermatofibroma and benign. Most of the skin lesion classification methods reported in the literature are mostly limited to binary classes either between melanoma and melanocytic nevi or between broader malignant and benign lesions. However, in recent times, some attempts have been made to solve multiclass skin lesion classifications [19,39–42]. Literature has recorded the use of multiple deep learning methods in an ensemble framework to mimic how a medical practitioner leverages a consensus opinion through the pearls of wisdom of other practitioners to corroborate the outcome of a medical diagnosis [43–46].

Ensemble learning belongs to the state-of-the-art approach that combines the predictions of manifold base learners into a strong algorithm to increase the overall classification

performance and decrease the peril of obtaining a local minimum [47]. It addresses the inherent challenges such as the high variance of being sensitive to training data and bias frequently associated with single learners in a deep learning framework. Several previous works have achieved improved results, reduced classification errors and obtained better generalization in skin lesion classification with ensemble learning [29,48–52]. Ensemble learning can be achieved by aggregating the prediction from each learner to build a meta-model that produces the final prediction. It includes boosting (AdaBoost, GradientBoost, and eXtreme GradientBoost), bagging (bootstrap aggregation like the random forest), voting (simple majority, averaged majority, or weighted majority), and stacking (stacked generalization for layer-2 meta-model), each with its intrinsic merits and demerits. The four approaches can use heterogeneous algorithms but bagging and boosting are typically used on multiple subsets of training data on homogenous learners, whereas voting and stacking mostly use heterogeneous learners. The aggregation of multiple predictions in ensemble learning can be done by a voting mechanism either by computing maximum voting, average voting, or weighted average voting of all the predicted classes.

The authors in [39] proposed an ensemble learning strategy termed the global-part CNN model with data-transformed ensemble learning (GP-CNN-DTEL). It involves training a global CNN (G-CNN) with downscaled dermoscopic images to produce a classification activation map (CAM) and the training of a network called part CNN (P-CNN) with the resulting CAM. In [51], an efficient lightweight melanoma classification network based on MobileNet and DenseNet121 was proposed to improve the ability of feature discrimination, and recognition accuracy of lightweight networks while utilizing a small number of model parameters. The study in [52] proposed a multi-scale multi-CNN (MSM-CNN) fusion based on a three-level ensemble strategy. This was achieved by fusing the results of three fine-tuned networks of EfficientNetB0, EfficientNetB1, and SeReNeXt-50 with cropped images at six scales. The authors asserted that image cropping is a better strategy when compared to image resizing for performing skin lesion classification based on transfer learning. The study in [53] fused the outputs of the classification layers of four different DNN architectures of GoogLeNet, AlexNet, residual network (ResNet), and visual geometry group network (VGGNet) into one output. The final classification was achieved based on the weighted outputs of the CNN members. In [54], an ensemble learning encompassing Inception-v4, ResNet-152, and DenseNet-161 was used for malignant classification in a manner reminiscent of mimicking the real-world approach. In the approach, a specialist will typically consult other specialists to cross-reference and double-check diagnosis before engaging with a patient.

Nevertheless, while the above studies have reported the success of ensemble learning, their methods present varying inefficiencies during model training or inference aggregation [55]. The overarching purpose of the present study, therefore, was to improve the overall performance of a multiclass skin lesion detection process using an ensemble deep learning that utilizes the strength of different heterogeneous models with the capability of parallel processing. The current study further strengthens the possibility of using an automated medical expert decision support system as a valuable device for soliciting a second opinion from dermatologists with the following cardinal contributions.

1. The introduction of two majority voting ensemble deep learning algorithms to accurately classify up to ten different skin lesion classes with improved classification performance.
2. The application of distributed computing paradigm to achieve a faster and more timely multiclass classification of skin lesions into one of three or multiple classes.
3. The comprehensive evaluation of the introduced majority voting ensemble deep learning algorithms against 28 state-of-the-art deep learning and ensemble learning algorithms as a way of demonstrating their prowess.

3. Materials and Methods

The materials for this study include the experimental datasets and computing devices used for classifying skin lesions. The computer has a capacity of 16 GB graphics process-

ing unit (GPU) memory, 2560 compute unified device architecture (CUDA) Cores, and 9 trillion floating point operations per second (TFLOPS). The study methods were based on the application of visual geometry group network (VGGNet) [56], residual network (ResNet) [57], residual network with neXt dimension (ResNeXt) [58], dense convolutional network (DenseNet) [59], dual path network (DPN) [60], efficient channel attention residual network (ECA ResNet) [61], instagram dataset-based residual network with neXt dimension (IG ResNeXt) [62], semi-weakly supervised learning residual network with neXt dimension (SWSL ResNeXt) [63], inception ResNet v2 [64], and rank expansion networks (ReXNet) [65]. These standard network models were applied in the present study to develop two new ensemble deep learning algorithms. The study methods also include the widely used metrics for performance evaluation and the experimentation method was to test the performance of the developed learning algorithms.

3.1. Dataset

This study uses 71,522 images congregated from 6 publicly available datasets of international skin imaging collaborations (ISIC) for experimentation. The training and validation datasets comprise the ISIC 2016 (ISIC_2016_TRN) [66], ISIC 2017 (ISIC_2017_TRN and ISIC_2017_VAL) [4], ISIC 2018 (ISIC_2018_TRN) [67,68], ISIC 2019 (ISIC_2019_TRN) [4,67,69], and ISIC 2020 (ISIC_2020_TRN) [70]. The datasets as described in Table 1 reflected 2, 3, 3, 7, 9, and 9 skin lesion classes respectively. It should be noted that images overlap across different ISIC datasets. Many studies on skin lesion segmentation and classification, for example, transformed lesion image data from original red, green, and blue (RGB) color to other color models such as the international commission on illumination perceptually uniform (CIELab) color model [1,14,39,71] and non-linear hue, saturation, value (HSV) color model [28,52]. However, the original RGB colors of skin lesion images were preserved in this study to avoid additional computation costs and possible tinting that can result from the specification of an inappropriate illuminant.

Table 1. Experimental image datasets.

Datasets	Descriptions	Sizes	Classes
ISIC_2017_TRN [4]	Training set for ISIC 2017 Classification Task	2000	3
ISIC_2017_VAL [4]	The validation set for ISIC 2017 Classification Task	150	3
ISIC_2016_TRN [66]	Training set for ISIC 2016 Classification Task	900	2
ISIC_2019_TRN [4,67,69]	Training set for ISIC 2019 Classification Task	25,331	9
ISIC_2018_TRN [67,68]	Training set for ISIC 2018 Classification Task	10,015	7
ISIC_2020_TRN [70]	Training set for ISIC 2020 Classification Task	33,126	9

Figure 1 shows the inclusion and exclusion protocol for certain skin lesion classes. The lesion classes like lentigo not otherwise specified (NOS), solar lentigo, lichenoid keratosis, cafe-au-lait macule, and atypical melanocytic proliferation in ISIC_2020_TRN were excluded because they present less than 50 lesion images. The number of such images in the sub-dataset was 90, thereby reducing the ISIC_2020_TRN experimental images from 33,126 to 33,036.

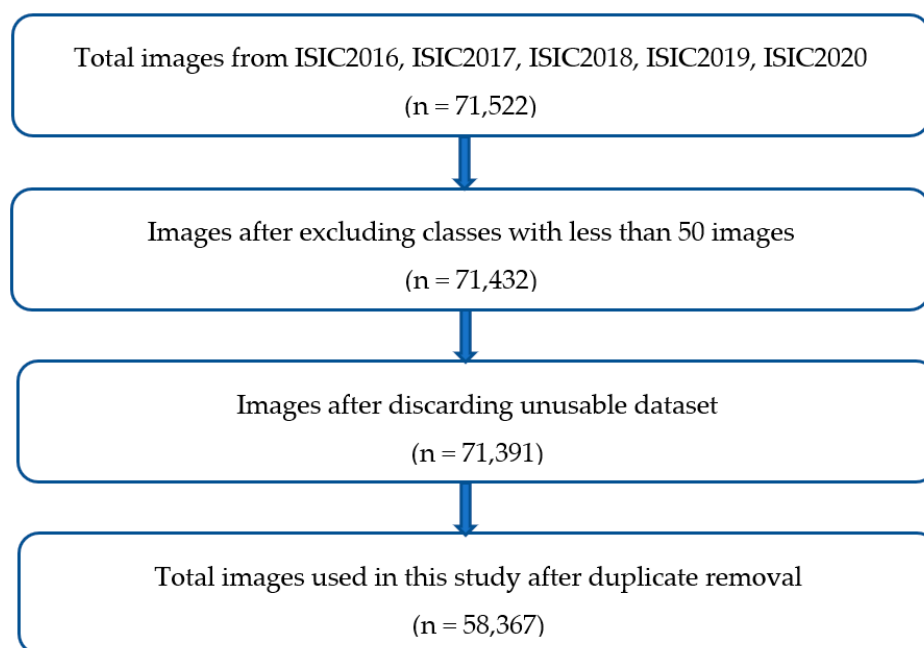


Figure 1. Inclusion and exclusion protocol for the selected skin lesion images.

Figure 2 displays the sample skin lesion images of ten different classes. The benign lesion has a total of seven classes which are actinic keratosis and intraepithelial carcinoma (AKIEC), benign keratosis-like lesions (BKL), dermatofibroma (DF), indeterminate benign (INDB), melanocytic nevi (NV), seborrheic keratosis (SK), and vascular lesions (VASC). The malignant lesion accounted for three classes which are melanoma (MEL), basal cell carcinoma (BCC), and squamous cell carcinoma (SCC). The images used in the dataset contain several noise attributes such as hair shafts (Figure 2a,c,e,f,g,k,l), vignettes (Figure 2a,d,l), ruler markings (Figure 2g,k), air bubble (Figure 2h), blood vessel (Figure 2f,h), and specular reflections (Figure 2e,l). The images equally contain early in-situ skin lesions with difficult border tracing (Figure 2b,c,g,i,j) and invasive skin lesions having different shades of color (Figure 2a,d-f,h,k,l). The use of in-situ lesion skin images can enhance the training of a learner for the early detection of skin diseases. Moreover, the use of invasive skin lesion images can ensure that the proposed ensemble learning methods can provide a valuable second opinion to dermatologists when investigating the diagnosis of a specific skin lesion.

It has been earlier mentioned that ISIC datasets contain duplicate images across different datasets. The duplicates were removed to obtain unique images as highlighted in Table 2. The number of benign images used was 49,310 out of the 58,367 lesion images prepared for training and cross-validation, while malignant classes accounted for a total of 9057 images. The 58,367 images selected were used in this study for training and validation after performing the required data cleansing. The subset of 46,694 skin lesion images that constituted about 80.00% of the selected images was used for training the base learners. The 11,673 skin lesion images which accounted for 20.00% of the selected images were used for cross-validation of each base learner. The validation dataset was equally used to check the strength of the introduced ensemble deep learning algorithms in comparison with the solitary deep learning algorithms considered.

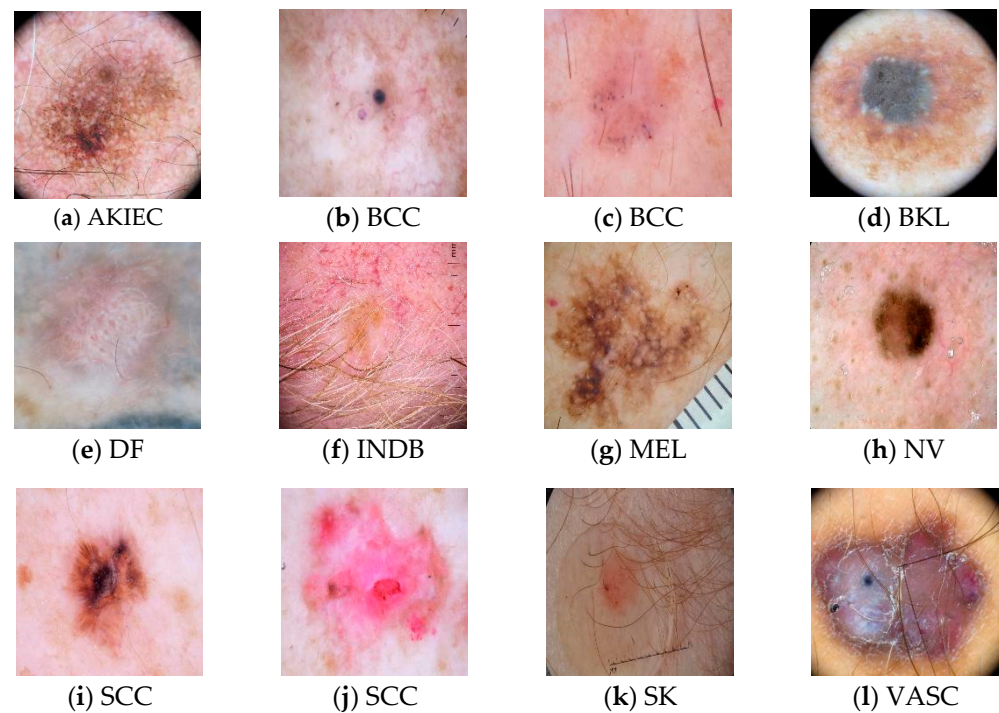


Figure 2. Samples of Skin Lesion Images.

Table 2. Experimental datasets for training and validation.

Database	Benign						Malignant				Total
	AKIEC	BKL	DF	INDB	NV	SK	VASC	BCC	MEL	SCC	
ISIC_2017_TRN [4]					1372	254			374		2000
ISIC_2017_VAL [4]					78	42			30		150
ISIC_2016_TRN [66]		1		29	688				182		900
ISIC_2019_TRN [4,67,69]	867	2624	239		12,875		253	3323	4522	628	25,331
ISIC_2018_TRN [67,68]	327	1099	115		6705		142	514	1113		10,015
ISIC_2020_TRN [70]				27,124	5193	135			584		33,036
Total	1194	3724	354	27,153	26,911	431	395	3837	6805	628	
Duplicates	327	1396	115		8843		142	514	1687		
Discarded				29					12		
Final Selection	867	2328	239	27,124	18,068	431	253	3323	5106	628	

3.2. Methods

The application of transfer learning has assisted significantly to resolve the general problem of insufficient data for the training of machine learning models [72]. The overarching goal was to ensure the realization of a learning algorithm that can accurately discriminate skin lesions in dermoscopic images. The investigated base learners were subjected to a two-stage process to determine the relevance of an image segmentation process. It was also to examine the discriminating prowess of the base learners for constructing novel ensemble deep learning algorithms. There are several approaches such as bagging, boosting, stacking, and voting previously used in the literature to improve the classification performance of ensemble learning. The approach of voting ensemble learning that follows the application of heterogeneous learners with multiclass labels was explored in this study to enhance the generalization of skin lesion classification. The impetus for the exploration of this approach includes its wide usage, simplicity, and efficiency [73,74].

The methods of this study were based on the VGGNet with 11, 13, 16, and 19 layers, ResNet with 18, 34, 50, 101, and 152 layers, ResNeXt having 50, and 101 layers, DenseNet with 121, 161, 169, and 201 layers, DPN with 68, 92, 98, 107, and 131 layers, ECA ResNet of 101 layers, IG ResNeXt having 101 layers, SWSL ResNeXt of 101 layers, ReXNet with $x1.5$ and $x2.0$ repeated layers. The VGGNet uses multiple non-linear rectification layers instead of a single rectification layer for a more discriminative result. The application of VGG-11, VGG-13, VGG-16, and VGG-19 networks was examined in this study [56]. ResNet applies an identity shortcut connection to solve the problem of vanishing gradient resulting from a deeper network [57]. The network uses a pre-activation residual block so that both forward and backward signals are directly propagated from one block to any other block with identity mapping. This mechanism has helped the use of an additive merging of previous layers into the future layers where the network can learn residuals that are the differences between the previous and current layers. The application of ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152 with other recent variations of ResNet was examined in this study. The evaluated variations of the network include residual networks with next dimension (ResNeXt-50, ResNeXt-101) [58], IG-ResNeXt-101 [62], SWSL-ResNeXt-101 [63], and efficient channel attention-based residual networks (ECA-ResNet-101) [61].

DenseNet is a logical extension of ResNet that concatenates the outputs from the previous layers into future layers, instead of using additive merges to sum residuals [58]. In contrast to ResNet, it explicitly differentiates the residual information preserved from the new information added to the network [58]. The intrinsic merits of the model include the reduction in the required hyperparameters by ignoring redundant feature maps and alleviating the vanishing gradient problem. Four types of DenseNet that were examined in this study are DenseNet-121, DenseNet-161, DenseNet-169, and DenseNet-201 [59]. DPN reforms the skip connection and inherits the intrinsic merits of feature reuse from a residual network (ResNet) and feature exploration from a densely connected network (DenseNet). These attributes make DPNs have good parameter efficiency, low computational cost, and low memory footprint. Five variations of DPN that were examined in this study are DPN-68, DPN-92, DPN-98, DPN-107, and DPN-131 [60]. The inception architecture combines the strength of the network in a network (NiN) with a repeated block paradigm. The basic convolutional block in the inception architecture is called the inception block. In this study, a variation of the inception architecture termed inception-ResNet-v2 that utilizes the residual connections was used to achieve an improved classification performance [64]. The ReXNet uses a search method for channel parameterization through the piecewise linear functions of the block index [65]. The two types of ReXNet examined in this study are ReXNet_1.5 and ReXNet_2.0, each corresponding to $x1.5$ and $x2.0$ repeated layers, respectively.

Public lesion image datasets were used for the training and validation of base learners to support a two-staged training process. The individual trainable input image was resized to 224×224 , while each batch of data was fed as input to a learner. In the first training stage, a balanced 2000 data subset of benign and malignant classes consisting of randomly selected 1000 benign and 1000 malignant images were used. In the training, 800 benign and 800 malignant images were used, while 200 benign and 200 malignant images were used for cross-validation. In the second training stage, the entire 58,367 multiclass skin lesion images were used of which about 80.00% of the entire dataset was for training, while the remaining 20.00% of images were used to validate the discriminating capability of each learner.

The purpose of the stage 1 evaluation was to identify the base learners with quality performance on ground truth segmented images or non-segmented images to improve the overall classification performance of skin lesions. This is because there is no consensus reached in the literature on the effect of image segmentation on performance of classification algorithms. Some studies have used segmentation as a prerequisite for performing skin lesion classification [40,75,76], while other have either argued against its necessity or completely ignored its usage for lesion classification [52,77,78]. Consequently, an experiment was performed at Stage 1 with a control group of unsegmented skin lesion images

and an experimental group of ground truth segmented skin lesion images. The 26 investigated deep learning algorithms were trained separately with experimental group images and control group images. The base learners that recorded a minimum dice coefficient (*DSC*) of 0.92 were selected to participate in the Stage 2 evaluation. This was to avoid the potential side effect of imbalanced accuracy and to select deep learners with a consistent discriminating ability across different images.

In stage 2, each base learner was trained to learn the discriminating attributes of AKIEC, BCC, DF, MEL, NV, SK, SCC, and VASC skin lesions. In addition, each learner was trained to classify a skin lesion as either BKL or INDB. A skin lesion is classified as BKL if it has keratosis-like features but cannot be recognized as belonging to a specific known keratosis class. Moreover, a lesion is categorized as INDB if it has benign features but cannot be discriminated from a specific known class. The base learner selection process at stage 2, uses the selected base learners, excluding those that could not be fitted because of the limited GPU capacity of 16 GB of the computer system used for the experimentation. Multiclass accuracy performance was computed for each base learner evaluated at stage 2 to further determine the discriminating strength of learners. The minimum threshold of 90.50% multiclass accuracy was used to select the models to be considered as components of an ensemble deep learning algorithm. In addition, it was used to compute the corresponding weights for a learner.

3.2.1. Performance Evaluation

The learning algorithms were evaluated using well-established performance evaluation metrics of sensitivity, specificity, accuracy, Jaccard index, dice coefficient, multiclass accuracy, precision, and Matthew correlation coefficient. Let *TP*, *TN*, *FN*, and *FP* represent the true positive classification, true negative classification, false positive classification, and false negative classification, respectively. In addition, let *D_{cm}* represents the diagonal for the confusion matrix of *n* total multiple classes. The performance evaluation metrics are computed as functions of true positive, true negative, false positive, and false negative as represented in Table 3.

Table 3. Performance Evaluation Metrics.

Metrics	Symbols	Formulae
Sensitivity	<i>Sn</i>	$Sn = \frac{TP}{TP+FN}$
Specificity	<i>Sp</i>	$Sp = \frac{TN}{TN+FP}$
Accuracy	<i>Acc</i>	$Acc = \frac{TP+TN}{TP+TN+FP+FN}$
Jaccard index	<i>Ji</i>	$Ji = \frac{TP}{TP+FP+FN}$
Dice coefficient	<i>DSC</i>	$DSC = \frac{2TP}{2TP+FP+FN}$
Multiclass accuracy	<i>MAcc</i>	$MAcc = \frac{\sum_i D_{cm}}{n}$
Precision	<i>Pr</i>	$Pr = \frac{TP}{TP+FP}$
Matthew correlation coefficient	<i>MCC</i>	$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

3.2.2. Experimental Setup

The application of an appropriate transformation during training can help deep learning algorithms to generalize better on future inputs. In this study, we applied a maximum of five rotations and deterministic random dihedral transformation as data augmentation methods. These transformations were applied to each data item and data batch fed into the learning algorithms at both stage 1 and stage 2 training because of the imbalance in a multiclass experimental dataset. Regularization and optimization hyperparameters are widely used to improve the performance of deep learning algorithms [79]. In this study, we applied regularization parameters of weight decay and dropout during training to facilitate

the reduction in the out-of-place fitting. Weight decay is typically used as a regularization parameter to minimize the loss function, reduce the possibility of the model over-fitting or under-fitting, and apply an appropriate penalty to large weights. Dropout is generally used as a regularization parameter that involves the probabilistic removal of network nodes to reduce over-fitting during the model training. Such a unit is temporarily removed from the network, including its corresponding connections, by dropping out.

The optimization parameters such as the learning rate and batch size were also applied to strengthen the generalization ability of the introduced voting ensemble deep learning algorithms. The learning rate typically outlines the sizes of the corrective steps that a learning model takes to adjust for errors at each epoch. During training, we used the cycling learning rate to determine the appropriate step size as the loss function gets minimized. This important mechanism has effectively assisted in determining the speed at which each network model can learn discriminating features of skin lesion classes. A high learning rate will shorten the training time, but ultimately with lower accuracy, while a lower learning rate might take longer, but with a potential for greater accuracy. The number of data samples that could be fed into the learning model per epoch differs in size. Batch size refers to the maximum amount of training data that could fit into a single iteration in batch, mini-batch, or stochastic modes. It is generally used to determine the error gradient during model training.

3.3. Proposed Ensemble Deep Learning Algorithm

The learning models of IG-ResNeXt-101, SWSL-ResNeXt-101, ECA-ResNet-101, and DPN-131 were aggregated to build two archetypes of ensemble deep learning algorithms. The ensembles are simple majority voting ensemble (SMVE) and weighted majority voting ensemble (WMVE). They are among the most popular voting ensembles with no definite consensus about which is better. Hence, we have decided to experiment with both of the ensembles to recommend the one that shows more promising results. The Ray distributed framework [80] was used in this study to resolve the slow convergence speed typically experienced by most ensemble learning algorithms because of the time taken to compute the result of an individual base learner [81–84]. The prediction of each handler of a learner was computed in a distributed fashion and the associated results were aggregated when available. This is to reduce the wait time as highlighted in both Algorithm 1 and Algorithm 2. It should be noted that preprocessing of the images is not performed during inference which makes the algorithms robust for practical real-world usage. The Ξ_j represents the predicted output of each handler (\mathfrak{b}_i) of a classifier, \bar{w}_j is the weighted confidence for each Ξ_j and κ_j is the lesion output class to be predicted for a given skin lesion image such that κ satisfies the following.

$$\kappa \in \{AKIEC, BCC, BKL, DF, INDB, NV, MEL, SK, SCC, VASC\} \quad (1)$$

3.3.1. Simple Majority Voting Ensemble Deep Learning Algorithm

The SMVE was achieved by simply combining the predictions of the individual learners. The class with the highest frequency is thereafter used as the final predicted class according to Algorithm 1. The simple majority voting ensemble is computed according to the following expression.

$$P(\kappa_j) = \max(\kappa_j) \quad (2)$$

Algorithm 1: Simple Majority Voting Ensemble Deep Learning

Given an input image i

1. Broadcast i to the respective handlers (b_1, b_2, \dots, b_n) of the learners.
2. Compute the prediction for each handler using distributed processing [80].
3. Compile responses from all handlers of the learners.
4. Aggregate the results of the handlers based on the maximum voting principle.
5. Determine the class prediction $P(\kappa_j)$ using Equation (2).

End

3.3.2. Weighted Majority Voting Ensemble Deep Learning Algorithm

The WMVE uses a confidence preservation mechanism to increase the accuracy of classifying skin lesions. The class prediction of the WMVE corresponding to the category of a skin lesion can be computed using Algorithm 2. The algorithm highlights the essential steps of classifying a given skin lesion image into one category from AKIEC, BCC, BKL, DF, MEL, NV, SK, SCC, VASC, and INDB. Due to the number of base learners used, hard and soft voting schemes were agglutinated to solve for the possibility of an even number of predicted output Ξ_j . The average weighted confidence probability $\mu_{\bar{w}_j}$ of each Ξ_j is given according to the following equation.

$$\mu_{\bar{w}_j} = \frac{1}{n} \sum_j^n \Xi_j \quad (3)$$

Algorithm 2: Weighted Majority Voting Ensemble Deep Learning

Given an input image i

1. Broadcast i to the respective handlers (b_1, b_2, \dots, b_n) of the learners.
2. Compute the prediction for each handler using distributed processing [80].
3. Compile responses from all handlers of the learners.
4. Compute \bar{w}_j for each Ξ_j .
5. Aggregate the results of the handlers with $\bar{w}_j \geq 0.25$.
6. If exactly one class κ_j has the highest predicted output Ξ_j

$$P(\kappa_j) = \kappa_j$$
 Else

$$P(\kappa_j) = \kappa_j \text{ with the maximum average weighted confidence } \mu_{\bar{w}_j}$$

End

4. Experimental Results

The model training was divided into two stages given the number of deep learning algorithms evaluated in this study. The evaluation of base learners was initiated at stage 1 while stage 2 evaluation involved 58,367 images to test 18 base learners that had a minimum threshold of 0.92 *DSC* during the stage 1 evaluation. The application of the ray-distributed framework assisted in reducing the inference process by a factor of 0.25 for a faster classification of skin lesions. The mean time to completion (MTTC) for the WMVE algorithm is 7.31 s but the application of distributed processing reduced the MTTC to 5.30 s. The experimental results of the regularization and optimization, stage 1 evaluation, stage 2 evaluation, and comparison with state-of-the-art algorithms are explicated in this section.

4.1. Regularization and Optimization Outcomes

This study utilizes weight decay, learning rate, batch size, and drop-out as hyperparameters to improve the quality of each base learner. The optimal hyperparameter values were determined as shown in Table 4 after several experiments. Multifarious decay weights within the interval (0.0001, 0.1) were investigated during training to determine the values that best fit each model. It was discovered after several experiments that values within the set {0.0001, 0.01} seem to produce the best results for most of the algorithms. The batch

size within {20, 24, 32} was used for most of the learning algorithms. The learning rate hyperparameters that range from 0.0004 to 0.020 were used in this study.

Table 4. Regularization and optimization hyperparameters.

Learning Algorithm	Weight Decay	Learning Rate	Batch Size	Drop Out ^a
VGG (11, 16) [56]	0.0100	0.0010	12	0.2500 to 0.5000
VGG-13 [56]	0.0100	0.0006	24	0.1250
VGG-19 [56]	0.0100	0.0007	24	0.1250
ResNet (18, 34, 50) [57]	0.0100	0.0010	12	0.2500 to 0.5000
ResNet (101 and 152) [57]	0.0100	0.0200	24	0.2500 to 0.5000
ResNeXt-50 [58]	0.0010	0.0020	24	0.2500 to 0.5000
ResNeXt-101 [58]	0.0100	0.0010	8	0.2500 to 0.5000
DenseNet (121, 169, 201) [59]	0.0100	0.0020	24	0.250 to 0.5000
DenseNet-161 [59]	0.0100	0.0010	20	0.250 to 0.5000
DPN (68, 92) [60]	0.0100	0.0010	12	0.250 to 0.5000
DPN-98 [60]	0.0100	0.0004	32	0.1250
DPN-107 [60]	0.0100	0.0008	32	0.1250
DPN-131 [60]	0.0100	0.0006	32	0.1250
ECA-ResNet-101 [61]	0.0100	0.0025 ^b , 0.0017 ^c	32	0.1250
IG-ResNeXt-101 [62]	0.0100	0.0044 ^b , 0.0007 ^c	24 ^d , 32 ^e	0.1250
SWSL-ResNeXt-101 [63]	0.0100	0.0001 ^b , 0.0006 ^c	32	0.1250
Inception-ResNet-v2 [64]	0.0100	0.0036	32	0.1250
ReXNet_1.5 [65]	0.0100	0.0020	24	0.1250
ReXNet_2.0 [65]	0.0100	0.0010	32	0.1250

^a Represents the lower and upper bound. ^b Represents the most optimal learning rate used at stage 1 of model training. ^c Represents the most optimal learning rate used at state 2 of model training. ^d Represents the batch size used at stage 1 model training. ^e Represents the batch size used at stage 2 of model training.

Figure 3a–c, and d, respectively, display the learning rates used for the DPN-131, ECA-ResNet-101, IG-ResNeXt-101, and SWSL-ResNeXt-101 learners. The dropouts ranging from 0.125 to 0.500 were applied to reduce the undesirable effect of large weights. The use of a constant lower limit of 0.125 and an upper limit of 0.500 appeared to have produced good results for most of the learning algorithms examined in this study.

4.2. Stage 1 Evaluation

Table 5 shows the result of each learning algorithm in classifying malignancy skin lesions over 2000 balanced datasets. The evaluation of the 26 learning algorithms yielded a total of 52 different experimental analyses for both the segmented and non-segmented skin lesion images. A total of 18 learning algorithms selected for further evaluation met the minimum requirement of a 0.92 dice coefficient. The chosen learning algorithms include VGG (13 and 19), ResNet (101 and 152), ResNeXt (50 and 101), ECA-ResNet-101, IG-ResNeXt-101, SWSL-ResNeXt-101, DenseNet (121, 161, 169, and 201), DPN (98, 107, and 131), Inception-ResNet-v2, and ReXNet_1.5. Most of the learning algorithms, interestingly, performed better without prior image segmentation when compared to the image segmentation episode. However, the DPN-107 algorithm achieved a dice coefficient of 0.93 on the segmented skin lesion images and a dice coefficient of 0.91 without prior image segmentation. This result seems to agree with the observation made in [52] that prior lesion segmentation did not significantly improve the classification performance for most of the base learners.

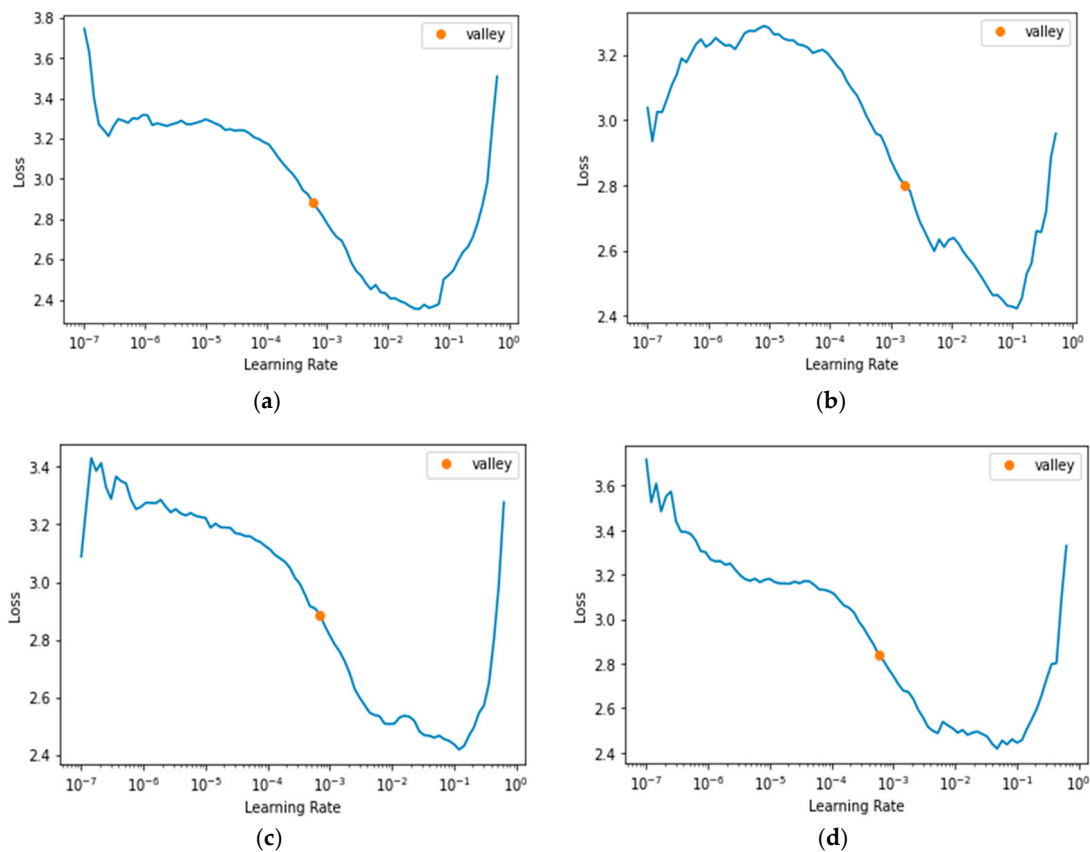


Figure 3. Learning Rate. (a) DPN-131 [60]; (b) ECA-ResNet-101 [61]; (c) IG-ResNet101 [62]; (d) SWSL-ResRet101 [63].

Table 5. Binary classification of malignant skin lesions.

Learning Algorithm	Segmented	Sn%	Sp%	Acc%	Ji	DSC
VGG-11 [56]	✓	70.63	99.38	90.63	0.70	0.82
VGG-11 [56]	x	75.00	98.75	90.31	0.74	0.85
VGG-13 [56]	✓	63.75	98.75	92.81	0.63	0.77
VGG-13 [56]	x	88.75	98.13	95.31	0.87	0.93
VGG-16 [56]	✓	71.25	98.13	91.25	0.70	0.82
VGG-16 [56]	x	66.25	99.38	93.44	0.66	0.79
VGG-19 [56]	✓	78.75	98.13	94.06	0.77	0.87
VGG-19 [56]	x	89.38	100.00	98.44	0.89	0.94
ResNet-18 [57]	✓	44.38	99.38	93.12	0.44	0.61
ResNet-18 [57]	x	75.63	100.00	89.06	0.76	0.86
ResNet-34 [57]	✓	78.75	98.75	97.50	0.78	0.88
ResNet-34 [57]	x	50.00	100.00	90.94	0.50	0.67
ResNet-50 [57]	✓	78.75	98.13	90.63	0.77	0.87
ResNet-50 [57]	x	65.63	100.00	91.56	0.66	0.79
ResNet-101 [57]	✓	50.00	99.38	94.69	0.50	0.66
ResNet-101 [57]	x	90.00	97.50	93.75	0.88	0.94
ResNet-152 [57]	✓	68.75	99.38	91.56	0.68	0.81
ResNet-152 [57]	x	89.38	99.38	94.38	0.89	0.94
ResNeXt-50 [58]	✓	35.63	100.00	85.63	0.36	0.53
ResNeXt-50 [58]	x	89.38	99.38	94.38	0.89	0.94
ResNeXt-101 [58]	✓	80.00	99.38	95.00	0.80	0.89
ResNeXt-101 [58]	x	85.00	99.38	95.00	0.84	0.92

Table 5. Cont.

Learning Algorithm	Segmented	Sn%	Sp%	Acc%	Ji	DSC
DenseNet-121 [59]	✓	80.00	99.38	89.69	0.80	0.89
DenseNet-121 [59]	x	88.75	100.00	94.38	0.89	0.94
DenseNet-161 [59]	✓	85.00	99.38	92.19	0.84	0.92
DenseNet-161 [59]	x	91.88	100.00	95.94	0.92	0.96
DenseNet-169 [59]	✓	70.63	99.38	85.00	0.70	0.82
DenseNet-169 [59]	x	88.75	100.00	94.38	0.89	0.94
DenseNet-201 [59]	✓	79.38	100.00	89.69	0.079	0.89
DenseNet-201 [59]	x	90.63	99.38	95.00	0.90	0.95
DPN-68 [60]	✓	69.38	100.00	84.69	0.69	0.82
DPN-68 [60]	x	77.50	99.38	88.44	0.77	0.87
DPN-92 [60]	✓	66.88	99.38	83.13	0.66	0.80
DPN-92 [60]	x	70.72	99.38	84.16	0.70	0.83
DPN-98 [60]	✓	76.25	98.75	87.50	0.75	0.86
DPN-98 [60]	x	91.25	100.00	95.63	0.91	0.95
DPN-107 [60]	✓	86.88	99.38	93.13	0.86	0.93
DPN-107 [60]	x	84.38	99.38	91.88	0.84	0.91
DPN-131 [60]	✓	72.50	99.38	85.94	0.72	0.84
DPN-131 [60]	x	88.13	100.00	94.06	0.88	0.94
ECA-ResNet-101 [61]	✓	48.75	100.00	74.38	0.49	0.66
ECA-ResNet-101 [61]	x	86.88	100.00	93.44	0.87	0.93
IG-ResNeXt-101 [61]	✓	76.25	98.13	87.19	0.75	0.86
IG-ResNeXt-101 [61]	x	89.38	98.75	94.06	0.88	0.94
SWSL-ResNeXt-101 [63]	✓	80.63	98.75	89.69	0.80	0.89
SWSL-ResNeXt-101 [63]	x	85.00	100.00	92.50	0.85	0.92
Inception-ResNet-v2 [64]	✓	75.63	95.63	85.63	0.72	0.84
Inception-ResNet-v2 [64]	x	85.63	99.38	92.50	0.85	0.92
ReXNet_1.5 [65]	✓	80.63	98.75	89.69	0.80	0.89
ReXNet_1.5 [65]	x	88.75	100.00	94.38	0.89	0.94
ReXNet_2.0 [65]	✓	51.25	98.75	75.00	0.51	0.67
ReXNet_2.0 [65]	x	53.13	99.38	76.25	0.53	0.69

4.3. Stage 2 Evaluation

The 18 selected learning algorithms from stage 1 were further evaluated across the entire datasets for training and validation in stage 2. Each algorithm was trained to classify skin lesion images into the known classes of AKIEC, BCC, BKL, DF, MEL, NV, SK, SCC, and VASC. The ResNeXt-101 and DPN-1s07 base learners could not be fitted to the training datasets because of the limited GPU capacity of 16 GB compared to other high-end computer vision hardware and were excluded from the stage 2 exercise. The results of a stage 2 evaluation on the entire skin lesion training and validation datasets are highlighted in Table 6. The IG-ResNeXt-101 and SWSL-ResNeXt-101 were the topmost performing models, while the three least performing models were DenseNet-121, DenseNet-161, and VGG-13 when compared to the rest of the learning models. Most of the base learners as specified in Table 6 performed above the average threshold of 0.50 in predicting AKIEC lesions relative to their DSC except for DenseNet-121 and DenseNet-161. The most precise base learner for AKIEC was SWSL-ResNeXt-101 with 82.55% precision which is closely followed by IG-ResNeXt-101 with 82.12% precision. Both IG-ResNeXt-101 and SWSL-ResNeXt-101 were the top two performing base learners whereas DenseNet-121 and DenseNet-161 accounted for the least two performing learners for malignant lesions of BCC, SCC, and MEL.

Table 6. Classification of skin lesions into one of the ten multiple classes.

Learning Algorithm	Class	Sn%	Sp%	Acc%	MAcc%	Pr%	Ji	DSC	MCC	Weight
VGG-13 [56]	AKIEC	49.43	99.43	98.67	86.53	56.86	0.36	0.53	0.52	0.87
	BCC	70.89	99.26	97.70		84.67	0.63	0.77	0.76	
	BKL	69.40	97.68	96.50		56.62	0.45	0.62	0.61	
	DF	59.57	99.92	99.76		75.68	0.50	0.67	0.67	
	MEL	60.50	97.56	94.39		69.94	0.48	0.65	0.62	
	NV	88.65	93.81	92.22		86.44	0.78	0.88	0.82	
	SCC	55.74	99.52	99.06		54.84	0.38	0.55	0.55	
	SK	47.06	99.55	99.17		43.48	0.29	0.45	0.45	
	VASC	88.00	99.91	99.86		81.48	0.73	0.85	0.85	
INDB	95.96	95.52	95.73	94.97	0.91	0.95	0.91			
VGG-19 [56]	AKIEC	51.14	99.69	98.95	88.24	71.43	0.42	0.60	0.60	0.88
	BCC	75.43	99.38	98.07		87.64	0.68	0.81	0.80	
	BKL	70.02	98.36	97.18		65.08	0.51	0.67	0.66	
	DF	63.83	99.91	99.77		75.00	0.53	0.69	0.69	
	MEL	63.30	97.65	94.71		71.61	0.51	0.67	0.64	
	NV	90.46	94.54	93.28		88.06	0.81	0.89	0.84	
	SCC	65.57	99.33	98.98		50.96	0.40	0.57	0.57	
	SK	48.24	99.59	99.22		46.59	0.31	0.47	0.47	
	VASC	84.00	99.96	99.89		89.36	0.76	0.87	0.87	
INDB	97.02	95.89	96.42	95.42	0.93	0.96	0.93			
ResNet-101 [57]	AKIEC	60.80	99.69	99.10	90.42	74.83	0.50	0.67	0.67	0.90
	BCC	85.29	99.42	98.65		89.49	0.78	0.87	0.87	
	BKL	79.06	98.47	97.66		69.24	0.59	0.74	0.73	
	DF	76.60	99.97	99.87		90.00	0.71	0.83	0.83	
	MEL	65.30	98.44	95.61		79.73	0.56	0.72	0.70	
	NV	92.24	95.32	94.37		89.76	0.83	0.91	0.87	
	SCC	63.93	99.74	99.37		72.22	0.51	0.68	0.68	
	SK	41.18	99.84	99.41		64.81	0.34	0.50	0.51	
	VASC	92.00	99.98	99.95		95.83	0.88	0.94	0.94	
INDB	97.84	96.00	96.86	95.57	0.94	0.97	0.94			
ResNet-152 [57]	AKIEC	62.50	99.75	99.19	90.29	79.14	0.54	0.70	0.70	0.90
	BCC	86.23	99.33	98.61		88.16	0.77	0.87	0.86	
	BKL	76.18	99.01	98.06		76.97	0.62	0.77	0.76	
	DF	85.11	99.91	99.85		80.00	0.70	0.82	0.82	
	MEL	67.10	98.56	95.86		81.33	0.58	0.74	0.72	
	NV	91.40	95.85	94.48		90.75	0.84	0.91	0.87	
	SCC	71.31	99.76	99.46		75.65	0.58	0.73	0.73	
	SK	45.88	99.82	99.43		65.00	0.37	0.54	0.54	
	VASC	94.00	99.98	99.96		95.92	0.90	0.95	0.95	
INDB	97.53	94.03	95.67	93.52	0.91	0.95	0.91			
ResNeXt-50 [58]	AKIEC	64.77	99.75	99.22	90.35	79.72	0.56	0.71	0.71	0.90
	BCC	85.76	99.43	98.68		89.69	0.78	0.88	0.87	
	BKL	76.18	98.78	97.83		73.03	0.59	0.75	0.73	
	DF	85.11	99.90	99.84		76.92	0.68	0.81	0.81	
	MEL	65.30	98.73	95.86		82.76	0.57	0.73	0.71	
	NV	92.43	95.19	94.34		89.52	0.83	0.91	0.87	
	SCC	74.59	99.74	99.48		75.21	0.60	0.75	0.75	
	SK	40.00	99.82	99.38		61.82	0.32	0.49	0.49	
	VASC	92.00	99.97	99.93		92.00	0.85	0.92	0.92	
INDB	97.72	95.40	96.49	94.94	0.93	0.96	0.93			

Table 6. Cont.

Learning Algorithm	Class	Sn%	Sp%	Acc%	MAcc%	Pr%	Ji	DSC	MCC	Weight
DenseNet-121 [59]	AKIEC	28.41	99.62	98.54	86.72	53.19	0.23	0.37	0.38	0.87
	BCC	83.41	97.61	96.83		66.88	0.59	0.74	0.73	
	BKL	59.55	98.06	96.45		57.20	0.41	0.58	0.57	
	DF	10.64	99.97	99.61		62.50	0.10	0.18	0.26	
	MEL	58.80	98.07	94.71		74.06	0.49	0.66	0.63	
	NV	89.84	94.75	93.24		88.39	0.80	0.89	0.84	
	SCC	12.30	99.85	98.94		46.88	0.11	0.19	0.24	
	SK	0.00	99.92	99.19		0.00	0.00	0.00	0.00	
	VASC	52.00	99.73	99.53		45.61	0.32	0.49	0.48	
INDB	98.43	94.60	96.39	94.15	0.93	0.96	0.94			
DenseNet-161 [59]	AKIEC	13.07	99.70	98.40	80.21	40.35	0.11	0.20	0.22	0.80
	BCC	80.13	95.27	94.44		49.52	0.44	0.61	0.60	
	BKL	30.60	98.52	95.69		47.45	0.23	0.37	0.36	
	DF	0.00	99.05	98.66		0.00	0.00	0.00	0.00	
	MEL	35.50	98.22	92.85		65.14	0.30	0.46	0.45	
	NV	87.65	90.36	89.52		80.17	0.72	0.84	0.76	
	SCC	0.00	100.00	98.95		0.00	0.00	0.00	0.00	
	SK	0.00	100.00	99.27		0.00	0.00	0.00	0.00	
	VASC	12.00	99.43	99.06		8.33	0.05	0.10	0.10	
INDB	94.43	92.84	93.58	92.09	0.87	0.93	0.87			
DenseNet-169 [59]	AKIEC	51.14	99.65	98.92	89.30	69.23	0.42	0.59	0.59	0.89
	BCC	79.34	99.44	98.34		89.10	0.72	0.84	0.83	
	BKL	75.56	98.19	97.25		64.56	0.53	0.70	0.68	
	DF	80.85	99.89	99.81		74.51	0.63	0.78	0.78	
	MEL	66.70	97.80	95.13		73.95	0.54	0.70	0.68	
	NV	90.46	95.46	93.92		89.86	0.82	0.90	0.86	
	SCC	52.46	99.75	99.25		68.82	0.42	0.60	0.60	
	SK	45.88	99.79	99.40		61.90	0.36	0.53	0.53	
	VASC	94.00	99.93	99.91		85.45	0.81	0.90	0.90	
INDB	97.81	95.66	96.67	95.22	0.93	0.96	0.93			
DenseNet-201 [59]	AKIEC	66.48	99.70	99.20	90.28	77.48	0.56	0.72	0.71	0.90
	BCC	86.23	99.45	98.72		90.03	0.79	0.88	0.87	
	BKL	77.82	98.69	97.82		72.19	0.60	0.75	0.74	
	DF	85.11	99.97	99.91		93.02	0.80	0.89	0.89	
	MEL	69.80	98.14	95.71		77.81	0.58	0.74	0.71	
	NV	91.40	95.78	94.43		90.60	0.83	0.91	0.87	
	SCC	72.13	99.85	99.56		83.81	0.63	0.78	0.78	
	SK	43.53	99.79	99.38		60.66	0.34	0.51	0.51	
	VASC	88.00	99.98	99.93		95.65	0.85	0.92	0.92	
INDB	97.97	96.05	96.95	95.63	0.94	0.97	0.94			
DPN-98 [60]	AKIEC	59.09	99.70	99.09	90.36	75.36	0.50	0.66	0.66	0.90
	BCC	85.76	99.47	98.71		90.28	0.79	0.88	0.87	
	BKL	77.21	98.76	97.86		73.01	0.60	0.75	0.74	
	DF	74.47	99.93	99.83		81.40	0.64	0.78	0.78	
	MEL	69.60	98.31	95.85		79.45	0.59	0.74	0.72	
	NV	91.51	95.67	94.39		90.38	0.83	0.91	0.87	
	SCC	69.67	99.68	99.37		69.67	0.53	0.70	0.69	
	SK	32.94	99.82	99.33		57.14	0.26	0.42	0.43	
	VASC	94.00	99.94	99.91		87.04	0.82	0.90	0.90	
INDB	97.57	95.32	96.38	94.85	0.93	0.96	0.93			

Table 6. Cont.

Learning Algorithm	Class	Sn%	Sp%	Acc%	MAcc%	Pr%	Ji	DSC	MCC	Weight
DPN-131 [60]	AKIEC	59.66	99.74	99.13	90.60	77.78	0.51	0.68	0.68	0.91
	BCC	87.17	99.42	98.75		89.69	0.79	0.88	0.88	
	BKL	78.03	98.91	98.04		75.70	0.62	0.77	0.76	
	DF	85.11	99.95	99.89		86.96	0.75	0.86	0.86	
	MEL	69.70	98.33	95.88		79.66	0.59	0.74	0.72	
	NV	92.46	95.32	94.44		89.79	0.84	0.91	0.87	
	SCC	75.41	99.72	99.47		74.19	0.60	0.75	0.75	
	SK	38.82	99.78	99.34		56.90	0.30	0.46	0.47	
	VASC	90.00	99.95	99.91		88.24	0.80	0.89	0.89	
INDB	96.91	95.87	96.36	95.40	0.93	0.96	0.93			
ECA-ResNet-101 [61]	AKIEC	64.20	99.68	99.14	90.50	75.33	0.53	0.69	0.69	0.90
	BCC	84.19	99.52	98.68		91.03	0.78	0.87	0.87	
	BKL	78.85	98.74	97.91		73.14	0.61	0.76	0.75	
	DF	80.85	99.93	99.85		82.61	0.69	0.82	0.82	
	MEL	72.20	97.88	95.68		76.16	0.59	0.74	0.72	
	NV	90.82	96.26	94.59		91.53	0.84	0.91	0.87	
	SCC	67.21	99.77	99.43		75.23	0.55	0.71	0.71	
	SK	40.00	99.78	99.35		57.63	0.31	0.47	0.48	
	VASC	96.00	99.96	99.94		90.57	0.87	0.93	0.93	
INDB	97.57	95.39	96.41	94.92	0.93	0.96	0.93			
IG-ResNeXt-101 [62]	AKIEC	70.45	99.77	99.32	91.80	82.12	0.61	0.76	0.76	0.92
	BCC	89.83	99.46	98.93		90.54	0.82	0.90	0.90	
	BKL	81.72	99.05	98.33		78.97	0.67	0.80	0.79	
	DF	78.72	99.94	99.85		84.05	0.69	0.81	0.81	
	MEL	77.00	98.13	96.32		79.38	0.64	0.78	0.76	
	NV	91.68	96.68	95.14		92.48	0.85	0.92	0.89	
	SCC	77.05	99.87	99.63		86.24	0.69	0.81	0.81	
	SK	44.71	99.72	99.31		53.52	0.32	0.49	0.49	
	VASC	88.00	100.00	99.95		100.00	0.88	0.94	0.94	
INDB	97.61	96.11	96.81	95.68	0.93	0.97	0.94			
SWSL-ResNeXt-101 [63]	AKIEC	69.89	99.77	99.32	91.46	82.55	0.61	0.76	0.76	0.91
	BCC	89.67	99.56	99.01		92.12	0.83	0.91	0.90	
	BKL	81.11	99.02	98.27		78.22	0.66	0.80	0.79	
	DF	95.74	99.85	99.83		71.43	0.69	0.82	0.83	
	MEL	75.30	98.01	96.07		78.03	0.62	0.77	0.75	
	NV	90.96	96.82	95.01		92.71	0.85	0.92	0.88	
	SCC	79.51	99.80	99.59		80.83	0.67	0.80	0.80	
	SK	42.35	99.70	99.28		50.70	0.30	0.46	0.46	
	VASC	88.00	99.97	99.91		91.67	0.81	0.90	0.90	
INDB	97.59	95.76	96.62	95.31	0.93	0.96	0.93			
Inception-ResNet-v2 [64]	AKIEC	43.18	99.83	98.97	87.74	79.17	0.39	0.56	0.58	0.88
	BCC	71.83	99.49	97.98		89.13	0.66	0.80	0.79	
	BKL	72.28	98.46	97.37		67.18	0.53	0.70	0.68	
	DF	80.85	99.91	99.84		79.17	0.67	0.80	0.80	
	MEL	62.10	98.35	95.25		77.92	0.53	0.69	0.67	
	NV	91.04	92.83	92.28		84.96	0.78	0.88	0.82	
	SCC	59.02	99.79	99.37		75.00	0.49	0.66	0.66	
	SK	30.59	99.78	99.27		50.00	0.23	0.38	0.39	
	VASC	88.00	99.90	99.85		78.57	0.71	0.83	0.83	
INDB	96.51	94.26	95.31	93.69	0.91	0.95	0.91			

Table 6. Cont.

Learning Algorithm	Class	Sn%	Sp%	Acc%	MAcc%	Pr%	Ji	DSC	MCC	Weight
ReXNet_1.5 [65]	AKIEC	64.77	99.51	98.99	89.20	67.06	0.49	0.66	0.65	0.89
	BCC	79.66	99.40	98.32		88.52	0.72	0.84	0.83	
	BKL	71.46	98.76	97.62		71.46	0.56	0.71	0.70	
	DF	68.09	99.97	99.84		88.89	0.63	0.77	0.78	
	MEL	63.60	98.12	95.16		75.99	0.53	0.69	0.67	
	NV	91.37	94.23	93.35		87.57	0.81	0.89	0.85	
	SCC	68.85	99.81	99.49		79.25	0.58	0.74	0.74	
	SK	34.12	99.89	99.41		60.05	0.30	0.46	0.48	
	VASC	88.00	99.95	99.90		88.00	0.79	0.88	0.88	
INDB	97.42	95.35	96.32	94.88	0.93	0.96	0.93			
SMVE (Ours)	AKIEC	70.45	99.83	99.38	92.33	86.11	0.63	0.78	0.78	
	BCC	90.61	99.56	99.07		92.34	0.84	0.91	0.91	
	BKL	80.70	99.20	98.43		81.54	0.68	0.81	0.80	
	DF	89.36	99.96	99.91		89.36	0.81	0.89	0.89	
	MEL	74.10	98.63	96.53		83.54	0.65	0.79	0.77	
	NV	93.16	96.47	95.45		92.16	0.86	0.93	0.89	
	SCC	81.15	99.89	99.69		88.39	0.73	0.85	0.85	
	SK	42.35	99.79	99.37		60.00	0.33	0.50	0.50	
	VASC	94.00	100.00	99.97		100.00	0.94	0.97	0.97	
INDB	98.10	95.73	96.84	95.30	0.94	0.97	0.94			
WMVE (Ours)	AKIEC	73.30	99.83	99.43	92.84	87.16	0.66	0.80	0.80	
	BCC	91.24	99.57	99.12		92.54	0.85	0.92	0.91	
	BKL	82.75	99.29	98.60		83.61	0.71	0.83	0.82	
	DF	93.62	99.96	99.93		89.80	0.85	0.92	0.92	
	MEL	76.90	98.63	96.77		84.04	0.67	0.80	0.79	
	NV	93.18	96.88	95.74		93.00	0.87	0.93	0.90	
	SCC	82.79	99.88	99.70		87.83	0.74	0.85	0.85	
	SK	41.18	99.78	99.35		57.38	0.32	0.48	0.48	
	VASC	92.00	99.98	99.95		95.83	0.88	0.94	0.94	
INDB	98.26	96.03	97.08	95.63	0.94	0.97	0.94			

Most of the base learners recorded low results for SK. ResNet-152 recorded the best DSC of 0.54 followed by DenseNet-169 with 0.53 DSC. This may be a result of its similar characteristics to melanoma that could easily confuse the model training [85]. NV lesion was one of the most correctly classified skin lesions by the base learners with Densenet-161 having the lowest DSC value of 0.84. INDB was correctly classified by averaging 0.95 for most of the base learners besides the ReXNet_1.5 which recorded a value of 0.84 for the SC. ECA-ResNet-101 recorded the most sensitivity of 96.00% for VASC. ResNet-152 had the best discriminating capability for VASC having a 0.95 DSC. However, both DenseNet-121 and DenseNet-161 performed poorly when compared to other base learners in the prediction of VASC lesions. The IG-ResNeXt-101 and SWSL-ResNeXt-101 were again the top two performing base learners in the classification of BKL, while DenseNet-161 recorded the worst performance of 0.37 DSC. The validation of the base learners against DF lesions indicated that DenseNet-201 provided the most discriminating power of 0.89 for DSC and DenseNet-121 had the least discriminating power of 0.18 for DSC. The metrics that provided the most guided evaluation for the base learners were DSC, Ji, and multiclass accuracy (MAcc). It can therefore be argued that the solitary application of simple balance accuracy, sensitivity, specificity, and precision without the compliment of DSC, Ji, and MAcc might not be a good representation of the discriminating capability of a given learning algorithm.

Both SMVE and WMVE algorithms were observed to outperform each base learner for most of the skin lesion classes because they recorded the best multiclass accuracies of 92.33% and 92.84%, respectively. The two algorithms shared the best specificity result of 99.83% with the Inception-ResNet-v2 base learner for the prediction of AKIEC. However, the WMVE algorithm provided the best result for sensitivity (73.30%), balance accuracy

(99.43%), precision (87.16%), Jaccard index (0.66), and Dice coefficient (0.80), and Matthew correlation coefficient (0.80) in the discrimination of AKIEC. It recorded the most promising result of sensitivity (91.24%), specificity (99.57%), balance accuracy (99.12%), precision (92.54%), and Jaccard index (0.85) with dice-coefficient (0.92) for malignant BCC skin lesions. The two algorithms had the best Matthew correlation coefficient result of 0.91 for the same BCC lesion class classification. Similarly, for the classification of the BKL lesion class, the WMVE algorithm chronicled the best performance based on the sensitivity (82.75%), specificity (99.29%), balance accuracy (98.60%), precision (83.61%), Jaccard index (0.71), dice-coefficient (0.83) and Matthew correlation coefficient (0.82). The WMVE algorithm had the best results for balance accuracy (99.93%), dice coefficient (0.92), and Matthew correlation coefficient (0.92) for the classification of DF skin lesions. It has equally demonstrated its prowess for the classification of malignant MEL by recording the best balance accuracy score of 96.77%, precision score of 84.04%, Jaccard index of 0.67, dice coefficient of 0.80, and Matthew correlation coefficient of 0.79.

The two ensemble algorithms had the best dice coefficient result of 0.93 for the classification of the benign NV lesion class. However, the WMVE algorithm recorded the best results of sensitivity (93.18%), specificity (96.88%), balance accuracy (95.74%), precision (93.00%), and Jaccard index (0.87) with Matthew correlation coefficient (0.90) for the classification of the same NV lesion. The classification of malignant SCC has equally shown the strength of both algorithms, with the SMVE algorithm recording the best precision value of 88.39%. They had the best dice coefficient (0.85) and Matthew correlation coefficient (0.85) for the same SCC skin lesion class prediction. But the WMVE algorithm outperformed the rest of the compared algorithms for SCC classification in terms of sensitivity (82.79%) and balance accuracy (99.70%). The SMVE algorithm had the same specificity (100.00%) and precision (100.00%) as the IG-ResNeXt-101 base learner for the classification of VASC lesions. However, the best results for balance accuracy (99.97%), Jaccard index (0.94), dice coefficient (0.97), and Matthew correlation coefficient (0.97) were credited to the SMVE algorithm for the discrimination of VASC lesions. The WMVE algorithm has equally recorded the best balance accuracy of 97.08% for classifying skin lesions categorized as INDB, whose specific benign properties could not be ascertained. The two algorithms have demonstrated equal discriminating capability for INDB lesions with the DenseNet-201 base learner considering the reported Jaccard index (0.94), dice coefficient (0.97), and Matthew correlation coefficient (0.94). Figure 4 shows the confusion matrix of the base learners used to build the ensemble deep learning algorithms for the classification of ten classes of skin lesions. The matrices have revealed that the strength of an ensemble deep learning algorithm is greatly influenced by how well each of its base learners was trained to generalize.

4.4. Comparison with State-of-the-Art Algorithms

A comparative analysis of the SMVE and WMVE algorithms against the investigated state-of-the-art learning algorithms was performed in this study (Table 7). The ensemble algorithms can be seen to outperform most of the comparative algorithms for classifying skin lesions into multiple classes. Most of the past studies that leverage the use of ensemble learning were observed to focus on a few classes of skin lesions such as melanoma, melanocytic nevi, and seborrheic keratosis. The study [75] is the result of the application of the kernel extreme learning machine (KELM) with ResNet101 and DenseNet201. However, the WMVE algorithm outstripped the compared ensemble algorithms with a mean specificity of 99.00%, mean balance accuracy of 98.54%, and multiclass accuracy of 82.84%. The SMVE and WMVE algorithms with the algorithm reported in [86] recorded the best DSC of 0.84. Most of the studies reported in the literature for classifying skin lesions generally used evaluation metrics such as sensitivity, specificity, balanced accuracy, and precision to examine the strength of their methods. While these are good, we have discovered that metrics such as the Jaccard index, multiclass accuracy, and DSC are more reliable and consistent in measuring the effectiveness of a learning algorithm for classifying skin lesion images.

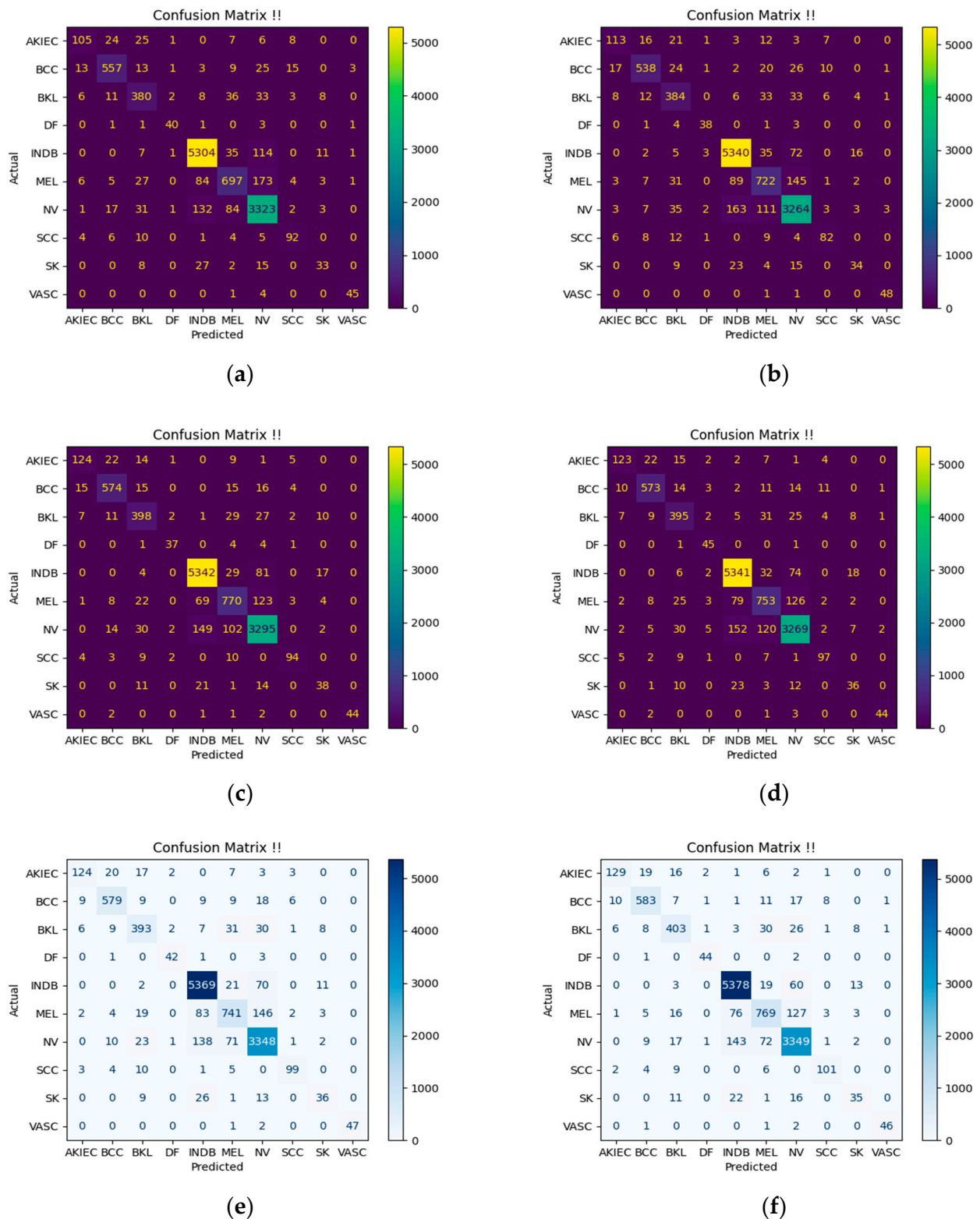


Figure 4. Confusion Matrices. (a) DPN-131 [60]; (b) ECA-ResNet-101 [61]; (c) IG-ResNeXt-101 [62]; (d) SWSL-ResNeXt-101 [63]; (e) Simple Majority Voting Ensemble; (f) Weighted Majority Voting Ensemble.

Table 7. Algorithm comparison with the state-of-the-art ensemble learning algorithms.

Reference	Algorithm	No of Classes	Data	Avg Sn (%)	Avg Sp (%)	Avg Acc (%)	MAcc (%)	Avg Pr (%)	Avg Ji	Avg DSC	MCC
[40]	Inception-v3, ResNet-50, Inception-ResNet-v2, and DenseNet-201	7	ISIC_2018			88.44					
[52]	EfficientNetB0, EfficientNetB1 and SeReNeXt-50	7	ISIC_2018			96.30	86.20	91.30			
[75]	Kernel Extreme Learning Machine, ResNet101 and DenseNet201	7	HAM_10000	90.20			90.67				
[76]	EW-FCM and EfficientNet-B0	8	HAM_10000	87.23	97.87	87.23					
[77]	GoogleNet and SVM	8	ISIC_2019	79.80	97.00	94.92		80.36			
[78]	KNN and SVM	8	ISIC_2019	66.45	97.85	97.35		91.61			
[86]	InceptionResNetV2 and ResNeXt101	7	ISIC_2019	85.00		92.83		83.00		0.84	
[87]	EfficientNets, SENet, and ResNeXt WSL	9	ISIC_2019	74.20							
[88]	DenseNet201 and MobileNetV2		ISIC_2018, ISIC_2019			94.50					
[89]	VGG16, VGG19, ResNet50, ResNet101, ResNet152, Xception, and MobileNet	7	HAM_10000				83.69				
[90]	AlexNet, ResNet18, ResNet50, ResNet101, ResNet152, SENet154, SqueezeNet1_0, VGG13BN and VGG16BN	7	ISIC_2018, ISIC_2019	48.30	97.70	92.40				0.49	
[91]	MobileNetV2 and GoogLeNet	7	HAM_10000	65.60	95.40	83.50		76.60		0.69	0.66
SMVE (Ours)	IG-ResNeXt-101, SWSL-ResNeXt-101, ECA-ResNet-101, and DPN-131	10	ISIC_2016, ISIC_2017, ISIC_2018, ISIC_2019, ISIC_2020	81.40	98.91	98.46	92.33	86.87	0.74	0.84	0.83
WMVE (Ours)	IG-ResNeXt-101, SWSL-ResNeXt-101, ECA-ResNet-101, and DPN-131	10	ISIC_2016, ISIC_2017, ISIC_2018, ISIC_2019, ISIC_2020	82.52	99.00	98.54	92.84	86.68	0.75	0.84	0.84

5. Conclusions

The weighted majority voting ensemble deep (WMVE) learning algorithm is suggested in this paper for classifying each of the ten different lesion classes. Eight of the classes have divulged known skin lesion diseases without compromising accuracy and speed. The known skin lesions are actinic keratosis and intraepithelial carcinoma, basal cell carcinoma, dermatofibroma, melanoma, melanocytic nevi, seborrheic keratosis, squamous cell carcinoma, and vascular lesions. The WMVE algorithm can classify a skin lesion as either benign keratosis-like or indeterminate benign. We have found the application of the WMVE algorithm to yield improved skin lesion classification results when compared to individual or majority voting algorithms investigated in this study. Moreover, executing the algorithm in a distributed way indicates a better opportunity for the timely classification of skin lesions into one of the ten multiple classes examined. It has been determined experimentally in this study that most of the existing learning algorithms investigated did well without the prior image segmentation of skin lesion images. The application of transfer learning and well-tuned hyperparameters such as dropout, weight decay, batch size, and learning rate has equally contributed to preventing model over-fitting plus increasing efficiency and generalization of the proposed learning algorithm.

The majority voting ensemble deep-learning algorithms of this study can improve the early diagnosis of skin lesions before they become invasive. In addition, for skin lesions that are already at the invasive stage, the proposed ensemble learning algorithms can act as a good second opinion for dermatologists to achieve effective discrimination of malignant tumors. It would be prudent as a future improvement to examine whether the discriminating capability of base learners can be improved with the use of hyperspectral images. These are images that typically have more spectral and spatial information on the ground features when compared to regular multispectral images. The spectral information reflects the unique physical properties of the ground features for a given image.

Author Contributions: Conceptualization by D.A.O. and O.O.O.; methodology by D.A.O.; data curation by D.A.O.; writing original draft by D.A.O.; writing—review and editing by O.O.O.; supervision by O.O.O.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available ISIC datasets used in this study can be found at <https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery?filter=%5B%5D>. Accessed on 20 February 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Olugbara, O.O.; Taiwo, T.B.; Heukelman, D. Segmentation of Melanoma Skin Lesion Using Perceptual Color Difference Saliency with Morphological Analysis. *Math. Probl. Eng.* **2018**, *2018*, 1524286. [[CrossRef](#)]
2. Okuboyejo, D.A.; Olugbara, O.O. Segmentation of Melanocytic Lesion Images Using Gamma Correction with Clustering of Keypoint Descriptors. *Diagnostics* **2021**, *11*, 1366. [[CrossRef](#)] [[PubMed](#)]
3. Ashraf, H.; Waris, A.; Ghafoor, M.F.; Gilani, S.O.; Niazi, I.K. Melanoma Segmentation Using Deep Learning with Test-Time Augmentations and Conditional Random Fields. *Sci. Rep.* **2022**, *12*, 3948. [[CrossRef](#)] [[PubMed](#)]
4. Codella, N.C.F.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.; Kittler, H.; 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 Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 168–172. [[CrossRef](#)]
5. Li, Z.; Fang, Y.; Chen, H.; Zhang, T.; Yin, X.; Man, J.; Yang, X.; Lu, M. Spatiotemporal Trends of the Global Burden of Melanoma in 204 Countries and Territories from 1990 to 2019: Results from the 2019 Global Burden of Disease Study. *Neoplasia* **2022**, *24*, 12–21. [[CrossRef](#)] [[PubMed](#)]
6. Saginala, K.; Barsouk, A.; Aluru, J.S.; Rawla, P.; Barsouk, A. Epidemiology of Melanoma. *Med. Sci.* **2021**, *9*, 63. [[CrossRef](#)]
7. Memon, A.; Bannister, P.; Rogers, I.; Sundin, J.; Al-Ayadhy, B.; James, P.W.; McNally, R.J.Q. Changing Epidemiology and Age-Specific Incidence of Cutaneous Malignant Melanoma in England: An Analysis of the National Cancer Registration Data by Age, Gender and Anatomical Site, 1981–2018. *Lancet Reg. Health Eur.* **2021**, *2*, 100024. [[CrossRef](#)]

8. Okuboyejo, D.A.; Olugbara, O.O. A Review of Prevalent Methods for Automatic Skin Lesion Diagnosis. *Open Dermatol. J.* **2018**, *12*, 14–53. [[CrossRef](#)]
9. van der Kooij, M.K.; Wetzels, M.J.A.L.; Aarts, M.J.B.; van den Berkortel, F.W.P.J.; Blank, C.U.; Boers-Sonderen, M.J.; Dierselhuis, M.P.; de Groot, J.W.B.; Hospers, G.A.P.; Piersma, D.; et al. Age Does Matter in Adolescents and Young Adults versus Older Adults with Advanced Melanoma; A National Cohort Study Comparing Tumor Characteristics, Treatment Pattern, Toxicity and Response. *Cancers* **2020**, *12*, 2072. [[CrossRef](#)]
10. Davis, D.S.; Robinson, C.; Callender, V.D. Skin Cancer in Women of Color: Epidemiology, Pathogenesis and Clinical Manifestations. *Int. J. Women's Dermatol.* **2021**, *7*, 127–134. [[CrossRef](#)]
11. Paulson, K.G.; Gupta, D.; Kim, T.S.; Veatch, J.R.; Byrd, D.R.; Bhatia, S.; Wojcik, K.; Chapuis, A.G.; Thompson, J.A.; Madeleine, M.M.; et al. Age-Specific Incidence of Melanoma in the United States. *JAMA Dermatol.* **2020**, *156*, 57–64. [[CrossRef](#)]
12. Popescu, D.; El-Khatib, M.; El-Khatib, H.; Ichim, L. New Trends in Melanoma Detection Using Neural Networks: A Systematic Review. *Sensors* **2022**, *22*, 496. [[CrossRef](#)] [[PubMed](#)]
13. Okuboyejo, D.A.; Olugbara, O.O.; Odunaiké, S.A. Automating Skin Disease Diagnosis Using Image Classification. *Proc. World Congr. Eng. Comput. Sci.* **2013**, *2*, 850–854.
14. Joseph, S.; Olugbara, O.O. Preprocessing Effects on Performance of Skin Lesion Saliency Segmentation. *Diagnostics* **2022**, *12*, 344. [[CrossRef](#)] [[PubMed](#)]
15. Papageorgiou, V.; Apalla, Z.; Sotiriou, E.; Papageorgiou, C.; Lazaridou, E.; Vakirlis, S.; Ioannides, D.; Lallas, A. The Limitations of Dermoscopy: False-Positive and False-Negative Tumours. *J. Eur. Acad. Dermatol. Venereol.* **2018**, *32*, 879–888. [[CrossRef](#)] [[PubMed](#)]
16. Son, H.M.; Jeon, W.; Kim, J.; Heo, C.Y.; Yoon, H.J.; Park, J.-U.; Chung, T.-M. AI-Based Localization and Classification of Skin Disease with Erythema. *Sci. Rep.* **2021**, *11*, 5350. [[CrossRef](#)] [[PubMed](#)]
17. Cui, X.; Wei, R.; Gong, L.; Qi, R.; Zhao, Z.; Chen, H.; Song, K.; Abdulrahman, A.A.A.; Wang, Y.; Chen, J.Z.S.; et al. Assessing the Effectiveness of Artificial Intelligence Methods for Melanoma: A Retrospective Review. *J. Am. Acad. Dermatol.* **2019**, *81*, 1176–1180. [[CrossRef](#)]
18. Ding, S.; Wu, Z.; Zheng, Y.; Liu, Z.; Yang, X.; Yang, X.; Yuan, G.; Xie, J. Deep Attention Branch Networks for Skin Lesion Classification. *Comput. Methods Programs Biomed.* **2021**, *212*, 106447. [[CrossRef](#)]
19. Lucieri, A.; Bajwa, M.N.; Braun, S.A.; Malik, M.I.; Dengel, A.; Ahmed, S. ExAID: A Multimodal Explanation Framework for Computer-Aided Diagnosis of Skin Lesions. *Comput. Methods Programs Biomed.* **2022**, *215*, 106620. [[CrossRef](#)]
20. Khan, S.; Rahmani, H.; Shah, S.A.A.; Bennamoun, M. Applications of CNNs in Computer Vision. In *A Guide to Convolutional Neural Networks for Computer Vision*; Khan, S., Rahmani, H., Shah, S.A.A., Bennamoun, M., Eds.; Synthesis Lectures on Computer Vision; Springer International Publishing: Cham, Switzerland, 2018; pp. 117–158.
21. Bhatt, D.; Patel, C.; Talsania, H.; Patel, J.; Vaghela, R.; Pandya, S.; Modi, K.; Ghayvat, H. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. *Electronics* **2021**, *10*, 2470. [[CrossRef](#)]
22. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *542*, 115. [[CrossRef](#)]
23. Haenssle, H.A.; Fink, C.; Toberer, F.; Winkler, J.; Stolz, W.; Deinlein, T.; Hofmann-Wellenhof, R.; Lallas, A.; Emmert, S.; Buhl, T.; et al. Man against Machine Reloaded: Performance of a Market-Approved Convolutional Neural Network in Classifying a Broad Spectrum of Skin Lesions in Comparison with 96 Dermatologists Working under Less Artificial Conditions. *Ann. Oncol.* **2020**, *31*, 137–143. [[CrossRef](#)] [[PubMed](#)]
24. Brinker, T.J.; Hekler, A.; Enk, A.H.; Klode, J.; Hauschild, A.; Berking, C.; Schilling, B.; Haferkamp, S.; Schadendorf, D.; Holland-Letz, T.; et al. Deep Learning Outperformed 136 of 157 Dermatologists in a Head-to-Head Dermoscopic Melanoma Image Classification Task. *Eur. J. Cancer* **2019**, *113*, 47–54. [[CrossRef](#)] [[PubMed](#)]
25. Tschandl, P.; Codella, N.; Akay, B.N.; Argenziano, G.; Braun, R.P.; Cabo, H.; Gutman, D.; Halpern, A.; Helba, B.; Hofmann-Wellenhof, R.; et al. Comparison of the Accuracy of Human Readers versus Machine-Learning Algorithms for Pigmented Skin Lesion Classification: An Open, Web-Based, International, Diagnostic Study. *Lancet Oncol.* **2019**, *20*, 938–947. [[CrossRef](#)] [[PubMed](#)]
26. Maron, R.C.; Weichenthal, M.; Utikal, J.S.; Hekler, A.; Berking, C.; Hauschild, A.; Enk, A.H.; Haferkamp, S.; Klode, J.; Schadendorf, D.; et al. Systematic Outperformance of 112 Dermatologists in Multiclass Skin Cancer Image Classification by Convolutional Neural Networks. *Eur. J. Cancer* **2019**, *119*, 57–65. [[CrossRef](#)]
27. Fujisawa, Y.; Otomo, Y.; Ogata, Y.; Nakamura, Y.; Fujita, R.; Ishitsuka, Y.; Watanabe, R.; Okiyama, N.; Ohara, K.; Fujimoto, M. Deep Learning Surpasses Dermatologists. *Br. J. Dermatol.* **2019**, *180*, e39. [[CrossRef](#)]
28. Codella, N.C.F.; Nguyen, Q.-B.; Pankanti, S.; Gutman, D.A.; Helba, B.; Halpern, A.C.; Smith, J.R. Deep Learning Ensembles for Melanoma Recognition in Dermoscopy Images. *IBM J. Res. Dev.* **2017**, *61*, 5:1–5:15. [[CrossRef](#)]
29. Goyal, M.; Knackstedt, T.; Yan, S.; Hassanpour, S. Artificial Intelligence-Based Image Classification Methods for Diagnosis of Skin Cancer: Challenges and Opportunities. *Comput. Biol. Med.* **2020**, *127*, 104065. [[CrossRef](#)]
30. Gupta, T.; Saini, S.; Saini, A.; Aggarwal, S.; Mittal, A. Deep Learning Framework for Recognition of Various Skin Lesions Due to Diabetes. In Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 19–22 September 2018; pp. 92–98.
31. Namozov, A.; Cho, Y.I. Convolutional Neural Network Algorithm with Parameterized Activation Function for Melanoma Classification. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 17–19 October 2018; pp. 417–419.

32. Hasan, M.K.; Roy, S.; Mondal, C.; Alam, M.A.; E Elahi, M.T.; Dutta, A.; Uddin Raju, S.M.T.; Jawad, M.T.; Ahmad, M. DermoDOCTOR: A Framework for Concurrent Skin Lesion Detection and Recognition Using a Deep Convolutional Neural Network with End-to-End Dual Encoders. *Biomed. Signal Process. Control* **2021**, *68*, 102661. [[CrossRef](#)]
33. Hasan, M.K.; Elahi, M.T.E.; Alam, M.A.; Jawad, M.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](#)]
34. Iqbal, I.; Younus, M.; Walayat, K.; Kakar, M.U.; Ma, J. Automated Multi-Class Classification of Skin Lesions through Deep Convolutional Neural Network with Dermoscopic Images. *Comput. Med. Imaging Graph.* **2021**, *88*, 101843. [[CrossRef](#)]
35. Tawhid, A.; Teotia, T.; Elmiligi, H. Chapter 13—Machine Learning for Optimizing Healthcare Resources. In *Machine Learning, Big Data, and IoT for Medical Informatics*; Kumar, P., Kumar, Y., Tawhid, M.A., Eds.; Intelligent Data-Centric Systems; Academic Press: Cambridge, MA, USA, 2021; pp. 215–239.
36. Sugiyama, M. Chapter 27—Support Vector Classification. In *Introduction to Statistical Machine Learning*; Sugiyama, M., Ed.; Morgan Kaufmann: Boston, MA, USA, 2016; pp. 303–320.
37. Song, M.; Shang, X.; Chang, C.-I. 3-D Receiver Operating Characteristic Analysis for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8093–8115. [[CrossRef](#)]
38. Pawara, P.; Okafor, E.; Groefsema, M.; He, S.; Schomaker, L.R.B.; Wiering, M.A. One-vs-One Classification for Deep Neural Networks. *Pattern Recognit.* **2020**, *108*, 107528. [[CrossRef](#)]
39. Tang, P.; Liang, Q.; Yan, X.; Xiang, S.; Zhang, D. GP-CNN-DTEL: Global-Part CNN Model with Data-Transformed Ensemble Learning for Skin Lesion Classification. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2870–2882. [[CrossRef](#)] [[PubMed](#)]
40. Al-masni, M.A.; Kim, D.-H.; Kim, T.-S. Multiple Skin Lesions Diagnostics via Integrated Deep Convolutional Networks for Segmentation and Classification. *Comput. Methods Programs Biomed.* **2020**, *190*, 105351. [[CrossRef](#)]
41. Mahbod, A.; Schaefer, G.; Ellinger, I.; Ecker, R.; Pitiot, A.; Wang, C. Fusing Fine-Tuned Deep Features for Skin Lesion Classification. *Comput. Med. Imaging Graph.* **2018**, *71*, 19–29. [[CrossRef](#)] [[PubMed](#)]
42. Afza, F.; Sharif, M.; Mittal, M.; Khan, M.A.; Jude Hemanth, D. A Hierarchical Three-Step Superpixels and Deep Learning Framework for Skin Lesion Classification. *Methods* **2022**, *202*, 88–102. [[CrossRef](#)]
43. Adetiba, E.; Olugbara, O.O. Lung Cancer Prediction Using Neural Network Ensemble with Histogram of Oriented Gradient Genomic Features. *Sci. World J.* **2015**, *2015*, e786013. [[CrossRef](#)]
44. Mondal, C.; Hasan, M.K.; Ahmad, M.; Awal, M.A.; Jawad, M.T.; Dutta, A.; Islam, M.R.; Moni, M.A. Ensemble of Convolutional Neural Networks to Diagnose Acute Lymphoblastic Leukemia from Microscopic Images. *Inform. Med. Unlocked* **2021**, *27*, 100794. [[CrossRef](#)]
45. Pérez, E.; Ventura, S. An Ensemble-Based Convolutional Neural Network Model Powered by a Genetic Algorithm for Melanoma Diagnosis. *Neural Comput. Appl.* **2021**, *34*, 10429–10448. [[CrossRef](#)]
46. Mitrea, D.; Badea, R.; Mitrea, P.; Brad, S.; Nedeveschi, S. Hepatocellular Carcinoma Automatic Diagnosis within CEUS and B-Mode Ultrasound Images Using Advanced Machine Learning Methods. *Sensors* **2021**, *21*, 2202. [[CrossRef](#)]
47. Sagi, O.; Rokach, L. Ensemble Learning: A Survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1249. [[CrossRef](#)]
48. Shahin, A.H.; Kamal, A.; Elattar, M.A. Deep Ensemble Learning for Skin Lesion Classification from Dermoscopic Images. In Proceedings of the 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, 20–22 December 2018; pp. 150–153.
49. Hilmy, M.A.; Sasongko, P.S. Ensembles of Convolutional Neural Networks for Skin Lesion Dermoscopy Images Classification. In Proceedings of the 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 29–30 October 2019; pp. 1–6.
50. Lima, J.P.D.O.; Filho, L.C.S.D.A.; Da Silva, F.S.; Figueiredo, C.M.S. Seródio Figueiredo Pigmented Dermatological Lesions Classification Using Convolutional Neural Networks Ensemble Mediated by Multilayer Perceptron Network. *IEEE Lat. Am. Trans.* **2019**, *17*, 1902–1908. [[CrossRef](#)]
51. Wei, L.; Ding, K.; Hu, H. Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network. *IEEE Access* **2020**, *8*, 99633–99647. [[CrossRef](#)]
52. Mahbod, A.; Schaefer, G.; Wang, C.; Dorffner, G.; Ecker, R.; Ellinger, I. Transfer Learning Using a Multi-Scale and Multi-Network Ensemble for Skin Lesion Classification. *Comput. Methods Programs Biomed.* **2020**, *193*, 105475. [[CrossRef](#)]
53. Harangi, B. Skin Lesion Classification with Ensembles of Deep Convolutional Neural Networks. *J. Biomed. Inform.* **2018**, *86*, 25–32. [[CrossRef](#)] [[PubMed](#)]
54. Pham, H.N.; Koay, C.Y.; Chakraborty, T.; Gupta, S.; Tan, B.L.; Wu, H.; Vardhan, A.; Nguyen, Q.H.; Palaparthi, N.R.; Nguyen, B.P.; et al. Lesion Segmentation and Automated Melanoma Detection Using Deep Convolutional Neural Networks and XGBoost. In Proceedings of the 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 20–21 July 2019; pp. 142–147.
55. Dong, X.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A Survey on Ensemble Learning. *Front. Comput. Sci.* **2020**, *14*, 241–258. [[CrossRef](#)]
56. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. *Identity Mappings in Deep Residual Networks*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 630–645.

58. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5987–5995.
59. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
60. Chen, Y.; Li, J.; Xiao, H.; Jin, X.; Yan, S.; Feng, J. Dual Path Networks. In Proceedings of the 2019 International SoC Design Conference (ISOCC), Jeju, Republic of Korea, 6–9 October 2019; p. 9.
61. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
62. Mahajan, D.; Girshick, R.; Ramanathan, V.; He, K.; Paluri, M.; Li, Y.; Bharambe, A.; van der Maaten, L. *Exploring the Limits of Weakly Supervised Pretraining*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 185–201.
63. Yalniz, I.Z.; Jegou, H.; Chen, K.; Ai, F.; Paluri, M.; Mahajan, D. Billion-Scale Semi-Supervised Learning for Image Classification. *arXiv* **2019**, arXiv:1905.00546.
64. 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 February 2017; pp. 4278–4284.
65. Han, D.; Yun, S.; Heo, B.; Yoo, Y. Rethinking Channel Dimensions for Efficient Model Design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 732–741.
66. Gutman, D.; Codella, N.C.F.; Celebi, E.; Helba, B.; Marchetti, M.; Mishra, N.; Halpern, A. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, Hosted by the International Skin Imaging Collaboration (ISIC). *arXiv* **2016**, arXiv:1605.01397v1. [[CrossRef](#)]
67. 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](#)]
68. Codella, N.; Rotemberg, V.; Tschandl, P.; Celebi, M.E.; Dusza, S.; Gutman, D.; Helba, B.; Kalloo, A.; Marchetti, M.; Kittler, H.; et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). *arXiv* **2018**, arXiv:1902.03368. [[CrossRef](#)]
69. Combalia, M.; Codella, N.C.F.; Rotemberg, V.; Helba, B.; Vilaplana, V.; Reiter, O.; Barreiro, A.; Halpern, A.C.; Puig, S.; Malvehy, J. BCN20000: Dermoscopic Lesions in the Wild. *arXiv* **2019**, arXiv:1908.02288. [[CrossRef](#)]
70. Rotemberg, V.; Kurtansky, N.; Betz-Stablein, B.; Caffery, L.; Chousakos, E.; Codella, N.; Combalia, M.; Dusza, S.; Guitera, P.; Gutman, D.; et al. A Patient-Centric Dataset of Images and Metadata for Identifying Melanomas Using Clinical Context. *Sci. Data* **2021**, *8*, 34. [[CrossRef](#)] [[PubMed](#)]
71. Okubojejo, D.A.; Olugbara, O.O.; Odunaike, S.A. *CLAHE Inspired Segmentation of Dermoscopic Images Using Mixture of Methods*; Kim, H.K., Ao, S.-I., Amouzegar, M.A., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 355–365.
72. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. *A Survey on Deep Transfer Learning*; Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 270–279.
73. Li, D.; Liu, Z.; Armaghani, D.J.; Xiao, P.; Zhou, J. Novel Ensemble Intelligence Methodologies for Rockburst Assessment in Complex and Variable Environments. *Sci. Rep.* **2022**, *12*, 1844. [[CrossRef](#)]
74. Kausar, N.; Hameed, A.; Sattar, M.; Ashraf, R.; Imran, A.S.; Abidin, M.Z.U.; Ali, A. Multiclass Skin Cancer Classification Using Ensemble of Fine-Tuned Deep Learning Models. *Appl. Sci.* **2021**, *11*, 10593. [[CrossRef](#)]
75. Khan, M.A.; Sharif, M.; Akram, T.; Damaševičius, R.; Maskeliūnas, R. Skin Lesion Segmentation and Multiclass Classification Using Deep Learning Features and Improved Moth Flame Optimization. *Diagnostics* **2021**, *11*, 811. [[CrossRef](#)] [[PubMed](#)]
76. Hoang, L.; Lee, S.-H.; Lee, E.-J.; Kwon, K.-R. Multiclass Skin Lesion Classification Using a Novel Lightweight Deep Learning Framework for Smart Healthcare. *Appl. Sci.* **2022**, *12*, 2677. [[CrossRef](#)]
77. Kassem, M.A.; Hosny, K.M.; Fouad, M.M. Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning. *IEEE Access* **2020**, *8*, 114822–114832. [[CrossRef](#)]
78. Molina-Molina, E.O.; Solorza-Calderón, S.; Álvarez-Borrego, J. Classification of Dermoscopy Skin Lesion Color-Images Using Fractal-Deep Learning Features. *Appl. Sci.* **2020**, *10*, 5954. [[CrossRef](#)]
79. Kotsilieris, T.; Anagnostopoulos, I.; Livieris, I.E. Special Issue: Regularization Techniques for Machine Learning and Their Applications. *Electronics* **2022**, *11*, 521. [[CrossRef](#)]
80. Moritz, P.; Nishihara, R.; Wang, S.; Tumanov, A.; Liaw, R.; Liang, E.; Elibol, M.; Yang, Z.; Paul, W.; Jordan, M.I.; et al. Ray: A Distributed Framework for Emerging AI Applications. *arXiv* **2018**, arXiv:1712.05889.
81. Wang, Z.; Pan, X.; Wei, G.; Fei, J.; Lu, X. A Faster Convergence and Concise Interpretability TSK Fuzzy Classifier Deep-Wide-Based Integrated Learning. *Appl. Soft Comput.* **2019**, *85*, 105825. [[CrossRef](#)]
82. Padmakala, S.; Subasini, C.A.; Karuppiah, S.P.; Sheeba, A. ESVM-SWRF: Ensemble SVM-Based Sample Weighted Random Forests for Liver Disease Classification. *Int. J. Numer. Methods Biomed. Eng.* **2021**, *37*, e3525. [[CrossRef](#)] [[PubMed](#)]
83. Shafieian, S.; Zulkernine, M. Multi-Layer Stacking Ensemble Learners for Low Footprint Network Intrusion Detection. *Complex Intell. Syst.* **2022**. [[CrossRef](#)]

84. Sueki, K.; Nishizawa, S.; Yamaura, T.; Tomita, H. Precision and Convergence Speed of the Ensemble Kalman Filter-Based Parameter Estimation: Setting Parameter Uncertainty for Reliable and Efficient Estimation. *Prog. Earth Planet. Sci.* **2022**, *9*, 47. [[CrossRef](#)]
85. Janowska, A.; Oranges, T.; Iannone, M.; Romanelli, M.; Dini, V. Seborrheic Keratosis-like Melanoma: A Diagnostic Challenge. *Melanoma Res.* **2021**, *31*, 407–412. [[CrossRef](#)] [[PubMed](#)]
86. Chaturvedi, S.S.; Tembhumne, J.V.; Diwan, T. A Multi-Class Skin Cancer Classification Using Deep Convolutional Neural Networks. *Multimed. Tools Appl.* **2020**, *79*, 28477–28498. [[CrossRef](#)]
87. Gessert, N.; Nielsen, M.; Shaikh, M.; Werner, R.; Schlaefer, A. Skin Lesion Classification Using Ensembles of Multi-Resolution EfficientNets with Meta Data. *MethodsX* **2020**, *7*, 100864. [[CrossRef](#)] [[PubMed](#)]
88. Attique Khan, M.; Sharif, M.; Akram, T.; Kadry, S.; Hsu, C.-H. A Two-Stream Deep Neural Network-Based Intelligent System for Complex Skin Cancer Types Classification. *Int. J. Intell. Syst.* **2021**, 1–29. [[CrossRef](#)]
89. Swetha, R.N.; Shrivastava, V.K.; Parvathi, K. Multiclass Skin Lesion Classification Using Image Augmentation Technique and Transfer Learning Models. *Int. J. Intell. Unmanned Syst.* **2021**. *ahead-of-print*. [[CrossRef](#)]
90. Gong, A.; Yao, X.; Lin, W. Classification for Dermoscopy Images Using Convolutional Neural Networks Based on the Ensemble of Individual Advantage and Group Decision. *IEEE Access* **2020**, *8*, 155337–155351. [[CrossRef](#)]
91. Thurnhofer-Hemsi, K.; López-Rubio, E.; Domínguez, E.; Elizondo, D.A. Skin Lesion Classification by Ensembles of Deep Convolutional Networks and Regularly Spaced Shifting. *IEEE Access* **2021**, *9*, 112193–112205. [[CrossRef](#)]