

Skin-NeT: Skin Cancer Diagnosis Using VGG and ResNet-Based Ensemble Learning Approaches



Abdullah Alshehri 

Department of Information Technology, Al Baha University, Al Aqiq 7738-65799, Saudi Arabia

Corresponding Author Email: aashehri@bu.edu.sa

Copyright: ©2024 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410405>

ABSTRACT

Received: 14 June 2023
Revised: 10 March 2024
Accepted: 23 April 2024
Available online: 31 August 2024

Keywords:

skin cancer, ensemble learning, machine learning, feature extraction, VGG-16, ResNet-50, VGG-19, Xception

To address the critical issue of late skin cancer diagnosis and its severe implications, this study leverages the latest in Computer Aided Diagnosis (CAD) and machine learning technologies. Despite the alignment of these technologies with professional medical diagnostics, challenges such as data imbalance, management of extensive datasets, and the need for high-quality images for superior feature extraction continue to pose significant hurdles. To overcome these challenges, this work introduced a novel approach utilizing ensemble learning, which significantly enhances the accuracy of early skin cancer detection. This research elaborates on the creation of two distinct ensemble models: one that combines the capabilities of VGG-16 and ResNet-50, and another that utilizes VGG-19 and Xception. These combinations were specifically chosen for their complementary strengths in deep learning and feature extraction, which are crucial for improving diagnostic accuracy. The models were trained on a comprehensive dataset of over 3000 skin images, achieving a groundbreaking training accuracy of 100% and a testing accuracy that reaches up to 85%. The rationale behind selecting these models for ensemble approach is their proven effectiveness in deep learning tasks. VGG models are renowned for their deep convolutional networks that excel in capturing intricate details, while ResNet models effectively address the vanishing gradient problem, enabling deeper network training without compromising performance. This strategic amalgamation enhances the ability to tackle the complexities of skin cancer detection. In comparative analysis, I eschew specific study references for a broader perspective on performance enhancement. The accuracy of these proposed models shows a substantial increase over existing methods, with testing accuracies advancing from the typical range of 75% to 84% observed in prior works, to as high as 85% in my models. This improvement not only demonstrates the superiority of ensemble learning approach over single-model methods but also establishes a new benchmark in the accuracy and reliability of skin cancer diagnostic tools.

1. INTRODUCTION

Skin cancer is one of the most common types of cancer that has rapidly spread in recent years. According to the report in [1], skin cancer can be diagnosed in one out of three cases of cancer, and it is estimated that more than 0.1 million new cases will be diagnosed globally by the end of 2022, with nearly 7,000 deaths per year from this disease. Nevertheless, skin cancer is typically classified into two categories, namely: i) melanoma (malignant) and ii) non-melanoma (benign). Melanoma is a life-threatening type of skin cancer and is responsible for more than 70% of deaths of patients globally diagnosed with melanoma [2]. Skin cancer is caused due to the abnormal growth of the cells and spreads to other body parts. Most cases of skin cancer are caused by prolonged exposure to the sun's ultraviolet (UV) radiation [3]. This increases the risk of all major types of skin cancers, such as basal cell skin cancer (BCC), melanoma, and squamous cell skin cancer (SCC) [4, 5]. However, a decrease in exposure to UV radiation is one of the effective methods for preventing BCC and non-

melanoma types of skin cancer, which is curable by surgical techniques, whereas melanoma is generally treated by radiation therapy or chemotherapy [6]. Melanoma has a high survival rate amongst all other cancer types if diagnosed early [3]. A typical procedure for diagnosing melanoma is employing the ABCDE rule, which can classify melanoma and non-melanoma from benign skin lesions [7]. Thus, it becomes essential to use computational techniques to improve the diagnosis process. A computational diagnosis of melanoma will help facilitate the early diagnosis process and save the patients' lives. At present, a biopsy of skin samples is performed by doctors to diagnose melanoma. However, this process is invasive, painful, and time-consuming [8]. Thus, the computational diagnosis will help in the noninvasive and effective diagnosis of melanoma.

Recently, artificial intelligence techniques (AI), such as machine learning (ML) and deep learning (DL), have been explored in various studies for the analysis of skin images for early diagnosis of melanoma. These techniques not only help dermatologists in the early diagnosis of skin cancer but also

help in avoiding unnecessary expenditures in imaging tests and biopsies [9]. Computational systems for melanoma diagnosis help in providing high accuracy diagnose of melanoma. For initial analysis of skin cancer, dermatologists often analyze macroscopic images of the skin. However, it becomes challenging to analyze these images due to the low quality of images and the presence of skin lines and shadows in the images [8]. Computational techniques can help resolve these challenges through image processing, data cleaning, and feature extraction [7-9]. Skin cancer is one of the leading causes of death in recent times. Various cases of death have been reported worldwide due to the inaccuracy of the medical diagnosis system. A person suffering from skin cancer can be saved if early diagnosis occurs with highly efficient diagnosis techniques [10]. Thus, various machine learning techniques have been suggested in recent studies to diagnose skin cancer to resolve this challenge. However, challenges need to be resolved to harness the full potential of machine learning in diagnosing skin cancer. Machine learning techniques often require high-quality data. In this context, high-quality image data helps achieve high accuracy in detecting, segmentation, and feature extraction of skin cancer. Retrieving high-quality image datasets is challenging as datasets are often heterogeneous with variable quality of images [11-13]. Furthermore, the segmentation of clinical images is a time-consuming and tedious task. These challenges make us search for more advanced techniques, such as ensemble learning, which can make the diagnosis and segmentation process faster. The challenges of machine learning techniques are summarized in Table 1.

Table 1. Summary of challenges of various machine learning models

Type of Model	Functions of Model	Challenges of Model
Linear	These functions as additive model which is capable of computing risk by using a weighted combination of features.	Additive nature of the model fails to detect the interactions between the variables.
Neural Network	It is a highly non-linear model which predicts based on the network of weighted transformation of features of input.	This technique is challenging to interpret, which increases the overall complexity of the model and to generate more accurate outcomes, parameter optimization is also required.
Decision Tree	It is also a non-linear algorithm that helps in the representation of the interaction between the variables.	This technique fails to establish a continuous relationship between the variables and outcomes.

Ensemble Learning (EL) helps resolve the challenges described in Table 1 by improving the predictive performance of a single model by combining outcomes from multiple trained models [14]. There are various deep learning and ensemble learning techniques suggested by other studies which are not suitable to be implemented at large scale, for example, the authors [15] have suggested a technique based on machine learning and dynamic training-testing augmentation to predict skin cancer. Although the technique achieved a

higher accuracy, but this technique can't be implemented in real world as this technique takes 300 hours of GPU time of Tesla V100 GPU to complete predictions. Because of high time consumption, this technique is not suitable to be used in cases of emergency. Furthermore, the authors [16] have developed a machine learning technique based on 6-class classification using support vector machine (SVM) algorithm to classify acne, eczema, psoriasis, and skin cancer from 1800 skin images with an accuracy of 83%. However, this technique is not reliable as a single model is used to classify various skin diseases that too with a low accuracy. The authors [17] have used ensemble deep learning technique to detect skin cancer from skin images. This technique achieved good results in detecting skin cancer but failed to evaluate the role of optimizers such as Adam optimizer. Therefore, to resolve the challenges faced by these studies and to suggest a more focused approach to detect skin cancer, this paper implemented EL through VGG-16 and ResNeT-50 and VGG-19 and Xception models. This paper proposes a novel approach for diagnosing melanoma from skin images through Ensemble Learning (EL). In concise, EL is a technique in which various machine learning models are integrated to generate one optimum model. Thus, EL helps generate more accurate results than any of the single ML and DL models, which is further explained in the upcoming sections of this paper. The main contributions of this paper are summarized as follows:

- This paper suggests an EL model with the help of VGG-16 + ResNeT-50 and VGG-19 + Xception models for melanoma diagnosis from skin images, along with a comparative analysis between the two models.
- A comparative analysis between various ML and DL models is also conducted with the EL model to show the efficiency of ensemble learning.

The remainder of this paper is structured as follows. The literature review is introduced in Section 2. This is followed by Section 3, where the methodology is described. The results and analysis are shown in Section 4. Finally, the paper is concluded with remarks for future directions in Section 5.

2. RELATED WORK

Recently, ML techniques have played a significant role in the diagnosis of cancer; however, it remains not entirely accepted by dermatologists due to concerns that ML techniques will replace dermatologists [18]. However, this is not the case, as ML techniques are applied to make the diagnosis faster. Technological advancements make ML more widely used for melanoma classification and diagnosis. Since early detection is necessary for treating skin cancer, computer-aided diagnosis with the help of ML is suggested by various studies. The authors [19] have analyzed more than 5000 skin images of more than 3000 patients using the FRCNN model for skin cancer detection. The accuracy of this system for a 6-class classification was 86.2% and for a 2-class classification was 91.5%. The authors [20] suggested using a faster region-based convolutional neural network (RCNN) combined with K-Means clustering to diagnose melanoma. This method achieved an average accuracy of 95.4%, 93.1%, and 95.6% on the datasets collected from ISIC-2016-2017 and PH2. Furthermore, the authors [21] have used various transfer learning models like ResNeT-50, Inception V3, and Inception ResNeT with ESRGAN processing for detecting and

classifying skin cancer on the ISIC2018 dataset. This technique has achieved an accuracy score of 83.7%, 85.8%, and 84%, respectively. The authors [22] have suggested an explainable CNN-based stacked ensemble framework for detecting melanoma at its early stages. This framework achieved an accuracy score of 95.76% in the detection of melanoma. The authors [23] have presented various CNN architectures for detecting skin cancer on the HAM10000 dataset. DenseNet-169 has achieved the highest accuracy of 92.25% for all architectures. The authors [24] have suggested a pretrained deep learning model, MobileNet, for detecting skin cancer. This model achieved an accuracy of 80.81% over the HAM10000 dataset. The authors [25] have developed a raw deep transfer learning model for classifying melanoma into seven categories. This method achieved an accuracy of 82.9%. A novel deep learning technique, InSiNeT, is suggested to detect benign and malignant skin lesions from the HAM10000 dataset [26]. This technique achieved an accuracy of 94.59%, 91.89%, and 90.54%, respectively. The authors [27] have developed a hybrid deep learning model based on the fusion of 3D wavelet transformation and tested this non-invasive technique on the PH2 database. This technique achieved 99.3% accuracy in detecting and classifying skin cancer. The authors [28] have suggested a CNN model to detect and classify skin cancer from the ISIC dataset of 2637 images. This model achieved an accuracy of 88% in classifying skin cancer images as benign or malignant. The authors [29] have suggested using the VGG-16 model to improve the accuracy of the diagnosis of skin cancer. The authors [30] have suggested employing transfer learning based on MobileNet-V2 for detecting melanoma from skin images from the ISIC 2020 dataset. The authors [31] have proposed a CNN-based architecture with magnitude-based weight pruning for detecting skin cancer with an accuracy of 99%. The authors [32] have suggested a diagnostic tool based on computer vision and machine learning algorithms for detecting skin cancer with an accuracy of 79.96%. The authors [33] have suggested an improved deep learning model with CNN for detecting skin cancer in skin images. The presented model outperforms various other models, as per the authors. The authors [34] have suggested a novel data augmentation technique to resolve the challenge of class imbalance and data scarcity based on covariant synthetic minority oversampling technique (SMOTE) along with detection of skin cancer. This technique achieved an accuracy of 92.18%. The authors [35] have implemented the VGG-SegNet scheme to extract the SM section from DDI image to further conduct a relative assessment between the segmented SM and ground truth. This technique was tested on the ISIC2016 database and achieved good results to detect and classify skin cancer. The authors [36] have implemented the VGG-UNet scheme to extract and evaluate the abnormal sections in dermoscopy images with high accuracy. The performance of the suggested scheme is verified using optimizers like Adam or SGD with average pooling. The scheme is tested on the ISIC 2016 dataset and evaluation was done using Jaccard Index, Dice coefficients and accuracy score. The exploration of machine learning (ML) and deep learning (DL) techniques in recent studies has significantly advanced the field of skin cancer diagnosis, showcasing the potential for these technologies to augment and expedite traditional diagnostic processes. Despite these advancements, the integration of ML and DL into clinical practice has been met with challenges that have limited their widespread adoption [37]. A critical examination of previous

studies reveals several limitations inherent to the current methodologies employed in skin cancer detection [38].

Firstly, a recurring issue is the imbalance in training datasets, where the quantity and variety of skin cancer images are insufficient or skewed towards certain types or stages of the disease. This imbalance hampers the model's ability to generalize across the diverse manifestations of skin cancer, leading to potential biases in diagnosis. Moreover, the adaptability of these models across different domains or datasets—referred to as cross-domain adaptability—remains a significant hurdle. Many models are fine-tuned for specific datasets and struggle to maintain their accuracy when applied to data from different sources or demographics. Additionally, the real-world application of these ML and DL techniques often encounters logistical and technical obstacles. The complexity of deploying these models in clinical settings, coupled with the need for substantial computational resources, poses a barrier to their practical use, especially in under-resourced environments. The requirement for high-quality, annotated datasets for training also underscores the need for collaboration between technologists and medical professionals, further complicating the path from research to clinical implementation. In response to these challenges, this paper advocates for the use of ensemble learning as a robust solution to enhance the accuracy and reliability of skin cancer diagnosis. By integrating multiple models, ensemble learning can mitigate the issues of data imbalance and model overfitting or underfitting, offering a more nuanced and comprehensive approach to diagnosis. This technique also presents a solution to the problem of model complexity, as it leverages the strengths of various models to improve diagnostic predictions without the need for excessively high computational resources. In this research, I meticulously address the limitations observed in previous studies by crafting an ensemble learning framework that is specifically tailored for the nuanced task of skin cancer detection. This paper delineates a novel approach where multiple diagnostic models, each with proven efficacy in certain aspects of skin cancer identification, are harmonized to work in concert. This collaborative model framework is not arbitrarily chosen but is the result of an in-depth analysis aimed at selecting models that complement each other's detection capabilities. The ensemble method stands apart due to its deliberate focus on enhancing the ensemble's collective intelligence, thereby significantly improving diagnostic accuracy. I achieved this by employing a rigorous optimization process, ensuring that each model within the ensemble contributes optimally to the final diagnosis outcome. This process involves adjusting and fine-tuning the models based on their performance metrics, with a keen focus on minimizing errors that individual models might introduce. Furthermore, the approach in this work innovatively addresses common challenges in skin cancer detection, such as data imbalance and the model's adaptability to diverse datasets. By integrating models that are robust across various conditions and data representations, the ensemble framework exhibits superior generalizability. This adaptability is critical for real-world applications where data variability is a given. The contribution of this work to the field is underscored by a comprehensive validation process, demonstrating that the ensemble learning framework not only achieves higher diagnostic accuracy but also maintains consistent performance across different testing scenarios. This validation is crucial for establishing the reliability of my method in clinical settings, where consistency and accuracy are paramount.

Table 2. Summary of literature review

Reference	Model/Technique	Key Techniques	Accuracy (%)	Computational Efficiency	Robustness to Noise
[19]	FRCNN	Skin Cancer Detection	86.2 (6-class), 91.5 (2-class)	Medium	High
[20]	RCNN + K-Means	Melanoma Diagnosis	95.4, 93.1, 95.6	High	Medium
[21]	Transfer Learning (ResNeT-50, Inception V3, Inception ResNeT) + ESRGAN	Skin Cancer Detection	83.7, 85.8, 84	High	High
[22]	Explainable CNN-based Stacked Ensemble	Early Melanoma Detection	95.76	Medium	High
[23]	Various CNN Architectures (DenseNeT-169 highlighted)	Skin Cancer Detection	Up to 92.25	High	Medium
[24]	Pretrained MobileNeT	Skin Cancer Detection	80.81	Low	Medium
[25]	Deep Transfer Learning	Melanoma Classification	82.9	Medium	Medium
[26]	InSiNeT	Benign and Malignant Lesion Detection	94.59, 91.89, 90.54	Medium	High
[27]	Hybrid Model (3D Wavelet Transformation)	Skin Cancer Classification	99.3	Medium	High
[28]	CNN Model	Skin Cancer Classification	88	Medium	Medium
[29]	VGG-16	Diagnosis Improvement	Not Specified	High	Medium
[30]	Transfer Learning (MobileNeT-V2)	Melanoma Detection	Not Specified	Low	Medium
[31]	CNN with Weight Pruning	Skin Cancer Detection	99	Medium	High
[32]	Computer Vision and ML Algorithms	Skin Cancer Detection	79.96	Low	Medium
[33]	Improved Deep Learning with CNN	Skin Cancer Detection	Not Specified	High	Medium
[34]	Data Augmentation (SMOTE)	Addressing Data Imbalance	92.18	Low	Medium
[35]	VGG-SegNet	Extracting SM Section	Not Specified	Medium	Medium
[36]	VGG-UNet	Evaluating Dermoscopy Images	Not Specified	Medium	High

In essence, this paper goes beyond presenting another ensemble learning application in the realm of medical diagnostics. It introduces a carefully constructed, optimized ensemble learning framework designed to overcome specific challenges in skin cancer detection. This targeted approach, backed by empirical evidence of its efficacy, marks a significant step forward in the application of machine learning techniques for improving health outcomes. A summary of literature review is discussed under Table 2.

Moreover, the decision to employ ensemble learning is underpinned by its proven efficacy in numerous studies across different domains of cancer diagnosis, where it has consistently outperformed single-model approaches in accuracy, reliability, and generalizability. Reference to similar works in the literature further solidifies the rationale behind my choice. For instance, a study on cervical cancer prediction [39] leveraged ensemble learning to integrate multiple machine learning techniques, resulting in a significant decrease in variance and bias, and an improvement in performance, achieving an accuracy of 87.21%. This aligns with my objective to enhance diagnostic accuracy through the reduction of model errors and the consolidation of strengths from various algorithms. Similarly, in the domain of lung and colon cancer detection, a research work [40] introduced a hybrid ensemble feature extraction model that combined deep feature extraction with ensemble learning, achieving remarkable accuracy rates of 99.05% for lung cancer, 100% for colon cancer, and 99.30% for combined detection. The success of this hybrid model in efficiently identifying cancer from histopathological datasets underscores the potential of ensemble learning in handling complex, image-based

diagnostics by merging deep learning's feature extraction capabilities with ensemble learning's robust classification performance. Moreover, a meta-learning study [41] addressed the challenge of training efficient models with limited labelled data, a common hurdle in medical imaging, by proposing a metric-based meta-learning model that integrates attention mechanisms with ensemble learning. This approach not only enhanced the feature extraction ability but also reduced the risk of overfitting, showcasing ensemble learning's versatility in improving model adaptability and performance in scenarios with scarce data. These studies underscore the adaptability, efficiency, and improved performance of ensemble learning techniques in cancer detection tasks, mirroring the goals of skin cancer detection paper. By drawing on the strengths of various algorithms to minimize weaknesses inherent in single-model approaches, ensemble learning offers a robust solution to the challenges previously outlined.

3. PROPOSED MODEL

This section gives an overview of the suggested methodology. Despite advancements, typical machine learning approaches may fail to achieve high accuracy in dealing with an imbalanced dataset as it is difficult for these techniques to handle multiple data features [42]. To tackle these challenges, ensemble learning is used. This technique helps explore machine learning techniques to get better results by extracting features and fusing the outcomes with various voting mechanisms [43, 44]. The methodology is explained in the upcoming subsections.

3.1 Working of ensemble learning

EL is a technique in which multiple machine learning models (base learners) are integrated, and only a single optimum prediction output is produced, irrespective of their inputs [45-47]. More specifically, the EL technique is classified into generative and non-generative techniques. Non-Generative techniques combine the predictive outcomes of the pretrained models, and these models are trained independently of each other. The ensemble algorithm directs them on how their predictive outcomes are integrated to generate one single output [48-50]. Generative techniques can influence the base learners they are using. These models can tune their algorithms to achieve a higher predictive accuracy [45]. This is explained by Algorithm 1.

Algorithm 1: Ensemble Learning (EL) Algorithm

Pre-Requisite: $X \in$ Training Data, $N \in$ Base Classifiers, $T \in$ Test Data

Output: combined prediction, skin cancer or no skin cancer

1. **for** $j = 1$ to N **do**
 2. Training Set, X_j is created from X
 3. Base classifier, Y_j is created from X
 4. **end**
 5. **for** every test data, $t \in T$ **do**
 6. $Y^*(t) = \text{Vote}(Y_1(t), Y_2(t), \dots, Y_N(t))$
 7. **End**
-

Two of the major non-generative techniques are voting and stacking. Voting is a technique in which various models are allowed to vote to generate a single predictive outcome. In contrast, stacking combines multiple machine learning models through a meta-classifier or regressor [46, 47]. In stacking, the base models are trained on the entire training data, and the meta-model is trained on the predictive outcomes obtained from those base models as features [48, 49]. Generative techniques such as bagging, and boosting are also helpful in achieving high accuracy. Bagging is also defined as bootstrap aggregation, as this technique integrates bootstrapping or sampling of data and aggregation to form an ensemble model [49]. Each base model is trained over different subsets of data, and then their outcomes are combined to form a single output and, in this way, the final outcome is less overfitted. This technique is helpful in bringing stability to the model and, thus, reducing variance [50, 51]. Boosting technique helps in converting weak base learners to strong base learners, with less bias and better accuracy. If an observation is incorrectly classified, the weight of that observation is increased to improve accuracy and vice versa [52]. EL helps in resolving the challenges of bias and variance as faced by various machine learning techniques and is preferred for this study. The bias and variance are defined by Eqs. (1)-(4) as:

$$X[a - b]^2 = bias^2 + \frac{1}{Y} variance + \left(1 - \frac{1}{Y}\right) covariance \quad (1)$$

$$bias(m) = \frac{1}{Y} \sum_j (X[a_j] - b) \quad (2)$$

$$variance(V) = \frac{1}{Y} \sum_j X[a_j - X[a_j]]^2 \quad (3)$$

$$covariance(C) = \frac{1}{Y(Y-1)} \sum_k \sum_{j \neq k} X[a_j] - X[a_j] [a_j - X[a_j]] \quad (4)$$

where, $b =$ target, $a_j =$ output of j th model, $X =$ ensemble size. The bias calculates the average difference between the base learner and the predictive outcome of the model. Variance measures their average variance and covariance helps in measuring the difference between the base learners pairwise.

To build an Ensemble Model, consider a dataset consisting of 'e' examples and 'f' features. This is further defined by $Z = \{a_i, b_i\}$, and $(|Z| = e, a_i \in Y^b, b_i \in Y)$, such that an ensemble model, λ utilizes an aggregate function A , aggregating ϕ inducers, $i_1, i_2, i_3 \dots i_\phi$ and generates a single predictive outcome given by Eq. (5) as:

$$\tilde{b}_i = \phi(a_i) = A(i_1, i_2, i_3 \dots i_\phi) \quad (5)$$

where, $\tilde{b}_i \in Y$ for regression problems and $\tilde{b}_i \in C$ for classification problems [38]. This represents the general framework of EL model. Furthermore, Ensemble Learning (EL) is an advanced technique in machine learning where multiple models, known as base learners, collaborate to produce a singular, optimized prediction output. This technique capitalizes on the strengths of various models to improve predictive performance, especially in complex tasks like skin cancer diagnosis. EL is broadly categorized into generative and non-generative techniques, each with distinct mechanisms for integrating the predictions of base learners. Non-generative ensemble techniques, such as voting and stacking, do not alter the base learners but focus on the strategic combination of their outputs. In voting, the prediction outcome is determined through a majority vote among the models. For example, if three models predict 'skin cancer' and two predict 'no skin cancer,' the final verdict will be 'skin cancer.' Stacking, on the other hand, employs a meta-classifier (or meta-regressor) that learns how to best combine the predictions from multiple models. The base models are trained on the full training dataset, and the meta-model is trained on the outputs of these base models. Generative techniques like bagging and boosting actively modify the training process of base learners to enhance prediction accuracy. Bagging (Bootstrap AGGREGATING) involves training each model on different data subsets, combining their predictions to reduce variance and improve stability. Boosting sequentially trains models, focusing more on instances that previous models misclassified, thereby converting weak learners into strong ones by iteratively adjusting the weights of observations. In this paper, two EL models have been used: i) Ensembling VGG-16 and ResNeT-50 and ii) Ensembling VGG-19 and Xception Model for early detection of Skin cancer. These techniques are explained further in the upcoming subsections.

3.2 Model selection 1: Ensembling based on VGG-16 and ResNeT-50

Lately, it has been found that machine learning helped us save people's lives during the COVID-19 pandemic. With the help of machine learning tools, I was able to diagnose fatal diseases in a less possible time [20]. Similarly, skin cancer is also one of the most fatal diseases and the most challenging to diagnose. This challenge in diagnosis is due to the color

images of the skin, which are almost similar for both benign and malignant cases [53]. Thus, a reliable approach to early diagnosis of skin cancer is required. In this paper, the EL approach is suggested, which can fasten the diagnosis process.

Let the first layer of the EL model be VGG-16. In VGG-16, there are 16 layers, with 3 being used as a convolutional filter, and the rest 13 convolutional layers are used for extracting features. The convolutional process of the layers is defined by Eq. (6) as:

$$F(a, b) = (X * Y)(a, b) \sum_m \sum_n P(m, n) Q(a - m, b - n) \quad (6)$$

Here, m and n are defined as the dimensions of the kernel. a and b are the coordinates of the convolutional matrix [54]. In this network, a ReLU layer works after each convolution layer, which has a maximum pooling layer and is used for sampling.

Consider the dataset with ‘s’ samples with $\alpha(1), \beta(1), \dots, \alpha(s), \beta(s)$ used for training the dataset. Then the overall cost function a (T) is defined by Eq. (7) as:

$$T(F, x) = \left[\frac{1}{s} \sum_{\lambda=1}^s \left(\frac{1}{2} \|X_{f,x}(\alpha^\lambda - \beta^\lambda)\|^2 \right) \right] + \frac{\Phi}{2} \sum_{i=1}^{j_{k-1}} \sum_{\lambda=1}^{L_k} \sum_{n=1}^{L_{k+1}} (F_{\lambda n}^{(i)})^2 \quad (7)$$

Here, $X_{f,x}$ is defined as the model of neural network, $(F_{\lambda n}^{(i)})$ is defined as the connection weight in between the λ th element of the first layer and the n th element of the $k+1$ layer. x is defined as the bias of the neuron of the hidden layer. This equation helps in resolving the challenge of overfitting [55-57]. The Network architecture of VGG-16 is given by Table 3.

Table 3. Network architecture for VGG-16

Block Number	Convolutional Layers	Dimensions	No. of Parameters
Xxxxx	Input	(224 × 224 × 3)	0
1	Conv-Layer 1	(224 × 224 × 64)	1792
1	Conv-Layer 2	(224 × 224 × 64)	36928
1	Max-Pool 1	(112 × 112 × 64)	0
2	Conv-Layer 1	(112 × 112 × 128)	73856
2	Conv-Layer 2	(112 × 112 × 128)	147584
2	Max-Pool 2	(56 × 56 × 128)	0
3	Conv-Layer 1	(56 × 56 × 256)	295168
3	Conv-Layer 2	(56 × 56 × 256)	590080
3	Conv-Layer 3	(56 × 56 × 256)	590080
3	Max-Pool 3	(28 × 28 × 256)	0
4	Conv-Layer 1	(28 × 28 × 512)	1180160
4	Conv-Layer 2	(28 × 28 × 512)	2359808
4	Conv-Layer 3	(28 × 28 × 512)	2359808
4	Max-Pool 4	(14 × 14 × 512)	0
5	Conv-Layer 1	(14 × 14 × 512)	2359808
5	Conv-Layer 2	(14 × 14 × 512)	2359808
5	Conv-Layer 3	(14 × 14 × 512)	2359808
5	Max-Pool 5	(7 × 7 × 512)	0
Total Parameter			14714688

Table 4. General architecture of ResNeT-50

Layer	Size of Output	50 Layers
Convolution-1	(112 × 112)	(7 × 7, 64, stride-2) (3 × 3, max pool, stride-2)
Convolution-2-X	(56 × 56)	1 × 1 64
		3 × 3 64 × 3
Convolution-3-X	(28 × 28)	1 × 1 256
		1 × 1 128 × 4
Convolution-4-X	(14 × 14)	3 × 3 128
		1 × 1 512
Convolution-5-X	(7 × 7)	1 × 1 256
		3 × 3 256 × 6
	(1 × 1)	1 × 1 1024
		1 × 1 512
		3 × 3 512 × 3
		1 × 1 2048
		Average Pooling 1000-d FC, SoftMax

According to the previous study [55], In VGG-16, the input given to the network is in the form of an image with dimension 224 × 224 × 3. The initial 2 layers of the network have 64 channels with a filter size of 3 × 3 with the same padding as explained in Table 2. A max pooling layer of 2 × 2 is added to the network along with two different layers of convolutional filter size of 128 and 3 × 3. Along with this, the input image is

passed through two other sets of three convolutional layers and a max pooling layer of 512 with a 3 × 3 size with the same padding. After this processing, the final feature map of 7 × 7 × 2 is acquired. The output generated by the network is in the form of a vector, $y = y_1, y_2, \dots, y_{999}$ which defines the probability of classification of the class. To make sure that the obtained probabilities sum up to 1, a SoftMax function [58] is

used, which is defined by Eq. (8) as:

$$Z(\lambda_x) = \frac{\exp(\lambda_x)}{\sum_s \exp(\lambda_s)} \quad (8)$$

where, λ_x is defined as the values obtained from the neurons of the output layer. Exp is defined as a nonlinear function. After defining the SoftMax function, I have to minimize between the actual and predicted outcomes. This is defined using the error function or loss function [59] defined by Eq. (9) as:

$$\delta = \frac{1}{y} \sum_q \min_i \beta(\lambda_i, L_Q) \quad (9)$$

where, the values of $\beta = 0$, when $\lambda_i = L_Q$, else the value of $\beta = 1$, λ_i is defined as the ground truth (GT) vectors and L_Q is defined as the probable vector. Once the VGG-16 model is developed, I need to combine it with ResNeT-50.

ResNeT-50 is a neural network, belonging to the class of residual network which is 50 layers deep. In this paper, a pretrained model which is trained on more than a million images from the ImageNet dataset [58] has been used. This network can classify images into a thousand different categories like animals, plants, etc. The general representation of the ResNeT-50 architecture is given by Table 4.

According to the previous study [58], the general network of ResNet-50 contains 3 convolutional layers of size 1×1 , 3×3 , and 1×1 , respectively, and these are passed through a ReLU activation function for generating results. As explained in Table 3, the ResNeT-50 model consists of five convolutional layers with each smaller convolutional layer of type 1,3,4, and 6 layers. The network takes an input as a RGB image of size 224×224 pixels. The first block will generate an output of 64 feature maps with 112×112 pixels. As the convolutional process progresses, there is an increase in the number of features along with the depth of the network. Finally, an output of 7×7 pixels with a feature size of 2048 is

extracted from the network and further classification is done by average pooling of the fully connected layers with SoftMax function as explained by the previous study.

Once the models are developed, the outputs generated by these models need to be integrated into a single output. Two methods which are followed to combine this output are weighing method and meta-learning method. The weighing method is most appropriate for the cases where the performance of the base model is comparable [60, 61]. This approach is applied to the voting mechanism. In the voting mechanism, the arithmetic mean of the predictive outcomes from multiple models are taken if the problem involves regression and for classification, the statistical model is calculated [62]. Another approach for combining the outputs into a single form is meta-learning. Meta learning is defined as the learning algorithm which learn from other learning algorithms. This means that the learning model already learn how to integrate the predictions generated from the base model to a single model [63]. Single machine learning model suffer from various challenges like lack of high-quality training datasets and high time complexity of the algorithms to generate predictions. Meta Learning helps in resolving these challenges by optimizing the base model algorithms and helps in the development of more generalized models. One of the major meta-learning processes which is also utilized in this study is stacking. In the stacking process, the learning models are trained by using the data available. Afterwards, a combiner algorithm is developed to combine all predictions generated, which is also called as the ensemble members, and the final predictions are generated based on this combiner algorithm.

3.3 Model selection 2: Ensembling VGG-19 and Xception

VGG-19 is similar to VGG-16. The only difference is in the number of layers. VGG-19 consists of 19 layers as 16 convolution layers, 3 fully connected layers, 5 max pooling layers, and 1 SoftMax layer. The general architecture of VGG-19 is explained in Table 5.

Table 5. General network architecture of VGG-19

Block No.	Convolutional Layer Type	Output Size	No. of Parameter
xxxx	Input	$(512 \times 512 \times 3)$	0
1	Layer-1	$(512 \times 512 \times 64)$	1,792
1	Layer-2	$(512 \times 512 \times 64)$	36,928
1	Max-Pool	$(256 \times 256 \times 64)$	0
2	Layer-1	$(256 \times 256 \times 128)$	73,856
2	Layer-2	$(256 \times 256 \times 128)$	1,47,584
2	Max-Pool	$(128 \times 128 \times 128)$	0
3	Layer-1	$(128 \times 128 \times 256)$	2,95,168
3	Layer-2	$(128 \times 128 \times 256)$	5,90,080
3	Layer-3	$(128 \times 128 \times 256)$	5,90,080
3	Layer-4	$(128 \times 128 \times 256)$	5,90,080
3	Max-Pool	$(64 \times 64 \times 256)$	0
4	Layer-1	$(64 \times 64 \times 512)$	11,80,160
4	Layer-2	$(64 \times 64 \times 512)$	23,59,808
4	Layer-3	$(64 \times 64 \times 512)$	23,59,808
4	Layer-4	$(64 \times 64 \times 512)$	23,59,808
4	Max-Pool	$(32 \times 32 \times 512)$	0
5	Layer-1	$(32 \times 32 \times 512)$	23,59,808
5	Layer-2	$(32 \times 32 \times 512)$	23,59,808
5	Layer-3	$(32 \times 32 \times 512)$	23,59,808
5	Layer-4	$(32 \times 32 \times 512)$	23,59,808
5	Max-Pool	$(16 \times 16 \times 512)$	0
Total No. of Parameters-		20,02,4384	No. of Non-Trainable Parameters- 0
No. of Trainable Parameters-		20,02,4384	

The VGG-19 architecture, renowned for its depth and simplicity, employs convolutional layers as its core component to analyze visual inputs [64]. These layers work by applying a series of filters (or kernels) across the input image to extract high-level features. Each filter in VGG-19 has a fixed size of 3×3 pixels, which is a deliberate choice to capture the essence of the image's spatial hierarchy with minimal computational complexity. The stride, which refers to the step size the filter moves across the image, is typically set to 1 pixel. This small stride ensures a thorough and detailed scanning of the image, allowing the network to capture fine-grained details by progressively building a comprehensive feature map. As a filter traverses the image, it performs a convolution operation at each position, computing the dot product between the filter's weights and the corresponding input pixels. This process generates a feature map for each filter, highlighting specific attributes in the image, such as edges, textures, or patterns, depending on the filter's learned weights. The VGG-19 model employs multiple such convolutional layers stacked together, each layer capable of detecting increasingly complex features as the depth increases. After the initial feature extraction, the network utilizes pooling layers, specifically max pooling, to downscale the feature maps. Max pooling operates by selecting the maximum value in a 2×2 pixel window with a stride of 2, effectively reducing the dimensionality of the feature maps by half. This reduction serves not only to decrease the computational load and memory usage but also to introduce an element of translation invariance to the model's representation. Following the convolutional and pooling layers, VGG-19 transitions to fully connected layers. The transition is facilitated by a flattening operation, which transforms the 2D feature maps into a 1D vector. This vector serves as input to the fully connected layers, where the network performs high-level reasoning based on the extracted features. The fully connected layers are equipped with weights that are learned during training, allowing the network to combine the detected features in various ways to make predictions. The final layer in the VGG-19 architecture is a fully connected output layer, which employs a softmax activation function to convert the network's outputs into probabilities. Each neuron in this layer corresponds to a class label, and the softmax function ensures that the output values sum up to 1, thus providing a probabilistic interpretation of the model's predictions. Building on the foundational explanation provided for the VGG-19 model, the Xception model introduces a unique approach to convolutional neural networks (CNNs) by incorporating the concept of depth wise separable convolutions into its architecture. With a total of 71 layers and designed to process input images of size 299×299 , the Xception model stands as a deep and complex network structured to enhance feature extraction capabilities beyond traditional CNNs. At the heart of the Xception architecture are 14 distinct groups, each composed of multiple convolutional layers that contribute to a total of 36 convolutional layers dedicated to extracting a wide range of features from the input data. Unlike conventional convolutional layers that simultaneously learn spatial and channel-wise features, the Xception model employs depth wise separable convolutions, a technique that decouples the learning of spatial correlations and channel-wise correlations within the image data. This separation occurs in two stages: first, the model applies depth wise convolutions that separately learn spatial features for each input channel, followed by pointwise convolutions (1×1 convolutions) that combine these features across the channels.

This approach not only increases the model's efficiency by reducing the number of parameters and computational complexity but also enhances its ability to capture more nuanced patterns within the data. A distinguishing feature of the Xception architecture is its use of residual connections, a method that facilitates the flow of information across layers by connecting the input of a group to its output. This is achieved by adding the input to the output of a block before applying the activation function. However, it is noteworthy that these residual connections are present in all but the first and last groups of the model. The absence of residual connections in these groups is by design, to allow the model to perform initial and final transformations of the data without the additive influence from the input. Residual connections are crucial for combating the vanishing gradient problem in deep networks, ensuring that gradients can flow through the network during training, thus enabling the effective training of deeper models without degradation in performance. In the Xception model, each group is structured to perform a specific sequence of operations that progressively increase the representational power of the network. The groups vary in their internal configuration, with differences in the number of depth wise separable convolutions and the presence of intermediate activation functions and pooling layers. This variability allows the model to efficiently learn a broad spectrum of features, from basic to highly complex, by adjusting the focus and depth of feature extraction at each stage. Upon processing through the convolutional groups and depth wise separable convolutions, the Xception model employs an output fusion technique to integrate the extracted features into a coherent output. This technique involves aggregating the outputs from various stages of the model, leveraging both the depth and breadth of the learned features to make a comprehensive prediction. This final output encapsulates the model's interpretation of the input image, categorizing it with a high degree of accuracy based on the learned representations. The schematic representation of Xception model is shown in Figure 1.

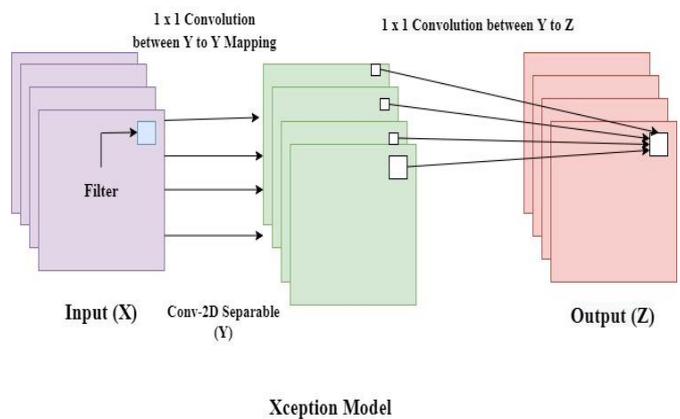


Figure 1. Schematic representation of Xception model

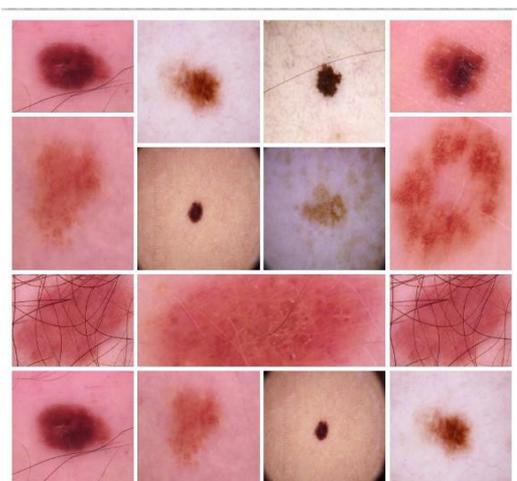
4. RESULT ANALYSIS

This section explains the performance of the models developed with a comparative analysis of the model. The first step for the generation of results is data collection. The data is collected from the previous study [65], which is publicly available. As training an efficient ensemble learning model

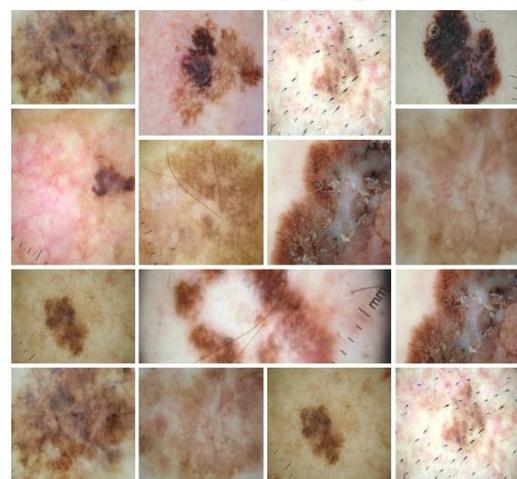
requires high-quality data, this dataset serves the image quality requirements for this study. This dataset contains well-balanced data in the form of images classified into benign and malignant skin images. The dataset includes 2 folders, test and train, consisting of 2 subfolders, benign and malignant, with more than 3000 images of skin with a size of 224×224 . The samples of the data are shown in Figure 2. No data preprocessing is required in this dataset as the images used are already pre-processed. This has saved time and reduced the complexity of the proposed model.

4.1 Results analysis for model 1 (VGG-16 + ResNeT-50)

The first ensemble model is concerned with VGG-16 and ResNeT-50. Initially, the images from the training dataset have been loaded, and prior passing images through the model, data augmentation is performed. The efficiency of the model depends on the quality of the images. However, it has been observed that the representability of the dataset is lost during data preprocessing which might affect the accuracy of the model. Data augmentation helps in resolving the challenge of data representation by reducing data overfitting and class imbalance of the data, which makes the model efficient and yields better prediction accuracy. The data augmentation parameters used for Model-1 (VGG-16 + ResNeT-50) are given in Table 6.



a) Benign images



b) Malignant images

Figure 2. Sample images of benign and malignant skin cancer from the dataset

Table 6. Data augmentation parameters used during model building

Parameters	Values Passed	Remarks
Rotation Range	20	This parameter helps in rotating the images with angle between 0-20 degree.
Width Shift Range	0.01	This parameter helps in shifting the image along the X-axis by the value passed.
Height Shift Range	0.01	This parameter helps in Vertical Shift by the value passed.
Horizontal Flip	False	This parameter helps to stop the flipping of the rows and columns horizontally.
Vertical Flip	False	This parameter helps to stop the flipping of the rows and columns vertically.

After data augmentation, the data is ready to be passed through the model. Initially, the data is passed through the VGG-16 model with default ImageNet weights as explained by the network architecture in Table 2. After this procedure, fine tuning is done. This technique helps us in initializing a new model to be trained on the image data from the same dataset domain by using the weights of a previously trained network. This technique helps in increasing the efficiency of the data training process and helps in overcoming the small data size. Various approaches that are used for fine tuning of models. The first approach is flattening of the layers in the network. This method is used for the conversion of all 2-dimensional arrays from a feature map into a continuous single layer of linear vectors. This flattened layer is fed into the fully connected layer for image classification. Another approach is based on Dense Layer, which is used for image classification based on the outcomes received from the Convolution Layers. This is supported using ReLU function which will give the output only if the input is positive, else it will give a negative value. At last, a dropout function along with sigmoid is used for ignoring the random neurons in the network temporarily during forward passing and the weights are not applied to that neuron during the backward passing. These techniques are further also applied to the ResNeT-50 network and then both models are combined to generate a single ensemble model. In an ensemble network model, an average layer is applicable, which takes the average of each of the feature maps and then sends this average value to the activation layer for further processing. This is summarized in Table 7.

Table 7. Summary of average layer ensemble model-1

Layer	Output Size	Parameter No.	Connection
Input	None \times 512 \times 512 \times 3	0	[]
Functional Model-1	None \times 2	157838850	['input_3[0][0]']
Functional Model	None \times 2	48302530	['input_3[0][0]']
Average	None \times 2	0	['model_1[0][0]', 'model [0][0]']
Total # of Parameters-206,141,380	No. of Trainable Parameters-173,614,340	No. of Non-Trainable Parameters-32,527,040	

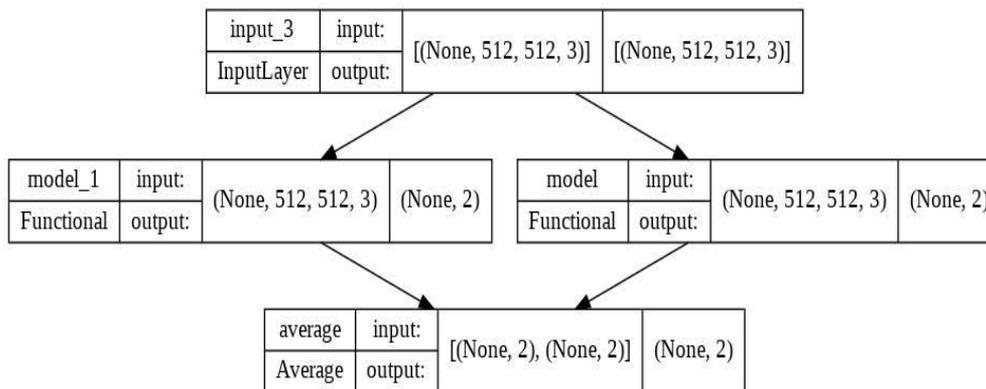


Figure 3. Schematic representation of the model-1 obtained after application of Adam optimizer

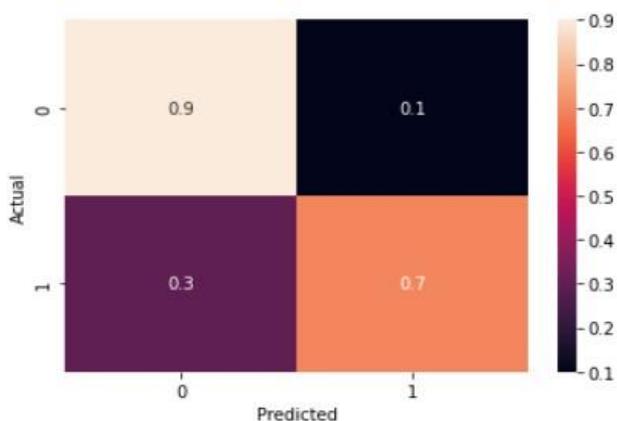


Figure 4. Normalized confusion matrix for the model 1

The ensemble model developed is based on various deep learning algorithms. This model tries to generalize the data received and generates predictions on new unseen data. During this process, I need to take care of minimizing the loss function which might affect the overall accuracy of the model. To resolve this challenge, optimizers are used. There are various optimizers available for various neural network models, but it has been decided to proceed with the Adam optimizer at a learning rate of 0.0001 as it is the most widely used optimizer. Adam optimization algorithm is a replacement algorithm for the SGD optimization algorithm for training various deep learning models. This algorithm can help in resolving the challenges like the handling of sparse gradients on noisy dataset problems by utilizing the properties of AdaGrad algorithm and RMSProp algorithm in a better way [66, 67]. The model obtained after application of Adam optimizer is given in Figure 3.

After optimization, this model is finally trained for 50 epochs of 32 batch size with 0.10 validation split. The results obtained are recorded in Table 8 and recorded in Figures 4 and 5, respectively.

From Table 8, it can be observed that the model achieved 100% accuracy for most of the cases during the training phase. This model achieved a testing accuracy of 80% which is also good, but it can get better. Thus, another ensemble model, based on VGG-19 and Xception, has been employed. Figure 4 tells us about the representation of class groupings having 1.0 samples. This means that the sum of each row represents 100% of the components present in that particular class of dataset.

Table 8. Results obtained with training ensemble model 1

No. of Epochs	Loss	Accuracy	Val_Loss	Val_Accuracy
1	7.9184	0.9722	8.2701	0.7500
2	7.7049	0.8333	8.7855	0.7500
3	7.7997	0.9167	10.4405	0.7500
4	7.2230	0.8889	11.1027	0.7500
5	7.4586	0.8889	10.9496	0.7500
6	7.2470	0.8889	10.8756	0.7500
7	7.3378	0.8333	10.8523	0.7500
8	7.1824	0.8611	10.7030	0.7500
9	6.4601	0.9167	10.3991	0.7500
10	6.7294	0.9167	7.3404	0.7500
11	6.1865	0.9167	6.3422	0.7500
12	6.1645	0.9167	6.2744	0.7500
13	5.9988	0.9444	10.0257	0.7500
14	6.0222	0.9444	8.4231	0.5000
15	6.5633	0.9167	6.4708	0.2500
16	6.0402	0.8611	6.3414	0.2500
17	5.7226	0.9444	6.1048	0.7500
18	5.6395	0.9722	5.9043	0.7500
19	5.6295	0.9167	5.7007	0.7500
20	5.5063	0.9722	5.6630	1.0000
21	5.4763	0.9167	5.6177	1.0000
22	5.3802	1.0000	5.5539	1.0000
23	5.3622	0.9167	5.4564	1.0000
24	5.4855	0.9167	5.3865	1.0000
25	5.2050	1.0000	5.3476	1.0000
26	5.2728	0.9167	5.3121	1.0000
27	5.5851	0.9444	5.2792	1.0000
28	5.2454	0.8611	5.3130	1.0000
29	5.0941	0.9722	5.2884	1.0000
30	5.1055	0.9722	5.2588	0.7500
31	5.0464	1.0000	5.2295	0.7500
32	4.9917	0.9444	5.1960	0.7500
33	4.9486	0.9772	5.0407	0.7500
34	4.9029	0.9444	4.9484	1.0000
35	4.9018	0.9444	4.9198	1.0000
36	4.9155	0.9444	4.8951	1.0000
37	4.8327	0.9722	4.8678	1.0000
38	5.2751	0.9167	4.7622	1.0000
39	4.7475	1.0000	4.6577	1.0000
40	4.7621	1.0000	4.6287	1.0000
41	4.7208	1.0000	4.6070	1.0000
42	4.7405	0.9722	4.5794	1.0000
43	4.6232	1.0000	4.5562	1.0000
44	4.6359	0.9722	4.5329	1.0000
45	5.1113	0.9167	4.5091	1.0000
46	4.8092	0.9722	4.4880	1.0000
47	4.5444	1.0000	4.4694	1.0000
48	4.5612	0.9722	4.4523	1.0000
49	4.4999	1.0000	4.4308	1.0000
50	4.9379	0.9722	4.4109	1.0000

4.2 Evaluation parameters

In evaluating the models presented in this work, I utilized a comprehensive set of metrics each offering unique insights into the model's performance. Precision highlights the model's accuracy in predicting positive outcomes, indicating the likelihood that a positive prediction accurately reflects a true positive case. This metric is crucial in contexts where the cost of a false positive is significant, underscoring the model's reliability in its affirmative diagnoses. Recall or Sensitivity measures the model's capability to identify all actual positive cases, an essential metric in medical diagnostics where failing to detect a condition could have dire consequences. It underscores the model's effectiveness in capturing the entirety of positive cases within the dataset. The F-1 Score, by combining Precision and Recall into a single metric through their harmonic mean, provides a balanced view of the model's overall accuracy. This metric is particularly valuable in situations where the data is imbalanced, ensuring that both the model's precision and its ability to recall positive cases are considered in its evaluation. Sensitivity, also referred to in the context of Recall, is reiterated for its critical role in medical applications, reflecting the model's success in correctly identifying cases with the condition of interest from all cases that actually have the condition. This metric is pivotal for ensuring that patients requiring further investigation or treatment are correctly identified. Specificity complements Sensitivity by measuring the model's ability to correctly identify true negatives, indicating its proficiency in ruling out individuals without the condition. High specificity is vital in reducing the number of false alarms, which can save resources and prevent unnecessary anxiety for patients. Together, these metrics provide a holistic assessment of the models presented in this work, revealing not only their accuracy but also their robustness, reliability, and applicability to real-world scenarios. By carefully analyzing these parameters, I gained valuable insights into the strengths and limitations of models, guiding improvements and ensuring their suitability for clinical application. These are further defined by Eqs. (10)-(14) as:

$$\text{Precision } (P) = \frac{A}{(A + B)} \quad (10)$$

where, A = True Positive (TP) and B = False Positive (FP).

$$\text{Recall } (R) = \frac{X}{(X + Y)} \quad (11)$$

where, X = True Positive (TP) and Y = False Negative (FN).

$$F - \text{Score} = \frac{(2 * P * R)}{(P + R)} \quad (12)$$

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (13)$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \quad (14)$$

The results obtained through the evaluation parameters are recorded in Table 9.

Table 9. Model evaluation results

	Precision	Recall	F-1 Score	Support
Benign (0)	0.75	0.90	0.82	20
Malignant (1)	0.88	0.70	0.78	20
Macro Average	0.81	0.80	0.80	40
Weighted Average	0.81	0.80	0.80	40
Specificity = 0.7		Sensitivity = 0.9		

4.3 Results analysis for model 2 (VGG-19 + Xception)

To keep a fair analysis, all parameters of Model 2 are kept same as the parameters of Model 1. Thus, the same data augmentation techniques as described in Table 5 have been applied. In this model, the same Adam optimizer for results optimization has been used. The architecture of the VGG-19 and Xception Ensemble Average Mode is given by Table 10 and Figure 5.

Upon Training Model-2 (VGG-19 + Xception), it has been found that this model also achieved the maximum 100% accuracy for most cases over the training dataset and 85% accuracy over the test dataset. This is better than Model-1. This accuracy is acceptable as training an efficient model requires a high quality of the dataset. In the real world, there is a large dataset with high quality of skin images, which when trained with this model can increase the accuracy percentage of the model in the diagnosis of skin cancer. Due to the restricted capacity of computational resources and less amount of data compared to the real world, 85% accuracy is acceptable to diagnose skin cancer. Furthermore, the analysis carried out in this paper also shows that this model outperforms various other single model diagnosis algorithms, which is explained through a comparative analysis in the following subsections.

The results obtained after training of ensemble model 2 (VGG-19 + Xception) for 50 epochs of batch size 32 with a validation split of 0.10 are recorded in Table 11 and Figure 6 respectively.

These results clearly indicate that Ensemble Learning model 2 based on VGG-19 and Xception is more efficient than model 1, VGG-16, and ResNeT-50 in the diagnosis and classification of skin cancer. For evaluating the results of the proposed work, certain parameters are evaluated for the model. For a fair analysis, the same parameters as used for model 1 for evaluation have been used. These are specificity, sensitivity, f-1 score, recall, and precision. These are summarized in Table 12.

Table 10. Summary of average layer ensemble model-2

Layer	Output Size	Parameter No.	Connection
Input	None × 512 × 512 × 3	0	[]
Functional Model-1	None × 2	155112618	['input_3[0][0]']
Functional Model	None × 2	53612226	['input_3[0][0]']
Average	None × 2	0	['model_1[0][0]', 'model [0][0]']
Total # of Parameters-208,724,844	No. of Trainable Parameters-175,725,316	No. of Non-Trainable Parameters-32,999,528	

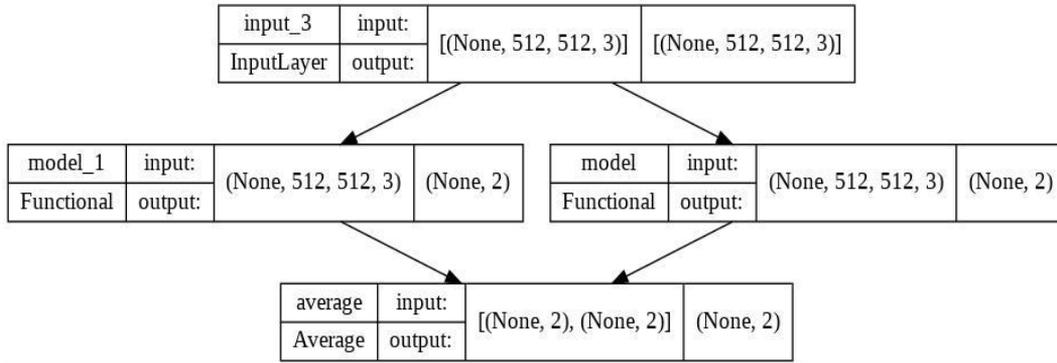


Figure 5. Schematic representation of the model 2 obtained after application of Adam optimizer

Table 11. Results obtained with training model 2

Epochs	Loss	Accuracy	Val_Loss	Val_Accuracy
1	14.3932	0.5278	13.6940	0.2500
2	13.1740	0.7222	12.1494	0.7500
3	12.2389	0.8611	11.2263	1.0000
4	11.4939	0.8889	10.6502	0.7500
5	10.8050	0.8889	10.0337	0.7500
6	10.4299	0.7222	9.4820	0.7500
7	9.5462	0.9167	8.9987	0.7500
8	8.9664	0.9444	8.5766	0.7500
9	8.4865	0.9444	8.2008	0.7500
10	8.5989	0.8333	7.8598	0.7500
11	7.8468	0.9444	7.5094	1.0000
12	7.5419	0.8889	7.0755	1.0000
13	7.2033	0.9722	6.8166	1.0000
14	6.9452	0.9722	6.5963	1.0000
15	6.7158	0.9722	6.3986	1.0000
16	6.6310	0.8889	6.2222	1.0000
17	6.3908	0.9444	6.0642	1.0000
18	6.2411	0.9444	5.9211	1.0000
19	6.0651	0.9722	5.7914	1.0000
20	6.1308	0.9444	5.6738	1.0000
21	5.8599	0.9444	5.5664	1.0000
22	5.6861	0.9444	5.4676	1.0000
23	5.6605	0.9444	5.3763	1.0000
24	5.5848	0.9722	5.2931	1.0000
25	5.4014	0.9722	5.2179	1.0000
26	5.3108	0.9722	5.1498	1.0000
27	5.2327	1.0000	5.0895	1.0000
28	5.2453	0.9444	5.0506	1.0000
29	5.0620	1.0000	5.0315	1.0000
30	5.0305	1.0000	5.0736	1.0000
31	4.9609	1.0000	5.3581	0.7500
32	4.9658	0.9167	5.3692	0.7500
33	4.8513	1.0000	5.3784	0.7500
34	4.8002	1.0000	5.6197	0.5000
35	4.7846	0.9722	5.6277	0.5000
36	5.1766	0.8889	5.2515	0.7500
37	4.6724	1.0000	5.0784	0.7500
38	4.6634	0.9444	5.0112	0.7500
39	4.6015	1.0000	4.9556	0.7500
40	4.6113	0.9444	4.9209	0.7500
41	4.5215	0.9722	4.9324	0.7500
42	4.4712	0.9722	4.7741	0.7500
43	4.4662	1.0000	4.4138	1.0000
44	4.4226	1.0000	4.3695	1.0000
45	4.3856	0.9722	4.3442	1.0000
46	4.3560	1.0000	4.3211	1.0000
47	4.3291	0.9722	4.2964	1.0000
48	4.2721	1.0000	4.2726	1.0000
49	4.2773	0.9722	4.2600	1.0000
50	4.2119	1.0000	4.4117	0.7500

Table 12. Model evaluation results

	Precision	Recall	F-1 Score	Support
Benign (0)	0.72	0.90	0.80	20
Malignant (1)	0.87	0.65	0.74	20
Macro Average	0.84	0.85	0.85	40
Weighted Average	0.84	0.85	0.85	40
Specificity = 0.65		Sensitivity = 0.9		

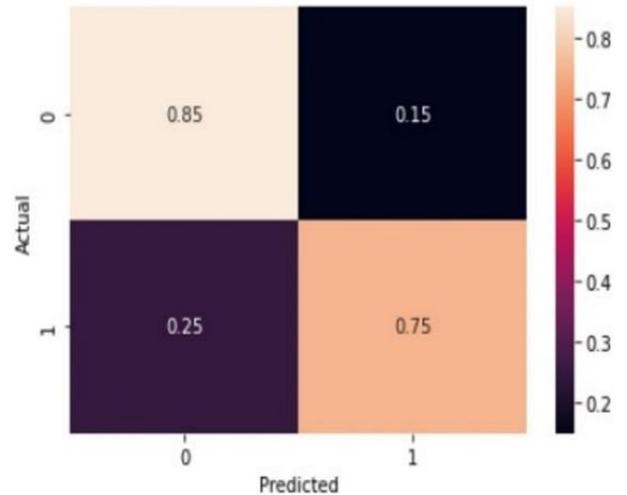


Figure 6. Normalized confusion matrix for model 2

To verify that EL technique can fasten the diagnosis process with good accuracy, I have conducted a comparative analysis of single developed models with the best performing model, that is, model 2 based on VGG-19 + Xception. In order to develop effective algorithms to detect and classify skin cancer, few criteria are required to be followed. These are; the model should be trained on a large dataset to ensure reliability of the model. The accuracy of the developed model should be good so that it could be used in real world. The model should be less complex so that its working should be easily understood by the users and to operate it to its full potential. The model developed by us fulfils all these criterias and is more efficient model than various closely related studies. For example, a skin cancer detection and classification model using transfer learning is suggested [68, 69]. The model follows a two-step process for classification of skin cancer. However, the classification accuracy decreases from 85% to 75% during two-step process and thus is not efficient enough to be used in real world. A mobile enabled classification tool based on computer vision is suggested for classification of skin cancer

in real time [66, 67]. The application achieved a sensitivity of 80% and specificity of 75%, whereas, the models suggested by us is more effective in terms of sensitivity, which is 90% for the best model. The suggested model achieved a classification accuracy of 84% which makes the technique unfit to be used for long term [70]. Various lightweight deep learning techniques are suggested to classify skin cancer [71]. However, the accuracy of the suggested models are lower, ranging from 78% to 84%, making them unreliable to be implemented in real world. In the previous study [72], the suggested technique achieved an overall classification accuracy of 84.09% and the technique suggested in the previous study achieved an accuracy of 76% [73]. The accuracy of the techniques suggested is lower than the technique suggested in the work done by us and thus the model presented in this work is more reliable to be used in real world. A comparison between the models is summarized by Table 13.

According to the table, I can conclude that the model suggested by us surpasses many other models and is more effective in detection and classification of skin cancer. Despite of the advantages of the technique suggested in the paper, there are certain limitations which are needed to be resolved. The ensemble learning technique, while offering significant advantages in skin cancer diagnosis, faces challenges such as high resource requirements and complex model architecture, which could hinder rapid deployment and troubleshooting. To address the resource intensity of training ensemble models, future work could explore the adoption of lightweight models or streamlined ensemble methods that maintain high accuracy while being less demanding computationally. This approach could make the technology more accessible, especially in resource-constrained settings where urgent diagnostic needs are prevalent. Improving the model's accuracy beyond the current threshold is another critical area for future development. Strategies such as augmenting the training dataset with more diverse and extensive data, or fine-tuning the models with advanced optimization techniques, could enhance the model's performance. Additionally, incorporating more sophisticated data augmentation methods could help the model learn from a wider variety of skin cancer manifestations, thereby increasing its generalizability and accuracy. Regarding the complexity of the ensemble model architecture, future efforts might focus on simplifying the model without compromising its predictive power. Simplification could involve identifying and retaining only the most impactful features and models within the ensemble, thereby easing the investigation process in case of model failure. Moreover, developing more interpretable models or employing explainable AI techniques could also aid in diagnosing and rectifying issues more efficiently, reducing time wastage and improving model trustworthiness. By tackling these limitations through targeted research and development, I can enhance the practicality and efficacy of ensemble learning models for skin cancer diagnosis, paving the way for their broader adoption and implementation in clinical settings. Furthermore, building on the identified limitations of current ensemble learning technique for skin cancer diagnosis, my future work will focus on concrete steps and methodologies aimed at overcoming these challenges. To address the high resource consumption and complexity inherent in training ensemble models, I plan to investigate and implement more efficient machine learning frameworks. Specifically, I will explore lightweight neural network architectures that are known for their reduced computational demand without

significantly compromising accuracy. This will involve comparative analysis to identify architectures that offer an optimal balance between performance and resource efficiency. To enhance the accuracy of the models presented in this work beyond the current levels, this work approach will include expanding the training datasets with a broader range of skin cancer images, encompassing more diverse skin types and cancer stages. This expansion aims to improve the model's generalization capabilities across various manifestations of skin cancers. Additionally, I will apply more sophisticated data augmentation techniques to artificially enlarge the dataset, thereby providing the models presented in this work with a more comprehensive learning experience.

In tackling the complexity of the model architecture, I will embark on a systematic simplification process. This process will involve pruning fewer effective components of the ensemble to streamline the architecture, making it more transparent and easier to analyze in case of failure. Concurrently, I will integrate explainable AI methodologies to enhance the interpretability of the models presented in this work. This effort will not only facilitate a deeper understanding of the model's decision-making processes but also significantly ease the investigation and rectification of issues should they arise. Future work will also include rigorous testing and validation of these improvements to ensure they effectively address the current limitations without introducing new challenges. Through these focused efforts, I aim to advance the development of ensemble learning models, making them more accessible, accurate, and user-friendly for clinical applications in skin cancer diagnosis.

Table 13. Comparison of the proposed models with the state-of-the-art

Recent Works	Accuracy (%)
[68]	85% for Step-1 Transfer Learning and 75% for Step-2 Transfer Learning
[69]	80% sensitivity and 75% specificity
[70]	84%
[71]	78.9% for SqueezeNet, 76% for ShuffleNet, 83.1% for ResNet, 82.9% for MobileNetV1 and 83.7% for DenseNet
[72]	84.09% for ResNet101 architecture
[73]	76%
[74]	76.87% using Random Forest Classifier
[75]	80%
[76]	80%
[77]	80% training and 70% Testing
Proposed models	80% for Model-1 and 85% for Model-2

5. CONCLUSIONS

Skin cancer stands as the most prevalent form of cancer, yet it remains highly treatable when identified in its nascent stages. It has been documented that over 75% of skin cancer fatalities stem from delayed detection. Presently, the diagnosis and classification of skin cancer predominantly rely on manual methods, which are hampered by several limitations including the expertise of radiologists, the complexity of laboratory setups, and the protracted duration required to obtain test results. In response to these challenges, this paper proposes the adoption of an Ensemble Learning technique to expedite and automate the diagnostic process. While not intended to

supplant existing diagnostic frameworks, this method significantly enhances the speed of diagnosis, achieving an impressive accuracy rate of 85%. This achievement, however, is currently bounded by the limitations of computational resources and the quality of the dataset available. With access to advanced computational resources and a high-quality image dataset, there is a tangible pathway to not only refine the accuracy of the model but potentially elevate it beyond the 85% threshold observed. The infusion of superior datasets would enable the model to learn from a more diverse and richer set of image features, thereby improving its diagnostic precision. In practical terms, the utility of the proposed Ensemble Learning method extends beyond theoretical applications, offering substantial benefits in real-world clinical settings. For instance, integrating this method within telemedicine platforms could drastically reduce the wait times for skin cancer screening results, making early intervention more feasible. Additionally, its application in remote areas, where specialist dermatological expertise is scarce, could democratize access to reliable diagnostic services. Looking ahead, the scope for further research is vast and promising. Future investigations could delve into the exploration of alternative ensemble learning strategies that may offer even greater efficiencies or accuracy improvements. Optimizing the current models to be more resource-efficient without sacrificing accuracy presents another fertile ground for exploration. Moreover, integrating the proposed technique with existing diagnostic tools could provide a more holistic and robust diagnostic procedure, blending the strengths of manual expertise with the precision of machine learning. The potential impact of this work extends beyond the realm of skin cancer, offering promising implications for other types of cancer and medical conditions. The adaptability of the Ensemble Learning approach, with appropriate modifications, could serve as a versatile tool in the broader medical diagnostic field, heralding a new era of rapid, accurate, and accessible disease detection and classification. In conclusion, this paper not only showcases the efficacy of Ensemble Learning in improving skin cancer diagnosis but also lays the groundwork for a broad spectrum of future research directions. By leveraging better resources and datasets, the proposed method has the potential to significantly advance the accuracy and reliability of skin cancer diagnostics, with far-reaching implications for healthcare delivery worldwide.

REFERENCES

- [1] Rezaoana, N., Hossain, M.S., Andersson, K. (2020). Detection and classification of skin cancer by using a parallel CNN model. In 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Bhubaneswar, India, 2020, pp. 380-386. <https://doi.org/10.1109/WIECON-ECE52138.2020.9397987>
- [2] Dildar, M., Akram, S., Irfan, M., Khan, H.U., Ramzan, M., Mahmood, A.R., Alsaiari, S.A., Saeed, A.H.M., Alraddadi, M.O., Mahnashi, M.H. (2021). Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health*, 18(10): 5479. <https://doi.org/10.3390/ijerph18105479>
- [3] Wikipedia contributors. (2022). Skin cancer. In Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Skin_cancer&oldid=1104284995, accessed on Sep. 3, 2022.
- [4] Höhn, J., Hekler, A., Kriehoff-Henning, E., Kather, J.N., Utikal, J.S., Meier, F., Gellrich, F.F, Hauschild, A., French, L., Schlager, J.G., Ghoreschi, K., Wilhelm, T., Kutzner, H., Heppt, M., Haferkamp, S., Sondermann, W., Schadendorf, D., Schilling, B., Maron, R.C., Schmitt, M., Jutzi, T., Fröhling, S., Lipka, D.B., Brinker, T.J. (2021). Integrating patient data into skin cancer classification using convolutional neural networks: Systematic review. *Journal of Medical Internet Research*, 23(7): e20708. <https://doi.org/10.2196/20708>
- [5] Attique Khan, M., Sharif, M., Akram, T., Kadry, S., Hsu, C.H. (2022). A two-stream deep neural network-based intelligent system for complex skin cancer types classification. *International Journal of Intelligent Systems*, 37(12): 10621-10649. <https://doi.org/10.1002/int.22691>
- [6] Daghbir, J., Tlig, L., Bouchouicha, M., Sayadi, M. (2020). Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. In 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, pp. 1-5. <https://doi.org/10.1109/ATSIP49331.2020.9231544>
- [7] Vidya, M., Karki, M.V. (2020). Skin cancer detection using machine learning techniques. In 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, pp. 1-5. <https://doi.org/10.1109/CONECCT50063.2020.9198489>
- [8] Murugan, A., Nair, S.A.H., Preethi, A.A.P., Kumar, K.S. (2021). Diagnosis of skin cancer using machine learning techniques. *Microprocessors and Microsystems*, 81: 103727. <https://doi.org/10.1016/j.micpro.2020.103727>
- [9] Javaid, A., Sadiq, M., Akram, F. (2021). Skin cancer classification using image processing and machine learning. In 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, pp. 439-444. <https://doi.org/10.1109/IBCAST51254.2021.9393198>
- [10] Kumaraswamy, E. (2022). Key challenges in the diagnosis of cancer using artificial intelligence methods. In *AIP Conference Proceedings*. AIP Publishing LLC, 2418(1): 030049. <https://doi.org/10.1063/5.0081712>
- [11] Cheung, H.M.C., Rubin, D. (2021). Challenges and opportunities for artificial intelligence in oncological imaging. *Clinical Radiology*, 76(10): 728-736. <https://doi.org/10.1016/j.crad.2021.03.009>
- [12] Bertsimas, D., Wiberg, H. (2020). Machine learning in oncology: Methods, applications, and challenges. *JCO Clinical Cancer Informatics*. <https://doi.org/10.1200/CCI.20.00072>
- [13] Tufail, A.B., Ma, Y.K., Kaabar, M.K., Martínez, F., Junejo, A.R., Ullah, I., Khan, R. (2021). Deep learning in cancer diagnosis and prognosis prediction: A minireview on challenges, recent trends, and future directions. *Computational and Mathematical Methods in Medicine*, 2021(1): 9025470. <https://doi.org/10.1155/2021/9025470>
- [14] Sagi, O., Rokach, L. (2018). Ensemble learning: A survey. *Data Mining and Knowledge Discovery*, 8(4): e1249. <https://doi.org/10.1002/widm.1249>

- [15] Putra, T.A., Rufaida, S.I., Leu, J.S. (2020). Enhanced skin condition prediction through machine learning using dynamic training and testing augmentation. *IEEE Access*, 8: 40536-40546. <https://doi.org/10.1109/ACCESS.2020.2976045>
- [16] Hameed, N., Shabut, A., Hossain, M.A. (2018). A computer-aided diagnosis system for classifying prominent skin lesions using machine learning. In 2018 10th Computer Science and Electronic Engineering (CEECE), Colchester, UK, pp. 186-191. <https://doi.org/10.1109/CEECE.2018.8674183>
- [17] Goyal, M., Oakley, A., Bansal, P., Dancey, D., Yap, M.H. (2019). Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *IEEE Access*, 8: 4171-4181. <https://doi.org/10.1109/ACCESS.2019.2960504>
- [18] Das, K., Cockerell, C.J., Patil, A., Pietkiewicz, P., Giuliani, M., Grabbe, S., Goldust, M. (2021). Machine learning and its application in skin Cancer. *International Journal of Environmental Research and Public Health*, 18(24): 13409. <https://doi.org/10.3390/ijerph182413409>
- [19] Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., Hamamoto, R. (2020). The development of a skin cancer classification system for pigmented skin lesions using deep learning. *Biomolecules*, 10(8): 1123. <https://doi.org/10.3390/biom10081123>
- [20] Nawaz, M., Mehmood, Z., Nazir, T., Naqvi, R.A., Rehman, A., Iqbal, M., Saba, T. (2022). Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. *Microscopy Research and Technique*, 85(1): 339-351. <https://doi.org/10.1002/jemt.23908>
- [21] Gouda, W., Sama, N.U., Al-Waakid, G., Humayun, M., Jhanjhi, N.Z. (2022). Detection of skin cancer based on skin lesion images using deep learning. *Healthcare*, 10(7): 1183. <https://doi.org/10.3390/healthcare10071183>
- [22] Shorfuzzaman, M. (2022). An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection. *Multimedia Systems*, 28(4): 1309-1323. <https://doi.org/10.1007/s00530-021-00787-5>
- [23] Kousis, I., Perikos, I., Hatzilygeroudis, I., Virvou, M. (2022). Deep learning methods for accurate skin cancer recognition and mobile application. *Electronics*, 11(9): 1294. <https://doi.org/10.3390/electronics11091294>
- [24] Agrahari, P., Agrawal, A., Subhashini, N. (2022). Skin cancer detection using deep learning. In *Futuristic Communication and Network Technologies*. Springer, Singapore, 792: 179-190. https://doi.org/10.1007/978-981-16-4625-6_18
- [25] Fraiwan, M., Faouri, E. (2022). On the automatic detection and classification of skin cancer using deep transfer learning. *Sensors*, 22(13): 4963. <https://doi.org/10.3390/s22134963>
- [26] Reis, H.C., Turk, V., Khoshelham, K., Kaya, S. (2022). InSiNet: A deep convolutional approach to skin cancer detection and segmentation. *Medical & Biological Engineering & Computing*, 60(3): 643-662. <https://doi.org/10.1007/s11517-021-02473-0>
- [27] Maniraj, S.P., Maran, P.S. (2022). A hybrid deep learning approach for skin cancer diagnosis using subband fusion of 3D wavelets. *The Journal of Supercomputing*, 78(10): 12394-12409. <https://doi.org/10.1007/s11227-022-04371-0>
- [28] Hasan, M.R., Fatemi, M.I., Monirujjaman Khan, M., Kaur, M., Zaguia, A. (2021). Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks. *Journal of Healthcare Engineering*, 2021(1): 5895156. <https://doi.org/10.1155/2021/5895156>
- [29] Tabrizchi, H., Parvizpour, S., Razmara, J. (2023). An improved VGG model for skin cancer detection. *Neural Processing Letters*, 55(4): 3715-3732. <https://doi.org/10.1007/s11063-022-10927-1>
- [30] Rashid, J., Ishfaq, M., Ali, G., Saeed, M.R., Hussain, M., Alkhalifah, T., Alturise, F., Samand, N. (2022). Skin cancer disease detection using transfer learning technique. *Applied Sciences*, 12(11): 5714. <https://doi.org/10.3390/app12115714>
- [31] Cabrejos-Yalán, V.M., Rosales-Huamani, J.A., Arenas-Ñiquin, J.L. (2022). Optimization of a deep learning model for skin cancer detection with magnitude-based weight pruning. In *World Conference on Information Systems and Technologies*. Cham: Springer International Publishing, pp. 624-629. https://doi.org/10.1007/978-3-031-04826-5_61
- [32] Parshionkar, S., Koshy, R., Sheikh, A., Phansalkar, G. (2022). Skin cancer detection and severity prediction using computer vision and deep learning. In *Second International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2021*, Springer Singapore, pp. 295-304. https://doi.org/10.1007/978-981-16-4641-6_25
- [33] Lafraxo, S., Ansari, M.E., Charfi, S. (2022). MelaNet: An effective deep learning framework for melanoma detection using dermoscopic images. *Multimedia Tools and Applications*, 81(11): 16021-16045. <https://doi.org/10.1007/s11042-022-12521-y>
- [34] Abayomi-Alli, O.O., Damasevicius, R., Misra, S., Maskeliunas, R., Abayomi-Alli, A. (2021). Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold. *Turkish Journal of Electrical Engineering and Computer Sciences*, 29(8): 2600-2614. <https://doi.org/10.3906/elk-2101-133>
- [35] Kadry, S., Taniar, D., Damaševičius, R., Rajinikanth, V., Lawal, I.A. (2021). Extraction of abnormal skin lesion from dermoscopy image using VGG-SegNet. In *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, Chennai, India, pp. 1-5. <https://doi.org/10.1109/ICBSII51839.2021.9445180>
- [36] Rajinikanth, V., Kadry, S., Damaševičius, R., Sankaran, D., Mohammed, M.A., Chander, S. (2022). Skin melanoma segmentation using VGG-UNet with Adam/SGD optimizer: A study. In *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, Kannur, India, pp. 982-986. <https://doi.org/10.1109/ICICICT54557.2022.9917848>
- [37] González-Cruz, C., Jofre, M.A., Podlipnik, S., Combalia, M., Gareau, D., Gamboa, M., Vallone M.G., Faride Barragán-Estudillo Z., Tamez-Peña A.L., Montoya J., América Jesús-Silva M., Carrera C., Malveyh J., Puig S. (2020). Machine learning in melanoma diagnosis. Limitations about to be overcome. *Actas Dermo-Sifiliográficas (English Edition)*, 111(4): 313-316. <https://doi.org/10.1016/j.adengl.2019.09.003>
- [38] Wu, Y., Chen, B., Zeng, A., Pan, D., Wang, R., Zhao, S.

- (2022). Skin cancer classification with deep learning: A systematic review. *Frontiers in Oncology*, 12: 893972. <https://doi.org/10.3389/fonc.2022.893972>
- [39] Ahishakiye, E., Wario, R., Mwangi, W., Taremwa, D. (2020). Prediction of cervical cancer basing on risk factors using ensemble learning. In 2020 IST-Africa Conference (IST-Africa), Kampala, Uganda, pp. 1-12.
- [40] Haq, I.U., Ali, H., Wang, H.Y., Lei, C., Ali, H. (2022). Feature fusion and ensemble learning-based CNN model for mammographic image classification. *Journal of King Saud University-Computer and Information Sciences*, 34(6): 3310-3318. <https://doi.org/10.1016/j.jksuci.2022.03.023>
- [41] Guo, N., Di, K., Liu, H., Wang, Y., Qiao, J. (2021). A metric-based meta-learning approach combined attention mechanism and ensemble learning for few-shot learning. *Displays*, 70: 102065. <https://doi.org/10.1016/j.displa.2021.102065>
- [42] Dong, X., Yu, Z., Cao, W., Shi, Y., Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2): 241-258. <https://doi.org/10.1007/s11704-019-8208-z>
- [43] Yang, Y., Lv, H., Chen, N. (2021). A survey on ensemble learning under the era of deep learning. *Artificial Intelligence Review*, 56(6): 5545-5589. <https://doi.org/10.1007/s10462-022-10283-5>
- [44] Gomes, H.M., Barddal, J.P., Enembreck, F., Bifet, A. (2017). A survey on ensemble learning for data stream classification. *ACM Computing Surveys (CSUR)*, 50(2): 1-36. <https://doi.org/10.1145/3054925>
- [45] Wikipedia contributors. (2022). Ensemble learning. In Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Ensemble_learning&oldid=1100411098, accessed on Sep. 7, 2022.
- [46] Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., Suganthan, P.N. (2021). Ensemble deep learning: A review. *arXiv preprint arXiv: 2104.02395*. <https://doi.org/10.48550/arXiv.2104.02395>
- [47] Liu, Y., Zhao, Q. (2022). Ensemble learning. In *HANDBOOK ON COMPUTER LEARNING AND INTELLIGENCE: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation*, pp. 635-660. https://doi.org/10.1142/9789811247323_0016
- [48] Talukder, M.A., Islam, M.M., Uddin, M.A., Akhter, A., Hasan, K.F., Moni, M.A. (2022). Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. *Expert Systems with Applications*, 205: 117695. <https://doi.org/10.1016/j.eswa.2022.117695>
- [49] Famitha, S., Moorthi, M. (2022). Intelligent and novel multi-type cancer prediction model using optimized ensemble learning. *Computer Methods in Biomechanics and Biomedical Engineering*, 25(16): 1879-1903. <https://doi.org/10.1080/10255842.2022.2081504>
- [50] Ngo, G., Beard, R., Chandra, R. (2022). Evolutionary bagging for ensemble learning. *Neurocomputing*, 510: 1-14. <https://doi.org/10.1016/j.neucom.2022.08.055>
- [51] Zounemat-Kermani, M., Batelaan, O., Fadaee, M., Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, 598: 126266. <https://doi.org/10.1016/j.jhydrol.2021.126266>
- [52] Guo, R., Fu, D., Sollazzo, G. (2021). An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree. *International Journal of Pavement Engineering*, 23(10): 3633-3646. <https://doi.org/10.1080/10298436.2021.1910825>
- [53] Hosny, K.M., Kassem, M.A., Foad, M.M. (2018). Skin cancer classification using deep learning and transfer learning. In 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, pp. 90-93. <https://doi.org/10.1109/CIBEC.2018.8641762>
- [54] Sahinbas, K., Catak, F.O. (2021). Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images. In *Data Science for COVID-19*, Academic Press, pp. 451-466. <https://doi.org/10.1016/B978-0-12-824536-1.00003-4>
- [55] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
- [56] Selimović, A., Meden, B., Peer, P., Hladnik, A. (2018). Analysis of content-aware image compression with VGG16. In 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOB), San Carlos, Costa Rica, pp. 1-7. <https://doi.org/10.1109/IWOB.2018.8464188>
- [57] Munshi, R.M., Cascone, L., Alturki, N., Saidani, O., Alshardan, A., Umer, M. (2024). A novel approach for breast cancer detection using optimized ensemble learning framework and XAI. *Image and Vision Computing*, 142: 104910. <https://doi.org/10.1016/j.imavis.2024.104910>
- [58] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [59] Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, 9(10): 143-150. <http://doi.org/10.29322/IJSRP.9.10.2019.p9420>
- [60] Chatzitheodoridou, E. (2022). Brain tumor grade classification in MR images using deep learning. Department of Computer and Information Science, The Division of Statistics and Machine Learning, Linköping University.
- [61] Krawczyk, B., Minku, L.L., Gama, J., Stefanowski, J., Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. *Information Fusion*, 37: 132-156. <https://doi.org/10.1016/j.inffus.2017.02.004>
- [62] Guan, D., Yuan, W., Lee, Y.K., Najeebullah, K., Rasel, M.K. (2014). A review of ensemble learning based feature selection. *IETE Technical Review*, 31(3): 190-198. <https://doi.org/10.1080/02564602.2014.906859>
- [63] Gu, K., Zhang, Y., Qiao, J. (2020). Ensemble meta-learning for few-shot soot density recognition. *IEEE Transactions on Industrial Informatics*, 17(3): 2261-2270. <https://doi.org/10.1109/TII.2020.2991208>
- [64] Abuared, N., Panthakkan, A., Al-Saad, M., Amin, S.A., Mansoor, W. (2020). Skin cancer classification model based on VGG 19 and transfer learning. In 2020 3rd International Conference on Signal Processing and Information Security (ICSPIS), DUBAI, United Arab Emirates, pp. 1-4. <https://doi.org/10.1109/ICSPIS51252.2020.9340143>

- [65] Singh, P., Kumar, M., Bhatia, A. (2022). A comparative analysis of deep learning algorithms for skin cancer detection. In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 1160-1166. <https://doi.org/10.1109/ICICCS53718.2022.9788197>
- [66] Moldovan, D. (2019). Transfer learning based method for two-step skin cancer images classification. In 2019 E-Health and Bioengineering Conference (EHB), Iasi, Romania, pp. 1-4. <https://doi.org/10.1109/EHB47216.2019.8970067>
- [67] Taufiq, M.A., Hameed, N., Anjum, A., Hameed, F. (2017). m-Skin Doctor: A mobile enabled system for early melanoma skin cancer detection using support vector machine. In eHealth 360°: International Summit on eHealth, Budapest, Hungary, pp. 468-475. https://doi.org/10.1007/978-3-319-49655-9_57
- [68] Zhang, Z. (2018). Improved Adam optimizer for deep neural networks. In 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, pp. 1-2. <https://doi.org/10.1109/IWQoS.2018.8624183>
- [69] Bock, S., Weiß, M. (2019). A proof of local convergence for the Adam optimizer. In 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, pp. 1-8. <https://doi.org/10.1109/IJCNN.2019.8852239>
- [70] Ameri, A. (2020). A deep learning approach to skin cancer detection in dermoscopy images. *Journal of Biomedical Physics and Engineering*, 10(6): 801-806. <https://doi.org/10.31661/jbpe.v0i0.2004-1107>
- [71] Wei, L., Ding, K., Hu, H. (2020). Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. *IEEE Access*, 8: 99633-99647. <https://doi.org/10.1109/ACCESS.2020.2997710>
- [72] Demir, A., Yilmaz, F., Kose, O. (2019). Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3. In 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, pp. 1-4. <https://doi.org/10.1109/TIPTEKNO47231.2019.8972045>
- [73] Codella, N.C., Nguyen, Q.B., Pankanti, S., Gutman, D.A., Helba, B., Halpern, A.C., Smith, J.R. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, 61(4/5): 5:1-5:15. <https://doi.org/10.1147/JRD.2017.2708299>
- [74] Manasa, K., Murthy, D.G.V. (2021). Skin cancer detection using VGG-16. *European Journal of Molecular & Clinical Medicine*, 8(1): 1419-1426.
- [75] Jain, M., Pulijal, S.V., Rajadhyaksha, M., Halpern, A.C., Gonzalez, S. (2018). Evaluation of bedside diagnostic accuracy, learning curve, and challenges for a novice reflectance confocal microscopy reader for skin cancer detection in vivo. *JAMA Dermatology*, 154(8): 962-965. <https://doi.org/10.1001/jamadermatol.2018.1668>
- [76] Farooq, M.A., Azhar, M.A.M., Raza, R.H. (2016). Automatic lesion detection system (ALDS) for skin cancer classification using SVM and neural classifiers. In 2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, Taiwan, pp. 301-308. <https://doi.org/10.1109/BIBE.2016.53>
- [77] Nugroho, A.A., Slamet, I., Sugiyanto. (2019). Skins cancer identification system of HAM10000 skin cancer dataset using convolutional neural network. In AIP Conference Proceedings, AIP Publishing LLC, 2202(1): 020039. <https://doi.org/10.1063/1.5141652>