



## OPEN Skin cancer classification leveraging multi-directional compact convolutional neural network ensembles and gabor wavelets

Omneya Attallah

Skin cancer (SC) is an important medical condition that necessitates prompt identification to ensure timely treatment. Although visual evaluation by dermatologists is considered the most reliable method, its efficacy is subjective and laborious. Deep learning-based computer-aided diagnostic (CAD) platforms have become valuable tools for supporting dermatologists. Nevertheless, current CAD tools frequently depend on Convolutional Neural Networks (CNNs) with huge amounts of deep layers and hyperparameters, single CNN model methodologies, large feature space, and exclusively utilise spatial image information, which restricts their effectiveness. This study presents SCaLiNG, an innovative CAD tool specifically developed to address and surpass these constraints. SCaLiNG leverages a collection of three compact CNNs and Gabor Wavelets (GW) to acquire a comprehensive feature vector consisting of spatial–textural–frequency attributes. SCaLiNG gathers a wide range of image details by breaking down these photos into multiple directional sub-bands using GW, and then learning several CNNs using those sub-bands and the original picture. SCaLiNG also combines attributes taken from various CNNs trained with the actual images and subbands derived from GW. This fusion process correspondingly improves diagnostic accuracy due to the thorough representation of attributes. Furthermore, SCaLiNG applies a feature selection approach which further enhances the model's performance by choosing the most distinguishing features. Experimental findings indicate that SCaLiNG maintains a classification accuracy of 0.9170 in categorising SC subcategories, surpassing conventional single-CNN models. The outstanding performance of SCaLiNG underlines its ability to aid dermatologists in swiftly and precisely recognising and classifying SC, thereby enhancing patient outcomes.

**Keywords** Skin cancer diagnosis, Dermoscopic imaging, Deep learning, Convolutional neural networks, Gabor wavelets, Feature fusion, Feature selection

Skin cancer (SC) is a common and serious disease caused by the uncontrollable proliferation of abnormal skin cells<sup>1</sup>. It is now a significant public health issue due to increasing rates of occurrence worldwide<sup>2</sup>. From 2008 to 2018, there was a 53% increase in the overall number of melanoma cases caused by excessive exposition to UV rays<sup>3</sup>. SC is categorised into various subtypes, with the most prevalent being basal cell carcinoma, squamous cell carcinoma, and the most severe form, melanoma<sup>4</sup>. Melanoma of the skin boasts a relatively high 5-year survival rate of 93.5%<sup>5</sup>. Identifying skin melanoma early leads to a 99.6% chance of surviving for 5 years<sup>5</sup>. Thus, timely identification and medical intervention greatly enhance the chances of survival, especially for melanoma<sup>6</sup>. Conventionally, the diagnosis of SC is mainly based on visual examination by dermatologists using dermoscopic images and occasionally requires biopsies for verification. There are still inherent drawbacks associated with manual diagnosis<sup>7</sup>. Subjective evaluations may result in inconsistencies, especially in situations involving subtle malignant characteristics or atypical manifestations because of complexities, noise, intensity variations, and

<sup>1</sup>Department of Electronics and Communications Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria 21937, Egypt. <sup>2</sup>Wearables, Biosensing, and Biosignal Processing Laboratory, Arab Academy for Science, Technology, and Maritime Transport, Alexandria 21937, Egypt. email: o.attallah@aast.edu

similarities among lesions. Besides, dermoscopic examinations encounter challenges in detecting and categorising early-stage malignancies because of their tiny dimensions and subtle differences in size, form, texture, and colour on various skin surfaces. Furthermore, the scarcity of dermatologists in certain areas could hinder visual detection and lead to delayed diagnosis<sup>8</sup>.

To address these obstacles, computer-aided diagnosis (CAD) systems are surfacing as a potentially beneficial tool for physicians. Deep learning (DL), specifically convolutional neural networks (CNN), has significantly transformed the field of medical image analysis<sup>9–11</sup>. CNNs have a remarkable ability to derive complex features from images<sup>12</sup>. DL techniques have also been extensively used to support the recognition of various diseases such as breast cancer<sup>13,14</sup>, cervical cancer<sup>15–17</sup>, lung diseases<sup>18–20</sup>, blood diseases<sup>21</sup>, heart diseases<sup>22</sup>, brain tumours<sup>23,24</sup>, and eye diseases<sup>25,26</sup>. The use of CAD systems, especially those powered by DL, offers a crucial avenue for early detection and improved diagnostic accuracy in SC<sup>27</sup>. Examining the literature revealed that although there are multiple diagnostic and classification methods for SC, there are still several gaps that need to be addressed. The factors involved are the detailed configurations, the higher complexities of the models used in specific studies, and the reduced level of diagnostic accuracy. Many studies used deep features extracted from large CNNs with many deep layers and huge parameters. Most previous CAD systems rely on separate CNNs for classification. However, combining more deep features from different CNNs with diverse structures could enhance diagnostic accuracy, as proposed by Karthik et al.,<sup>28</sup> Nagaraj and Subhashini,<sup>29</sup> Attallah<sup>30</sup>. Primarily, they used skin images as direct inputs to CNN models and relied solely on spatial information to predict the types of skin lesions. However, employing multi-resolution analysis such as Gabor Wavelets (GW) to decompose images and produce various subbands of images could boost the performance of CNNs. GW analysis can offer useful supplementary knowledge for CNNs in the context of SC diagnosis<sup>26</sup>. GW is proficient in isolating specific features from images that are localised, multi-scale, and orientation-specific<sup>31</sup>. This corresponds effectively to the visual attributes of skin lesions, in which texture, patterns, and edges at different scales are crucial diagnostic indicators<sup>32</sup>. CNNs can obtain a more detailed representation of skin lesions when using GW by breaking down the images into various texture, frequency, and orientation components (Gabor-filtered subbands). This may result in better performance, especially in situations involving subtle or intricate features of the lesion<sup>33</sup>. This is because, the texture and frequency information obtained by GW enhances the feature representation acquired by the CNN, potentially enhancing its diagnostic accuracy<sup>33</sup>.

In this study, a novel CAD based on three lightweight CNNs of distinct architectures and GW is proposed to diagnose SC called “SCaLiNG” which stands for skin cancer classification leveraging compact networks and Gabor wavelets. The proposed CAD involves breaking down input images into four directional sub-bands using GW. This approach offers multi-directional decomposition, with each sub-band producing distinct directional information in its corresponding channel. Five parallel CNNs having similar architectures are fed with these four subband photos and the original photos. Deep features are extracted for the five CNNs and combined. Each channel in a CNN generates learning specific to that direction, which can be considered independent from learning in other directions. Therefore, integrating the deep features extracted from these channels enhances the overall classification accuracy. Furthermore, SCaLiNG CAD merges deep spatial-textural-frequency features of the three lightweight CNNs benefiting from their unique architectures, thus improving performance. In addition, SCaLiNG CAD employs feature selection (FS) to select deep features that impact diagnostic accuracy.

The primary contributions and novelty of this article can be summarised as follows :

- Instead of directly using the input images to train the CNN, SCaLiNG CAD involves breaking down input images into four directional sub-bands and using them, along with the original image, as inputs for five parallel CNNs.
- Employ lightweight CNNs instead of CNNs with a huge number of parameters and deep layers
- Using a variety of CNNs with distinct structures, instead of traditional CAD models for SC diagnosis that depend on a single CNN.
- Combining deep features that are obtained from CNNs trained with different subbands of GW produces distinct directional, texture, and frequency information rather than using only spatial data, which is the case in existing CADs.
- In order to streamline and reduce the classification model complexity of the training process, a FS approach is employed to determine a more concise and significant subset of attributes from the merged deep features.
- SCaLiNG CAD simplifies the complexity of the CAD model by not requiring segmentation or enhancement operations.

This paper is structured as follows. Section “[Literature review](#)” reviews existing research on using DL for SC diagnosis. Section “[Methods and materials](#)” details the methodology employed in this study. The experimental setup is described in Section “[Experimental setting](#)”. Section “[Results](#)” presents the results of a comprehensive assessment and analysis, followed by Section “[Discussions](#)” which discusses the key findings and limitations of SCaLiNG CAD, and suggests potential areas for future research. Finally, Section “[Conclusion](#)” concludes the paper.

## Literature review

The literature on SC diagnosis shows that different DL models were used to recognise multiple subclasses of SC including melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, vascular lesion, and squamous cell carcinoma<sup>34</sup>. This section will discuss the studies that employed the DL technique to identify SC.

The authors of the article<sup>35</sup> developed a CAD for categorising 7 subtypes of SC. Pre-processing methods like contrast enhancement, colour space transformation, and brightness adjustment were used to improve the quality of pictures. Subsequently, preprocessed images were used to independently train five CNNs, which included ResNet-50, VGG-16, VGG-19, Xception, and AlexNet. An accuracy of 92.08% was achieved with the ResNet-50 CNN. The study<sup>36</sup> introduced a DL-based framework for detecting SC lesions. Data augmentation was operated to balance various groups and address data inequality. SC was categorised using InceptionV3, AlexNet, and RegNetY-320. Several combinations of hyperparameters were utilised to optimise the suggested structure. RegNetY-320 achieved the highest performance in F1 score and accuracy, achieving 88.1% and 91%, respectively. Research<sup>37</sup> developed a dependable method for categorising SC that can distinguish between various types of SC. The system used various image segmentation and classification methods. The combined segmentation paradigm incorporated the threshold approach, edge detection method, region-growing method, clustering method, U-Net model, and RP-Net model. Three CNNs, including ConvNeXtSmall, EfficientNetV2B3, and EfficientNetV2S, were utilised for classification. The accuracy and precision achieved values of 0.984 and 0.986, respectively.

The authors of reference<sup>38</sup> developed a deep residual neural network (RNN) to detect skin lesions using the wavelet transform. The suggested model utilised the discrete wavelet transform, pooling, and normalisation to eliminate extraneous information from SC photos and uncover more intricate details. Later, the RNN was employed to acquire attributes. Finally, the detection procedure was completed using the Extreme Learning Machine. The F1-Score, specificity, accuracy, precision, and attained values were 93.44%, 98.8%, 95.73%, and 95.84% respectively. The authors of reference<sup>34</sup> developed a reliable and automated method for detecting seven distinct types of skin cancer by utilising DenseNet and EfficientNet. The highest accuracy of 85.8% was attained by utilising EfficientNet.

The study<sup>39</sup> utilised various sampling techniques to address the issue of class imbalance. The authors then processed the SC images to eliminate noise. They segmented images using the encoders and decoders approach later on. They used two CNNs including ResNet-50 and DenseNet-169, separately for classification. The study<sup>40</sup> suggested the use of an effective affine transformation technique for data augmentation to enhance the SC dataset and improve SC identification. SC photos were categorised using Inception-Resnet-v2. After expanding the dataset in this study, the highest reported accuracy achieved was 95.09%. The authors of<sup>8</sup> suggested an ensemble method that exploited three CNNs, each tailored with unique dropout layers to improve feature-level learning. The combined network named DCENSnet accomplished a superior balance between bias and variance trade-off. The model surpassed the performance of cutting-edge networks on the widely used HAM10000 skin lesion dataset, attaining an accuracy of 99.53%.

In the study<sup>41</sup>, a system was suggested for multiclass SC classification. The CAD had both segmentation and classification steps. ResNet-50-based mask recurrent convolutional neural networks (Mask-R-CNN) and feature pyramid networks (FPN) were implemented during the segmentation phase. A CNN was applied for classification, achieving an accuracy of 86.5%. The study<sup>42</sup> processed the SC photographs by applying a decorrelation deformation technique and then used Mask-R-CNN. Important features were acquired and combined from two layers of DenseNet. An SVM classifier was utilised to select the optimal features, resulting in an accuracy of 88.5%. The study<sup>43</sup> created and evaluated a SC CAD system using DL. Utilising Gated Recurrent Unit Networks (GRU) and an enhanced Orca predation algorithm optimised the effectiveness of the proposed system. The authors also used the HAM10000 skin lesion dataset. On the other hand, the study<sup>44</sup> introduced an automated framework in which normalisation was applied to dermoscopic images to reduce the effects of noise, outliers, and pixel fluctuations. Malignant areas in the preprocessed images were identified using the Mask-RCNN model. The Mask-RCNN model provided segmentation by precisely outlining the boundary of objects. Discriminatory feature vectors were obtained from the segmented malignant regions using three pretrained CNN models: ResNeXt101, Xception, and InceptionV3. A modified Gated GRU model was used for SC classification, while the study<sup>45</sup> employed the sand cat swarm optimisation with ResNet50 (SCSO-ResNet50) technique to distinguish deep hidden features from known attributes, providing precise predictions. They use an enhanced harmony search method to optimise attributes and decrease data complexity. Ensemble classifiers such as Naive Bayes, random forest, k-nearest neighbour (k-NN), support vector machine (SVM), and linear regression were employed in the early detection of SC.

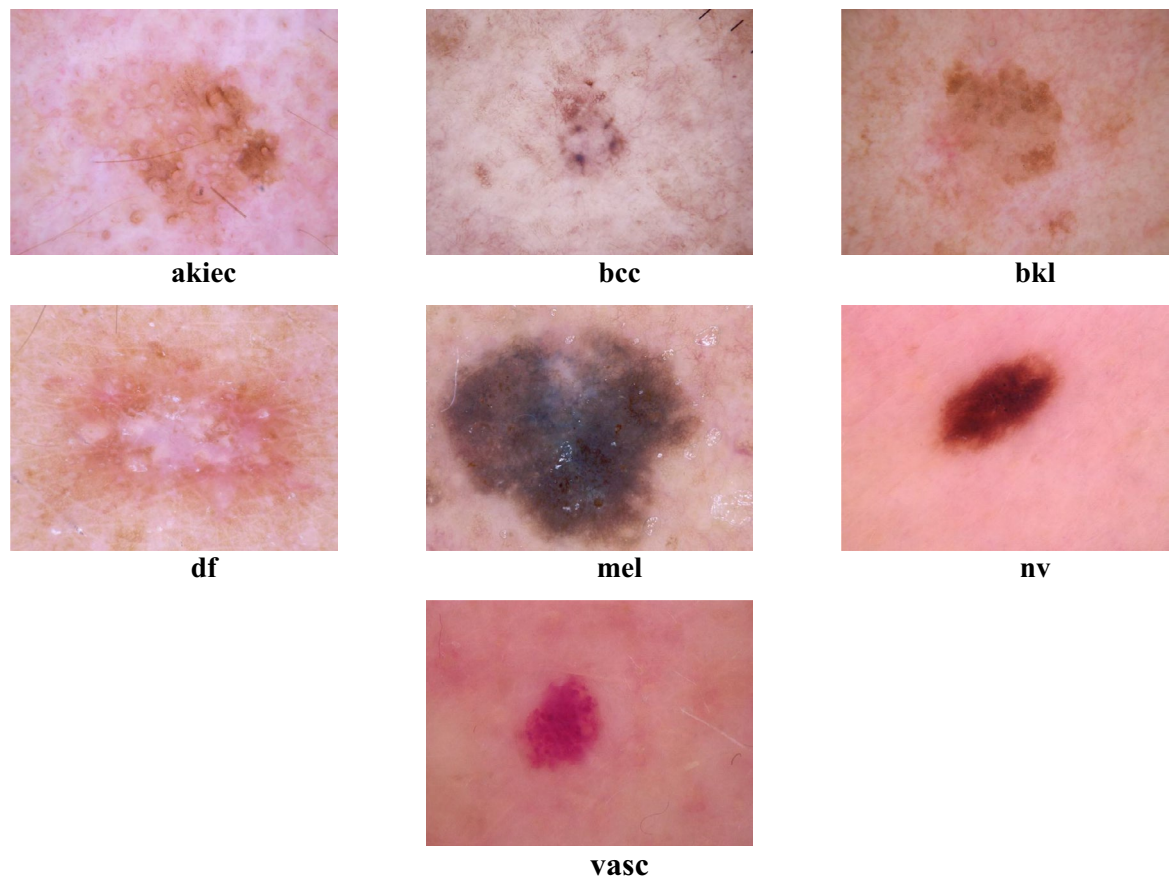
## Methods and materials

### HAM10000 Skin cancer dataset

The HAM10000 data is a benchmark dataset in the fields of SC image analysis and DL research. The dataset includes more than ten thousand dermoscopic pictures depicting seven distinct forms of SC subcategories. The dataset's value for training and evaluating DL models tasked with critical differential diagnoses lies in its combination of malignant and benign lesions. The images in HAM10000 are carefully labelled by skilled dermatologists, ensuring a reliable reference point for training and assessing models. The dataset shows diversity in patient age, sex, skin type, and lesion location, which promotes the creation of models that can handle real-world variations effectively. The HAM10000 dataset can be accessed for free on the ISIC archive website<sup>46</sup>. The database contains 142 cases of vascular lesions (vasc), 6705 cases of melanocytic nevus (nv), 1113 cases of melanoma (mel), 1099 cases of benign keratosis (bkl), 115 cases of cutaneous fibromas (df), 327 cases of actinic keratosis (akiec), and 514 cases of basal cell carcinoma (bcc). Figure 1 displays numerous images acquired from the HAM10000 database.

### Suggested SCaLiNG CAD

The SCaLiNG CAD framework consists of multiple successive phases: GW image generation and preparation, retraining pre-trained lightweight DL models, feature extraction and integration, feature selection, and classification of SC categories. The first step involves the application of GW analysis to dermoscopic pictures. This



**Figure 1.** Examples of the HAM10000 dataset.

process generates sub-images that have various orientations and scales, which capture the textural-frequency information. Afterward, both the images that are produced and the original photos are resized and augmented. The previously processed pictures are subsequently inputted into numerous simultaneous versions of pre-trained, compact CNNs of the same architecture. The deep attributes obtained from such CNNs are then combined. This process is repeated for the three CNNs employed in the ScaLiNG tool including ResNet-18, MobileNet, and ShuffleNet. Next, the resultant deep spatial-textural-frequency features, generated by the ResNet-18, ShuffleNet, and MobileNet structures, are merged. Subsequently, a FS procedure is utilised to determine the most distinctive attributes for classification. Ultimately, machine learning classifiers are employed to classify SC by considering the chosen variables. The stages of ScaLiNG CAD are briefly explained in Fig. 2.

#### *Gabor wavelets image generation and preparation*

The present research leverages GWs to break down SC photos into various frequencies and orientations. GWs produce a collection of transformed photos by identifying important visual characteristics at different sizes and angles. These pictures are to be used to feed pre-trained CNNs to acquire features and classify them. GWs provide a clear benefit compared to conventional image transformation algorithms by simultaneously acquiring textural and frequency domain characteristics. This detailed depiction enables CNNs to acquire a deeper and all-encompassing comprehension of the content within images, which might eventually result in better performance in diagnostic examinations. The GW images are generated using the following formulas:

The GW image is defined as:

$$g(x, y) = e^{[x-x_o] + [y-y_o]/\alpha^2} e^{-i[u_o(x-x_o) + v_o(y-y_o)]/\beta^2} \quad (1)$$

where, the location in a photo is presented as  $(x_o, y_o)$  and the spatial frequency is referred to as  $\omega_o = \sqrt{u_o^2 + v_o^2}$ . The orientation is defined as  $\theta = \tan^{-1}\left(\frac{v_o}{u_o}\right)$ .

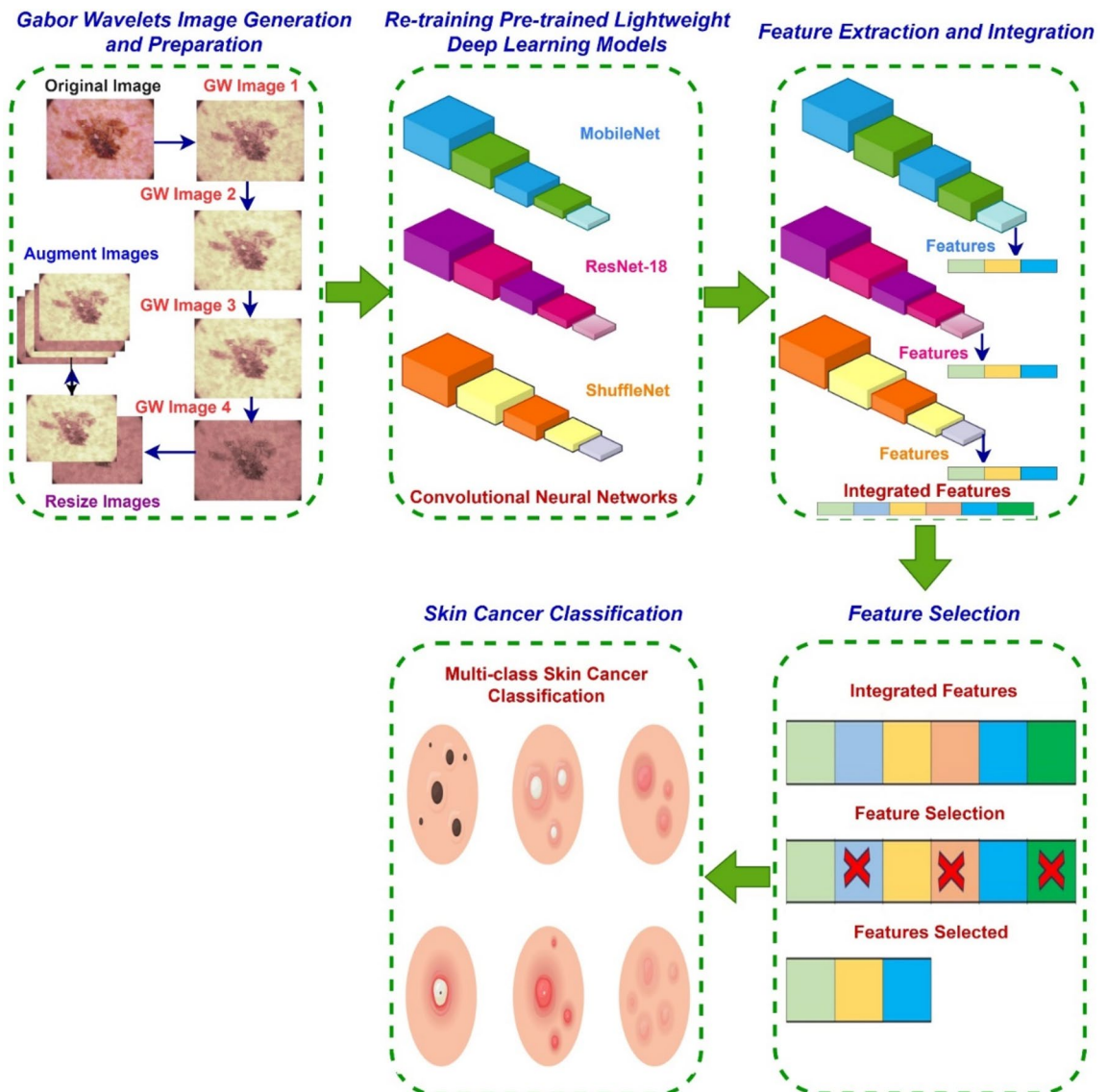
The GW image Fourier transform is given by

$$G(u, v) = e^{[u-u_o] + [v-v_o]/\alpha^2} e^{-i[x_o(u_o-v_o) + y_o(v-v_o)]/\beta^2} \quad (2)$$

The GW function of the real portion is found in the center of the photo  $x_o, y_o = (0, 0)$ .

GW analysis utilises two scales  $x, y$  (0.176, 0.25) and two orientations  $\theta$  (0, 90) in this study<sup>33,47</sup> producing four subband images named: GW 1, GW 2, GW 3, and GW 4. Samples of the images generated after the GW analysis are shown in Fig. 3.





**Figure 2.** The five stages of SCaLiNG.

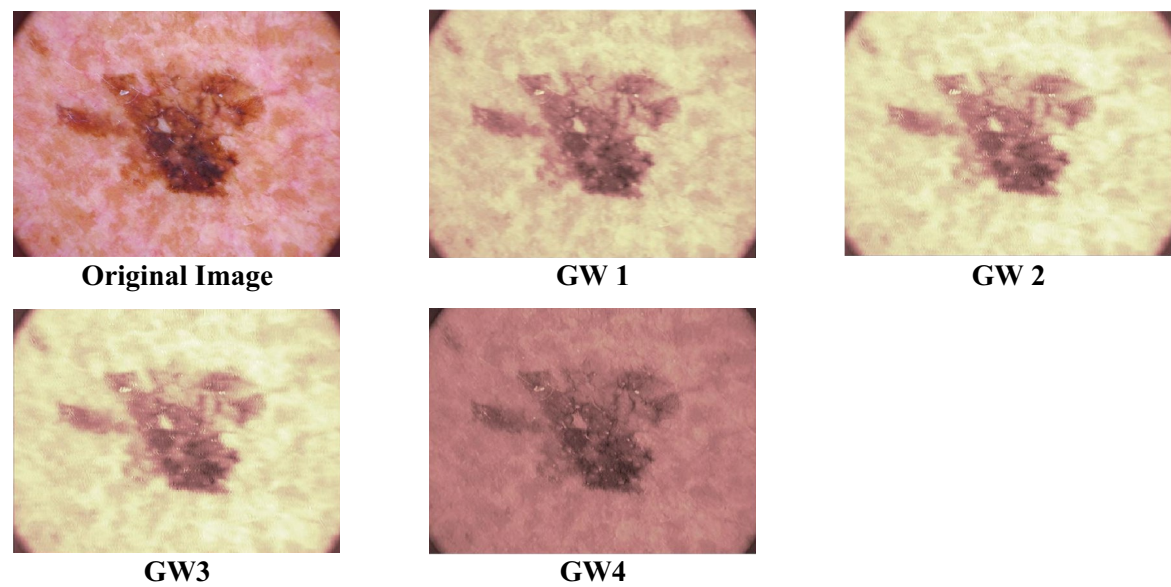
After generating the GW images, both the original and GW-generated photos undergo a resizing process, which involves trimming the lengths of such photographs to align with the input layer length of each CNN. The new dimensions for these images are  $224 \times 224 \times 3$  for ResNet-18, ShuffleNet, and MobileNet101. Next, the dataset is divided into sets for training and testing with a proportion of 70%-30%. Subsequently, the data used for training is augmented by applying different augmentation approaches to raise the number of training images, preventing overfitting and improving training performance. Table 1 illustrates the augmentation approaches in detail.

#### Retraining pre-trained lightweight DL models

Transfer learning (TL) is used in the current stage to develop and train three efficient pre-trained CNNs that had already been trained on the ImageNet dataset<sup>48</sup>. Specifically, ResNet-18, MobileNet, and ShuffleNet are the CNNs that are being discussed. To match the number of SC categories that are present in the HAM10000 dataset, the total number of fully connected layers of each CNN has been adjusted accordingly to seven. After that, a number of hyperparameters are modified in order to achieve optimal performance; additional information regarding this process can be found in the section titled "Experimental Setting." Following that, the process of retraining the three CNNs begins. While each of the three CNNs is being re-trained independently, individual versions of the created GW images as well as the original images are being used to re-train these CNNs. This indicates that each DL model is trained in parallel with images generated by four GW subbands as well as the original dermoscopic photos.

#### Feature extraction and integration

TL is utilised to extract deep features from the very last fully connected layers of each CNN once the retraining process has been finally completed. The feature vector extracted from each version of a CNN has a length



**Figure 3.** A sample of the four subband images generated after the GW analysis.

Augmentation technique	Range
Flipping in horizontal and vertical directions	– 45–45
Scaling	0.5–1.5
Shearing perpendicularly	– 50–50
Rotating in horizontal and vertical directions	– 60–60

**Table 1.** Augmentation techniques used and their range values.

of seven. Immediately after the process of feature extraction comes the procedure of feature integration, which is accomplished in two stages. In the beginning, the deep attributes of each CNN that belong to the four GW images are merged with the deep attributes of the original image. To take advantage of each CNN architecture, the integrated features of the CNNs are then incorporated. In this step, an investigation is carried out to determine whether or not combining deep features from multiple CNNs can improve diagnostic accuracy.

*Feature selection*

FS is an essential method in machine learning that determines the most pertinent features (variables) from a dataset containing a large number of them. The process aims to remove unnecessary and repetitive information, concentrating on features that have a substantial impact on the analysis or model construction. FS is crucial in the medical field due to the diverse range of data, including different measurements, patient histories, and other intricate factors<sup>49,50</sup>. It can be used to enhance model performance, removing irrelevant features allows the model to prioritise the most informative ones, resulting in improved accuracy and generalisability of predictions<sup>51,52</sup>. FS is also able to decrease computational complexity because having fewer features speeds up the training process of machine learning models and reduces resource usage, which is advantageous for analysing large medical datasets<sup>53</sup>. In addition, FS is capable of improving interpretability by choosing the most significant features that can offer valuable insights into the fundamental biological or clinical connections, assisting medical professionals in comprehending the factors that impact health outcomes.

In this study, the minimum Redundancy Maximum Relevance (mRMR) FS approach is used to select significant features. The mRMR algorithm is a popular method to select features in machine learning tasks. The task at hand involves choosing a subset of features that are not only informative for the target variable, but also have minimal redundancy between them<sup>54</sup>. The mRMR approach accomplishes this by utilising a greedy search method that systematically chooses features with the greatest trade-off between relevance and redundancy. This trade-off can be expressed in various formulations, such as maximising the disparity between relevancy and redundancy or maximising their ratio<sup>55</sup>. The main objective of this study is to ascertain the maximum level of interdependence among a set of attributes  $X$  and  $Y$  and the class label  $C$ , using mutual information (MI) represented as  $I$ . The MI between two variables can be determined using formula (3) when the marginal probabilities ( $p(x)$  and  $p(y)$ ) and the joint probability ( $p(x, y)$ ) for both variables are known<sup>56</sup>.

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \tag{3}$$

Implementing the maximum dependency criterion can be a complex task in multivariate domains. In particular, there is usually a lack of sufficient instances, and calculating the multivariate frequency often requires computationally expensive calculations. A different approach is to determine the criterion that has the greatest degree of significance. The idea of maximum relevance involves the identification of the features that satisfy formula (4). Given an example feature set  $F = \{f_1, f_2, \dots, f_n\}$

$$\max D(F, C); D = \frac{1}{|F|} \sum_{f_i \in F} I(f_i, C) \tag{4}$$

Selecting variables according to the maximum relevance metric may lead to a considerable amount of duplication. Therefore, it is imperative to incorporate the requirement of minimal redundancy as suggested by the formula (5) mentioned in reference<sup>57</sup>.

$$\min R(F); R = \frac{1}{|F|} \sum_{f_i, f_j \in F} I(f_i, f_j) \tag{5}$$

The combination and improvement of the two conditions  $D$  and  $R$  lead to the approach of mRMR. Pragmatically, a greedy approach may be employed, where  $S$  denotes the set of selected features:

$$\max_{X_i \notin S} \left[ I(X_i; C) - \frac{1}{|S|} \sum_{X_j \in S} I(X_i, X_j) \right] \tag{6}$$

*Skin cancer classification*

During the final stage of SCaLinG CAD, skin lesions are classified into one of the seven different categories presented. Specifically, for the purpose of multiclass classification, three SVM classifiers with distinct kernels, namely linear, quadratic, and Gaussian, have been developed, respectively. There are three settings in which the classification procedure is carried out. SVM classifiers are initially trained independently using textural-frequency features obtained from distinct CNNs learned with GW subbands image sets. After that, the SVM classifiers are learned with the aggregated textural-frequency features that are taken from the four GW sub-band images. In the second setting, the spatial features provided by each CNN are combined with the integrated textural frequency features, and the resulting combination is subsequently used as input for the three SVMs. The third setup involves the merging of the combined features of the three CNNs, followed by the utilisation of an FS method to select the most significant features that have an impact on accuracy. A further issue that is investigated in this setting is whether or not the combination of features from multiple CNNs can improve the accuracy of the classification. In addition, it investigates whether or not FS is capable of improving performance.

**Experimental setting**

The following section details the hyper-parameter values that have been adjusted and the evaluation metrics used to assess the effectiveness of SCaLinG CAD. Various hyperparameter values have been adjusted for each CNN. Table 2 shows the hyperparameters and the values that correspond to them. All other configurations are kept at their default state. The optimisation method employed to train the CNNs is SGDM, short for stochastic gradient descent with momentum. The codes for the ScaLinG CAD are available in the supplementary material.

Multiple performance metrics are employed for assessing the efficiency of machine learning classifiers. Among them, the accuracy which is a straightforward metric that shows the percentage of correctly classified samples out of the total samples. Precision quantifies the accuracy of positive labels assigned by the model, while recall (or sensitivity) evaluates the accuracy of correctly identified actual positives. Specificity is a metric that quantifies the model's capacity to accurately detect negative cases. It essentially signifies the fraction of actual negative samples that the model correctly identified as negative. The F1-score strikes a balance between precision and recall. The receiving characteristic curve (ROC) and the area under the ROC curve (AUC) are crucial for imbalanced datasets as they evaluate the model's capacity to differentiate between classes at different thresholds. The metrics offer valuable insights into the capabilities and limitations of a machine learning classifier, as well as its suitability for particular uses.

Before calculating performance metrics in machine learning, it is essential to define key indicators: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP is the properly determined positive cases, while TN is the accurately categorised negative cases. FP is cases that are mistakenly identified as

Hyper-parameter	Value
Mini batch	10
Epochs	40
Learning rate	0.001
Validation frequency	701

**Table 2.** The CNNs’ hyper-parameters altered and their values.

positive, while FN is cases that are inaccurately classified as negative. The metrics for assessment are computed using the subsequent equations:

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{10}$$

$$F1 - Score = \frac{2 \times TP}{(2 \times TP) + FP + FN} \tag{11}$$

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \tag{12}$$

Results

The section begins with presenting the results of setting I in Section "Setting I classification results" where SVM classifiers are trained independently using textural-frequency features that are extracted by individual CNNs from sets of textural-frequency GW images. In addition, the results of the SVM classifiers when trained using the combined textural frequency features extracted from the four subband images of the GW are illustrated. Afterward, Section "Setting II classification results" illustrates the results of the second setting where the spatial deep features acquired using each DL model are combined with the integrated textural-frequency features obtained by GW and used as inputs for the three SVMs. Finally, in Section "Setting III classification results" the results of setting III are shown where both combined features of all the three CNNs are further integrated, and the mRMR FS approach is employed to select the most significant features that impact accuracy. This setting also investigates the potential enhancement of classification accuracy by integrating features from multiple CNNs. It also investigates whether FS can improve performance.

Setting I classification results

The results of the SVM classification models trained with deep features obtained using each CNN architecture fed with individual GW subband images of the four subbands generated after the GW analysis and the combined deep features of the four GW subband images are shown in Tables 3, 4, 5 for ResNet-18, ShuffleNet, and MobileNet deep features, respectively. It can be seen from Table 3 that the Linear SVM kernel trained with deep ResNet-18 features exhibited the lowest accuracy among all GW subbands and the combined subbands except for the GW1 and GW3 subbands. The accuracy varies between 71.9% for the GW3 subband and 76.4% for all GW subbands. Furthermore, the quadratic kernel accuracy varies between 69.3% for the GW1 subband and 78.6% for all GW subbands. Furthermore, the Gaussian kernel verified the uppermost accuracy in all scenarios. Accuracy varies between 75.2% for the GW3 subband and 78.4% for all GW subbands.

According to Table 4 which exhibits the results using deep features of ShuffleNet, the SVM classifier with a linear kernel had the lowest accuracy in all scenarios. The accuracy varies from 72.0% for the GW3 subband to 80.0% for all GW subbands. For the quadratic SVM, the accuracy ranges between 73.0% for the GW3 subband and 82.2% for all GW subbands. Furthermore, the Gaussian kernel demonstrates a higher level of accuracy in all scenarios than the linear SVM classifier. The accuracy fluctuates between 74.6% for the GW3 subband and 82.2% for all GW subbands. The accuracy of the Gaussian SVM classifier outperforms the other two classifiers. The accuracy of the Gaussian kernel decreased marginally for the GW3 subband (74.6%) in comparison to the

Deep features	Linear	Quadratic	Gaussian
GW1 Subband	73.3	69.3	75.5
GW2 Subband	73.3	74.6	75.7
GW3 Subband	71.9	71.3	75.2
GW4 Subband	72.9	71.0	75.9
All GW Subbands	76.4	78.6	78.4

**Table 3.** Classification accuracy (%) achieved using deep features obtained using ResNet-18 trained with each individual GW subband images.



Deep features	Linear	Quadratic	Gaussian
GW1 Subband	76.3	77.7	78.9
GW2 Subband	75.6	77.5	78.5
GW3 Subband	72.0	73.0	74.6
GW4 Subband	74.8	75.8	78.6
All GW Subbands	80.0	82.2	82.2

**Table 4.** Classification accuracy (%) achieved using deep features obtained using ShuffleNet trained with each individual GW subband images.

Deep Features	Linear	Quadratic	Gaussian
GW1 Subband	79.2	81.6	81.6
GW2 Subband	79.0	80.0	81.6
GW3 Subband	79.1	81.0	81.5
GW4 Subband	79.3	81.1	81.5
All GW Subbands	84.0	84.4	84.7

**Table 5.** Classification accuracy (%) achieved using deep features obtained using MobileNet trained with each individual GW subband images.

other subbands. Across all three kernels, the accuracy of the combined GW subbands is the highest compared to that of each individual subband. This suggests that using all GW subbands together provided the most valuable features for the SVM classifiers.

In accordance with Table 5 which shows the results using deep features of MobileNet, the SVM classifier with the linear kernel has the least accuracy among all GW subbands and the combined subbands. The accuracy differs from 79.2% for the GW1 subband to 84.0% for all GW subbands. In addition, the quadratic kernel outperforms the linear kernel in all scenarios. The accuracy fluctuates from 80.0% for the GW2 subband to 84.4% for all GW subbands. In addition, the Gaussian kernel demonstrates a superior level of accuracy in all respects. The accuracy diverges from 81.5% for GW3 Subband to 84.7% for all GW subbands.

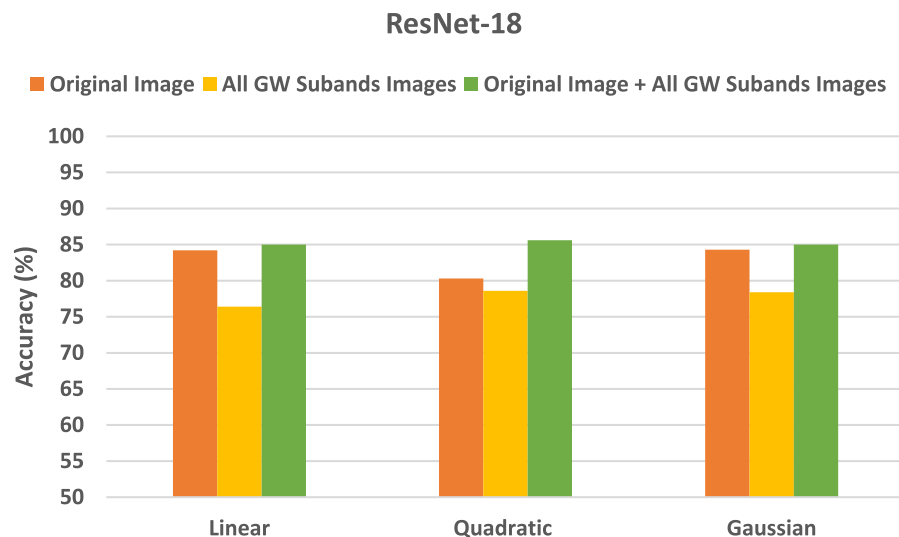
In summary, these findings of Tables 3, 4, 5 indicate that the combined subbands of all GW obtained the highest accuracy compared to each individual subband for all three kernels for deep features ResNet-18, ShuffleNet, and MobileNet. This implies that utilising all of four GW subbands collectively yielded the most informative attributes for the SVM classification model.

## Setting II classification results

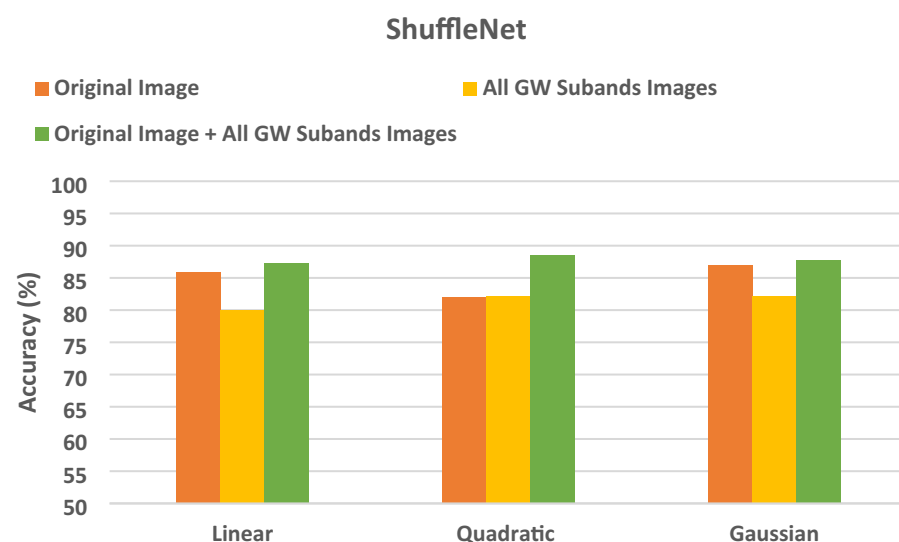
The section demonstrates the results of the SVM classifiers when trained with deep features extracted from every DL model learned individually with deep features of the original images, then the deep features of all GW subbands, and finally the deep of both the integrated deep features of all GW subbands fused with deep features of the original images. These results are shown in Figs. 4, 5, 6. Figure 4 reveals that SVM classifiers with different kernels achieved different classification accuracies when trained on deep features extracted from ResNet-18 CNN using various strategies. The Gaussian kernel achieved the highest accuracy in most configurations, except for the original image and the combined GW features, where the quadratic kernel performed better (85.6% and 78.6% for the quadratic compared to 85.0% and 78.4% for the Gaussian kernel, respectively). This indicates that the data displays non-linear relationships, which the non-linear kernels were able to capture more efficiently. Remarkably, the integration of the deep features of the original images with those from all GW subbands yielded a substantial enhancement in the outcomes, surpassing the utilisation of solely the combined GW subbands or the original image alone. Integrating textural-frequency information from GW images with spatial data from the original image could enhance performance.

According to the results in Fig. 5, SVM classifiers have different classification accuracies across various kernels trained on deep features extracted from ShuffleNet CNN. The quadratic and Gaussian kernels consistently outperformed the linear kernel in terms of accuracy across all scenarios. This implies that the data displays non-linear patterns, which the non-linear kernels were able to capture more efficiently. The Gaussian kernel demonstrated superior accuracy in all configurations, except for the integration of four GW subbands combined with the original image (87.7%), where the quadratic kernel (88.5%) exhibited a slightly higher performance. Significantly, the integration of the deep features of the original pictures with those from all GW subbands led to a substantial enhancement in the outcomes, surpassing the utilisation of solely the combined GW subbands or original images alone. This suggests that the combination of spatial and textural-frequency knowledge is more effective than relying solely on spatial data.

The results of Fig. 6 indicate that SVM classifiers with different kernels accomplish varying classification accuracies when learned on deep features obtained from Mobile CNN using different strategies. In all scenarios, the performance of non-linear kernels (quadratic and Gaussian) is superior to that of the Linear kernel. The Gaussian



**Figure 4.** Classification accuracies (%) for the SVM classifiers trained with deep features taken from ResNet-18 CNN trained separately using the original image's deep features, followed by the all-GW subbands' deep features, and lastly the combined deep features of all GW subbands fused with the original image's deep features.



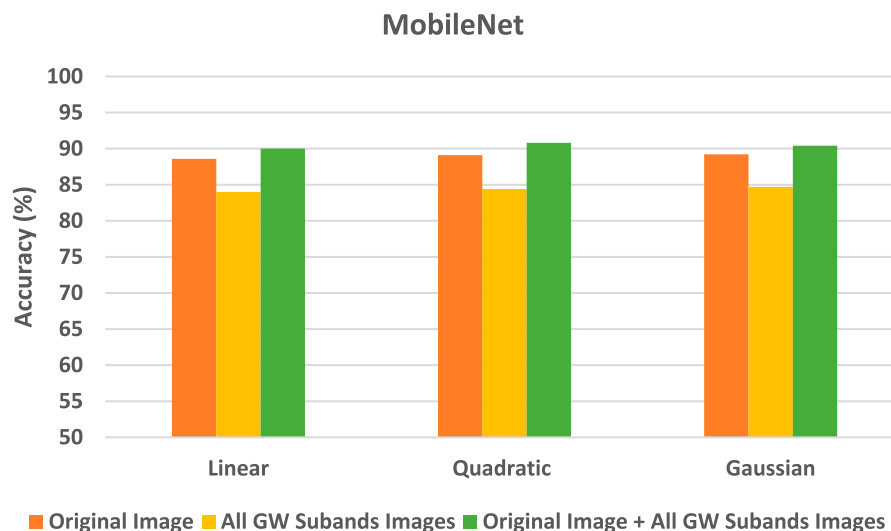
**Figure 5.** Classification accuracies (%) for the SVM classifiers trained with deep features taken from ShuffleNet CNN trained separately using the original image's deep features, followed by the all-GW subbands' deep features, and lastly the combined deep features of all GW subbands fused with the original image's deep features.

kernel demonstrates superior accuracy across all configurations, except for the original images combined with textural-frequency information of the four GW subbands, where the quadratic kernel exhibited a slightly higher performance (90.8% compared to 90.4%). This indicates that the data displays non-linear relationships, which were better captured by the Gaussian kernel. Outstandingly, when combining the deep features of the original image with those from all GW subbands, the outcomes are enhanced compared to using only the combined GW subbands or the original images. This demonstrates that the integration of spatial and textural-frequency data enhances performance.

d lastly the combined deep features of all GW subbands fused with the original image's deep features.

### Setting III classification results

This section illustrates the results of the mRMR FS technique applied to the combined deep features of the three CNNs, including those obtained with the four GW subbands and the original images. An ablation study is shown in Fig. 7 which represents the variation in classification accuracy with increasing number of variables. Figure 7 shows that the classification accuracy improves while increasing the number of features. For the linear kernel, the peak accuracy achieved is 91.0% with only 70 features. Whereas for the quadratic kernel, the maximum

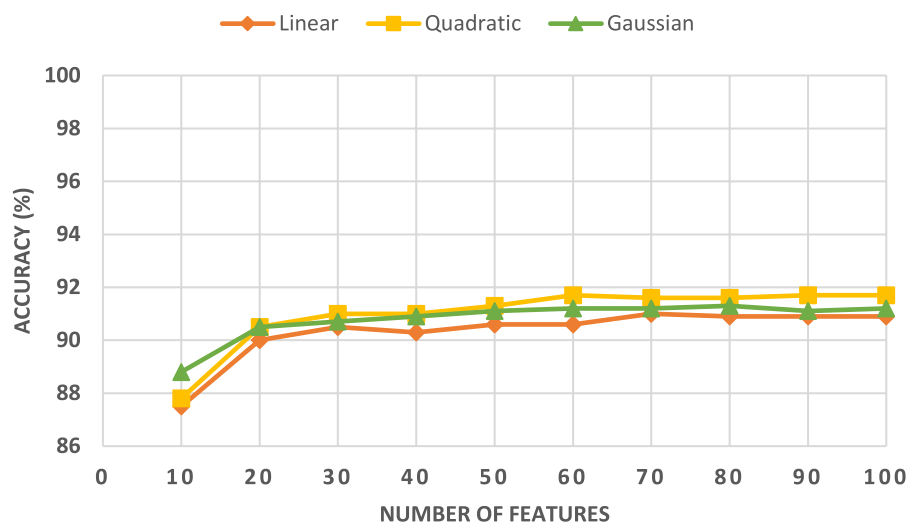


**Figure 6.** Classification accuracies (%) for the SVM classifiers trained with deep features taken from MobileNet CNN trained separately using the original image's deep features, followed by the all-GW subbands' deep features, and.

accuracy of 91.7% is accomplished with just 60 features. In contrast, the greatest accuracy for the Gaussian kernel is attained with 80 features. These results provide evidence that the mRMR FS is capable of selecting significant features while achieving better performance.

Additional evaluation metrics are employed for evaluating the efficacy of SCAliNG in achieving the highest level of accuracy after FS for each classifier. The evaluation metrics include precision, specificity, F1-score, MCC, and sensitivity. Table 6 describes these metrics. The linear SVM model achieved a precision of 0.9096, an F1-score of 0.9096, a specificity of 0.9849, an MCC of 0.8946, and a sensitivity of 0.9096. The quadratic SVM classifier yielded a precision of 0.9170, an F1 score of 0.9170, a specificity of 0.9862, an MCC of 0.9032, and a sensitivity of 0.9170. The Gaussian SVM classifier obtained a precision of 0.9126, an F1-score of 0.9126, a specificity of 0.9854, an MCC of 0.8981, and a sensitivity of 0.9126. In general, all three classifiers demonstrated outstanding precision, F1-score, specificity, and sensitivity, with all values exceeding 0.9. This indicates that the mRMR FS process successfully identified a subset of features that provided valuable information for classification purposes.

The confusion matrices for the three SVM classifiers, which were constructed using the selected deep features of the mRMR FS procedure, are displayed in Fig. 8. The diagram demonstrates that the linear, quadratic, and Gaussian SVM models of the SCAliNG system can accurately classify all categories of the HAM10000 dataset. For the linear, quadratic, and Gaussian SVM classifiers, a sensitivity of (76.6%, 82.9%, 78.3%) for the akiec



**Figure 7.** FS classification accuracy (%) of the three SVMs fed with the combined deep features of the three CNNs, including those obtained with the four GW subbands and the original images versus the number of features.

Classifier	Precision	F1-score	Specificity	MCC	Sensitivity
Linear	0.9096	0.9096	0.9849	0.8946	0.9096
Quadratic	0.9170	0.9170	0.9862	0.9032	0.9170
Gaussian	0.9126	0.9126	0.9854	0.8981	0.9126

**Table 6.** Evaluation measures achieved using the MRMR FS approach applied to the combined deep features of the three CNNs including those obtained with the four GW subbands and the original images.

class, (90.3%,90.9%,89.9%) for the bcc class, (83.0%, 83.8%,82.6%) for the bkl class, (80.9%, 87.0%, 83.5%) for the df class, (63.9%,70.8%, 67.8%) for the mel class, (97.6%, 97.0%,97.4%) for the nv class, and lastly (89.4%, 93.7%,93%) for the vasc class.

The ROC curves for the Linear, Quadratic, and Cubic SVM classifiers are displayed in Fig. 9. These classifiers are applied to the HAM10000 dataset that underwent FS using the mRMR method. The ROC curves plot the true positive rate (sensitivity) on the y-axis and the false positive rate (1-specificity) on the x-axis. Each curve represents a positive class. The SCaLiNG model demonstrates excellent performance across all categories, as evidenced by the ROC curves. All SVM classifiers achieved an AUC value higher than 0.95, which indicates that the SCaLiNG model is capable of accurately distinguishing between various subgroups within SC.

Discussions

The research article suggests a CAD called "SCaLiNG" for diagnosing SC, which is based on numerous light-weight CNNs and GW. SCaLiNG CAD analyses input images into four directional subbands using a GW-based approach. Each subband provides unique deep features that could be coupled to improve overall performance. Five parallel CNNs of the same topology are used to process the four sub-band images and the original images. Deep features are obtained from every CNN and then merged. Next, those deep features acquired from ResNet-18, ShuffleNet, and MobileNet are integrated. Afterwards, the mRMR FS procedure is implemented to identify the most influential attributes that impact the diagnostic performance. Ultimately, machine learning classification models were used to achieve the task of classification.

The procedure for classification is carried out in three distinct settings. At first, SVM classifiers are trained independently using textural-frequency features that are extracted by individual CNNs from sets of GW images. Afterwards, the SVM classifiers are fed with the combined textural-frequency features extracted from the four GW sub-band images. In the subsequent setting, the three SVMs receive input consisting of the combined spatial attributes of every CNN and the integrated textural-frequency features. The third setup combines the features of the three CNNs and employs the mRMR FS method to identify the most critical features that impact accuracy. This setting also investigates the potential enhancement of classification accuracy by combining features from different CNNs. It also examines whether FS has the ability to improve performance. The maximum accuracy achieved in each set-up is demonstrated in Fig. 10.

Figure 10 shows that the integration of both textural–frequency and spatial information (setting II) is superior to relying solely on textual frequency information, (setting I) or spatial information as proven in Figs. 4, 5, 6. This is clear in Fig. 10, since the classification accuracy in setting II (90.8%) is much higher than that reached in setting I (84.7%). Furthermore, the greatest accuracy achieved in setting III is superior to that obtained in settings I and II, which proves that merging deep features from several CNNs of different topologies is capable of enhancing performance. In addition, the increase in performance demonstrates that FS is capable of improving performance while lowering the number of features thus reducing training complexity.

Comparative evaluation

The outcomes of SCaLiNG CAD are contrasted with comparable studies that developed CAD systems for SC classification utilising the HAM10000 dataset. The superiority of SCaLiNG CAD over various prior CAD systems is demonstrated in Table 7. The accuracy, sensitivity, precision, and F1-score of SCaLiNG are all 0.9170, which is evident to be higher than those observed in previous research, with the exception of the study<sup>42</sup>, which achieved slightly higher accuracy. Although SCaLiNG CAD demonstrated the best overall efficiency, it is worth mentioning that the CAD system<sup>42</sup> displayed the greatest accuracy of 0.9208. This can be attributed to the use of preprocessing and presegmentation steps, which adds complexity to the model. The exploitation of advanced preprocessing methods and the robust segmentation powers of Mask-R-CNN enable accurate identification of tumour borders, which, in turn, provides useful data for the following feature extraction and classification steps but at the cost of complexity. The reason for the competing ability of SCaLiNG is that it uses both textural-frequency information acquired through the GW method and spatial information extracted from the original images to carry out the diagnosis. In addition, it employs combined deep features from multiple CNNs with diverse architectures. Furthermore, it uses the mRMR FS technique to reduce the dimensions of the feature space, select the most superior features, and eliminate unnecessary attributes.

Skin-CAD limitations and forthcoming directions

It is necessary to investigate the limitations of the suggested SCaLiNG CAD for SC diagnosis. One main issue revolves around the possible influence of an unequal distribution of classes in the training dataset. The lack of balance in the data could result in a shortage of training data for certain subclasses of SC, which may undermine the effectiveness of the classification process. The presence of class imbalance may have a substantial impact on



True Class	akiec	78.6%	4.9%	8.6%		4.9%	3.1%		78.6%	21.4%
	bcc	0.8%	90.3%	2.5%	0.2%	2.3%	3.9%		90.3%	9.7%
	bkl	1.0%	1.0%	83.0%		4.7%	10.3%		83.0%	17.0%
	df	3.5%	0.9%	1.7%	80.9%	4.3%	8.7%		80.9%	19.1%
	mel	0.4%	0.2%	4.8%	0.1%	63.9%	30.5%	0.1%	63.9%	36.1%
	nv	0.1%	0.2%	0.8%		1.3%	97.6%	0.0%	97.6%	2.4%
	vasc		2.1%	0.7%		4.2%	3.5%	89.4%	89.4%	10.6%
		akiec	bcc	bkl	df	mel	nv	vasc	TPR	FNR
Predicted Class										

(a) Linear SVM

True Class	akiec	82.9%	4.0%	6.7%		3.4%	2.8%	0.3%	82.9%	17.1%
	bcc	1.6%	90.9%	2.3%	0.2%	1.2%	3.5%	0.4%	90.9%	9.1%
	bkl	1.6%	0.5%	83.8%		5.0%	9.0%		83.8%	16.2%
	df	3.5%	0.9%	1.7%	87.0%	0.9%	6.1%		87.0%	13.0%
	mel	0.5%	0.4%	3.8%		70.8%	24.4%	0.1%	70.8%	29.2%
	nv	0.0%	0.2%	0.9%		1.8%	97.0%	0.0%	97.0%	3.0%
	vasc		1.4%	0.7%		1.4%	2.8%	93.7%	93.7%	6.3%
		akiec	bcc	bkl	df	mel	nv	vasc	TPR	FNR
Predicted Class										

(b) Quadratic SVM

**Figure 8.** Confusion matrices for the three SVMs fed with the selected features of the mRMR FS approach (a) Linear SVM, (b) Quadratic SVM, and (c) Gaussian SVM.

the effectiveness of machine learning algorithms, possibly resulting in predictions that are skewed toward the majority class. Within the framework of SCaLiNG, this may result in an excessive focus on categorising the most forms of skin cancer with a larger number of samples, while not performing as well in identifying subtypes, with a lower number of images. In order to address the aforementioned issue, multiple techniques are implemented throughout the SCaLiNG system. Initially, data augmentation approaches are employed to artificially increase the number of images. Furthermore, employing numerous CNN architectures in an ensemble manner may alleviate the effects of class imbalance by offering varied feature representations and possibly increasing the resilience of the model. In addition, the inclusion of feature selection is intended to determine the most distinguishing attributes, which can assist in mitigating the impact of imbalanced class variations on the classification procedure. Furthermore, the effectiveness of the SCaLiNG CAD is examined by employing measures such as precision, recall, and F1-score, in order to offer a more thorough evaluation of its efficacy on datasets with imbalanced

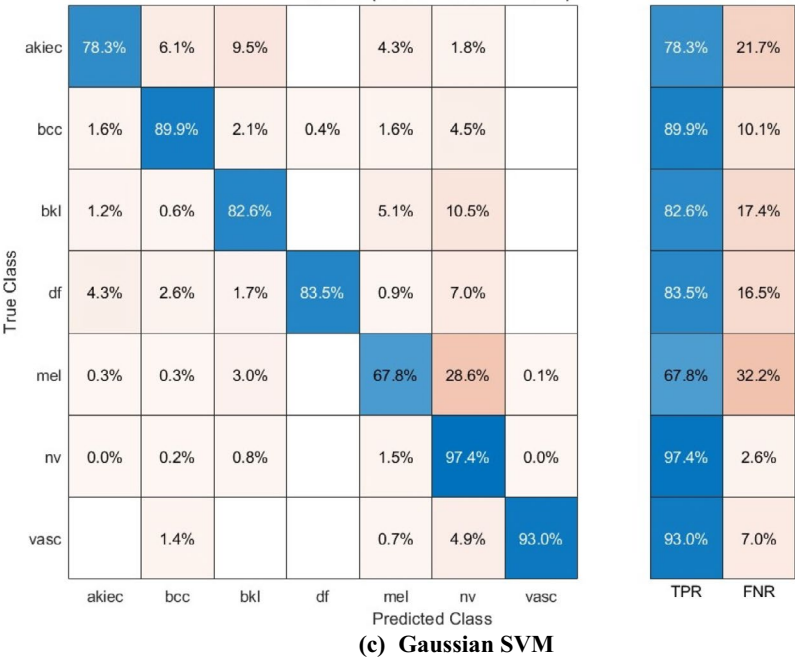


Figure 8. (continued)

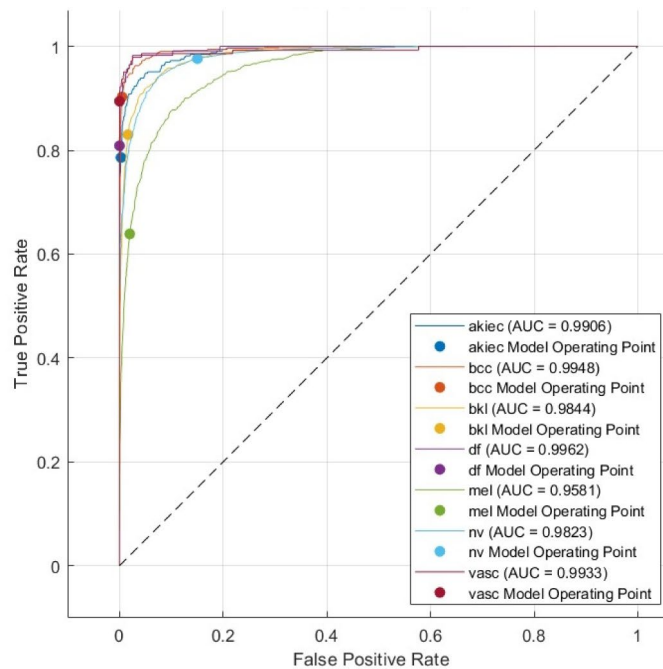
data. Although the above steps are put in place to tackle the issue of class imbalance, ScaLiNG shows varying performance across different SC classes. For instance, ScaLiNG exhibits higher performance for nv, bcc, and vasc in comparison to other SC categories. Therefore, it would be beneficial to conduct additional research on advanced methods, including cost-sensitive learning or class weighting, to improve the system's effectiveness on imbalanced datasets. Future research will investigate other sampling techniques to address class imbalance issues.

Though this study specifically examined the identification of different types of SC, the basic concepts and techniques of ScaLiNG could potentially be applied to various other skin disorders. The ScaLiNG system's ability to efficiently extract and merge spatial-textural-frequency features by integrating compact CNNs and GW is a flexible strategy that can be applied to various tasks involving the evaluation of dermatological images. In order to assess the model's capability to be applied to a wider range of skin disorders, future investigations might include utilising ScaLiNG on datasets that cover a more diverse array of skin conditions. Furthermore, the ScaLiNG model's adaptability might be enhanced further by considering the integration of various other imaging techniques, including dermoscopy and clinical photography.

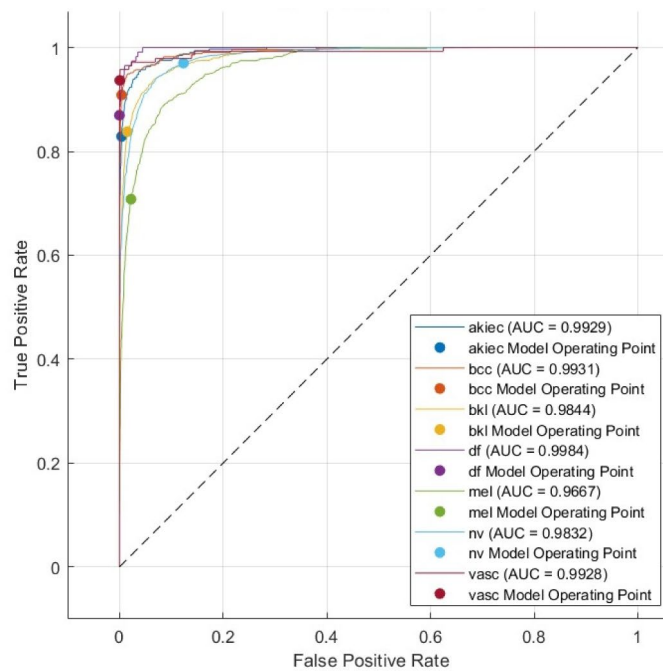
It is important to clarify that although the ScaLiNG CAD exhibits promising efficacy in categorising subtypes of SC, the main objective of the present study is to construct and assess a reliable classification model, without specifically taking into account individual patient features. Therefore, the model's ability to accurately recognise patients with specific characteristics including age, gender, or skin form, is not directly evaluated. This stated shortcoming can be addressed in future research by adding patient-specific data into the model, which has the possibility of improving its effectiveness and practicality. Furthermore, it would be worthwhile to investigate the establishment of submodels customised for particular patient populations.

ScaLiNG lacks a pre-segmentation stage. The choice to skip this stage is mainly driven by the aim of simplifying the procedure and decreasing the intricacy of computing. The goal behind ScaLiNG is to utilise the inherent capability of CNN networks to determine relevant characteristics without depending on intentional segmentation by simply entering the whole dermoscopic picture into the suggested CAD. Although removing pre-segmentation makes the CAD workflow less complex, it is recognised that this may create difficulties. For example, ScaLiNG could be required to acquire the ability to differentiate between the tumor and surrounding areas without explicit instruction. This has the potential to affect the precision of the system, particularly when the tumour only occupies a tiny area of the photo. To address this issue, the ScaLiNG CAD tool integrates numerous CNNs that operate across distinct image resolutions and scales. This approach enables the models to efficiently gather attributes at varying degrees of detail. Moreover, the utilisation of Gabor wavelets offers further texture data that may help identify and describe lesions.

The integration of any CAD framework, such as ScaLiNG, into an actual clinical environment presents numerous challenges. One significant obstacle is the attainment of generalisability. The performance of the ScaLiNG framework, which is developed on a particular set of images, may not be consistent across various patient groups that have various colours of skin, tumor forms, or imaging techniques used with various scanning devices. To address this issue, it is necessary to conduct external validation experiments that involve collecting data from a broader and more varied sample of patients from different scan centres. The wider range of data enables the model to effectively generalise and adjust to practical clinical situations. Moreover, it is essential to acquire trust and approval from medical professionals. To incorporate the ScaLiNG tool into the current clinical process, it is essential to engage together with dermatologists in order to comprehend their requirements and apprehensions.



(a) Linear SVM

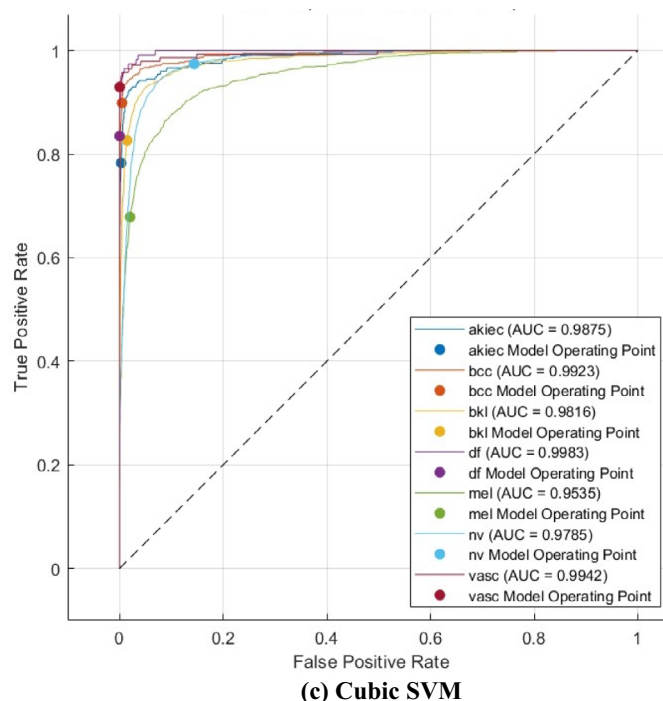


(b) Quadratic SVM

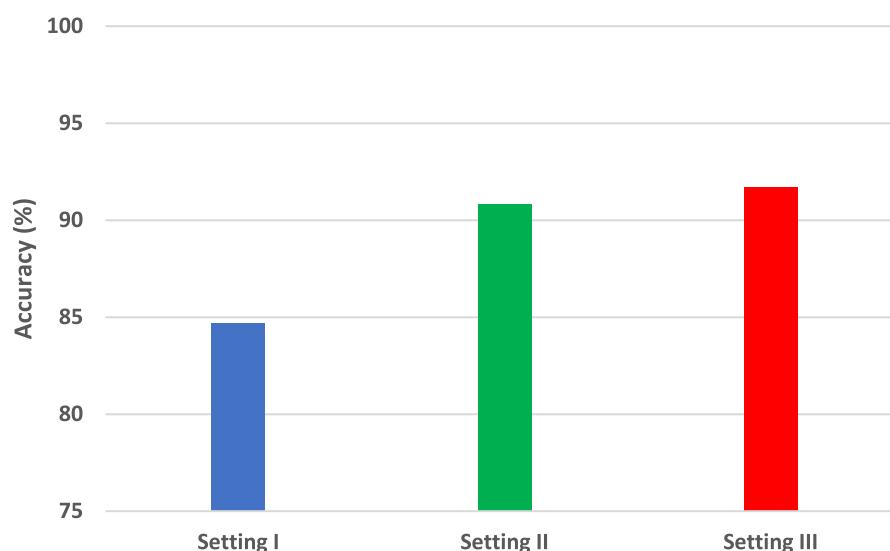
**Figure 9.** ROC curves for the three SVMs fed with the selected features of the mRMR FS approach (a) Linear SVM, (b) Quadratic SVM, and (c) Gaussian SVM.

This cooperation could assist in customising the design and capabilities of the system to seamlessly integrate into their diagnostic procedure. Furthermore, thorough clinical assessments conducted by qualified dermatologists are imperative to evaluate the efficacy in a real-life context and verify its dependability in aiding the diagnosis of skin cancer. To overcome such obstacles, SCaLiNG CAD could be strategically placed to gain greater acceptance and enhance its clinical effectiveness for a more diverse patient population.

In order to improve the dependability and suitability of the CAD in a clinical environment for a broader range of individuals, it is essential to collaborate with specialists and conduct external validation which will be done in future work. There are plans to collect data from a larger group of patients from different scan centers to enhance the generalisability and performance of the study. In the end, the system will be evaluated by professional dermatologists in a clinical environment. AD will be examined in clinical settings by experienced dermatologists.



**Figure 9.** (continued)



**Figure 10.** A comparison of the highest levels of accuracy achieved across all SCaLiNG settings.

## Conclusion

Although CNN-based CAD systems have been developed to assist dermatologists, these tools have certain limitations. Presently, many CAD systems depend on complex CNN structures, which may impede their widespread use. In addition, they often rely on an individual CNN model, which restricts the ability to capture a wide range of image features. Additionally, such CADs mainly focus on analysing the spatial information of the original images, while disregarding potentially valuable directional or textural details. This paper introduced a CAD called 'SCaLiNG' to identify and categorising SC at an early stage using a combination of three efficient CNNs of distinct structures (ResNet-18, ShuffleNet and MobileNet) and GW. SCaLiNG preserved four directional details by utilising GW to decompose the input photos into four directional subbands producing textural-frequency information. Five simultaneous CNNs of the same architecture were fed with the original image that produced spatial information and the four sub-bands that provided multiple textural-frequency representations. By merging the extracted deep spatial-textural-frequency features of each CNN architecture, a more comprehensive representation and enhanced classification accuracy were achieved. Additionally, SCaLiNG merged deep spatial-textural-frequency



Study	Approaches	Accuracy	Sensitivity	Precision	F1-score
58	U-Net based on Efficient Net B4 as a backbone	0.9103	0.8336	0.8444	0.8390
39	DenseNet-169	0.9120	–	–	0.9170
59	InceptionResNetV2 with Soft-Attention	0.900	0.820	0.810	0.810
35	Contrast enhancement, color space transformation illumination correction, ResNet-50	0.8850	–	0.8866	0.8860
42	Decorrelation deformation for preprocessing. Mask-R-CNN for segmentation LS-SVM for feature selection and classification	0.9208	0.9253	0.9373	0.9274
36	RegNet-320	0.9100	–	–	0.8810
60	Haar hair removal + EVGG	0.8900	0.8900	0.8900	0.8800
61	20-layer CNN for segmentation + 17 layered CNN for classification + Regula Falsi + Extreme learning machine (ELM)	0.8702	0.8698	–	–
SCaLiNG	GW + MobileNet + ShuffleNet + ResNet-18 + mRMR + Quadratic SVM	0.9170	0.9170	0.9170	0.9170

**Table 7.** Performance comparison of SCaLiNG and other existing CAD systems based on the HAM10000 dataset.

features of the three CNNs to combine the benefits of each unique structure, thus improving performance. It also used mRMR FS to choose the influential features. SCaLiNG CAD showcased its efficacy in the classification of multi-class skin lesions, attaining noteworthy outcomes. The proposed model attained an accuracy of 0.9170 using just 60 features. SCaLiNG CAD exhibited a superior enhancement in overall accuracy when compared with other existing CAD systems. These findings indicate that SCaLiNG is a reliable and effective CAD model for the detection, classification, and analysis of cancerous skin lesions. The results significantly enhance the progress of current knowledge in the early detection of skin cancer.

Data availability

The HAM10000 dataset analysed during the current study are available in the Harvard Dataverse repository, <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T> (accessed 30 May 2023).

Received: 20 June 2024; Accepted: 12 August 2024  
Published online: 04 September 2024

References

- Narayanan, D. L., Saladi, R. N. & Fox, J. L. Review: Ultraviolet radiation and skin cancer. *Int. J. Dermatol.* **49**, 978–986 (2010).
- Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Fabbrocini, G. *et al.* Epidemiology of skin cancer: Role of some environmental factors. *Cancers* **2**, 1980–1989 (2010).
- Nikolaou, V. & Stratigos, A. J. Emerging trends in the epidemiology of melanoma. *Br. J. Dermatol.* **170**, 11–19 (2014).
- Naqvi, M. *et al.* Skin cancer detection using deep learning—A review. *Diagnostics* **13**, 1911 (2023).
- Rigel, D. S., Russak, J. & Friedman, R. The evolution of melanoma diagnosis: 25 years beyond the ABCDs. *CA Cancer J. Clin.* **60**, 301–316 (2010).
- Hasan, M. K. *et al.* A survey, review, and future trends of skin lesion segmentation and classification. *Comput. Biol. Med.* **155**, 106624 (2023).
- Chanda, D. *et al.* DCENSnet: A new deep convolutional ensemble network for skin cancer classification. *Biomed. Signal Process. Control* **89**, 105757 (2024).
- Umirzakova, S. *et al.* Enhancing the super-resolution of medical images: Introducing the deep residual feature distillation channel attention network for optimized performance and efficiency. *Bioengineering* **10**, 1332 (2023).
- Pacal, I., Alaftekin, M., Zengul, F.D. Enhancing skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP. *J Digit Imaging Inform med.* Epub ahead of print 5 June 2024. <https://doi.org/10.1007/s10278-024-01140-8>.
- Pacal, I. & Kılıcarslan, S. Deep learning-based approaches for robust classification of cervical cancer. *Neural Comput. Appl.* **35**, 18813–18828 (2023).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Attallah, O. *et al.* Histo-CADx: Duo cascaded fusion stages for breast cancer diagnosis from histopathological images. *PeerJ Comput. Sci.* **7**, e493 (2021).
- Ghanem, N.M., Attallah, O., Anwar, F., *et al.* AUTO-BREAST: A fully automated pipeline for breast cancer diagnosis using AI technology. In: *Artificial Intelligence in Cancer Diagnosis and Prognosis, Volume 2: Breast and bladder cancer*. IOP Publishing, 2022.
- Attallah, O. Cervical cancer diagnosis based on multi-domain features using deep learning enhanced by handcrafted descriptors. *Appl. Sci.* **13**, 1916 (2023).
- Attallah, O. CerCan-Net: cervical cancer classification model via multi-layer feature ensembles of lightweight CNNs and transfer learning. *Expert Syst. Appl.* **229**, 120624 (2023).
- Pacal, I. MaxCerVixT: A novel lightweight vision transformer-based approach for precise cervical cancer detection. *Knowledge-Based Systems* 2024; 111482.
- Attallah, O. ECG-BiCoNet: An ECG-based pipeline for COVID-19 diagnosis using Bi-Layers of deep features integration. *Comput. Biol. Med.* 2022; 105210.
- Attallah O. RADIC: A tool for diagnosing COVID-19 from chest CT and X-ray scans using deep learning and quad-radiomics. *Chemom. Intell. Lab. Syst.* 2023; 104750.
- Attallah, O. A computer-aided diagnostic framework for coronavirus diagnosis using texture-based radiomics images. *Digit. Health* **8**, 20552076221092544 (2022).
- Attallah, O. Acute lymphocytic leukemia detection and subtype classification via extended wavelet pooling based-CNNs and statistical-texture features. *Image Vis. Comput.* **147**, 105064 (2024).

22. Attallah, O. An intelligent ECG-based tool for diagnosing COVID-19 via ensemble deep learning techniques. *Biosensors* **12**, 299 (2022).
23. Attallah, O. MB-AI-His: Histopathological diagnosis of pediatric medulloblastoma and its subtypes via AI. *Diagnostics* **11**, 359–384 (2021).
24. Attallah, O. CoMB-Deep: Composite deep learning-based pipeline for classifying childhood medulloblastoma and its classes. *Frontiers in Neuroinformatics* **15**, 663592 (2021).
25. Attallah, O. DIAROP: Automated deep learning-based diagnostic tool for retinopathy of prematurity. *Diagnostics* **11**, 2034 (2021).
26. Attallah, O. GabROP: Gabor wavelets-based CAD for retinopathy of prematurity diagnosis via convolutional neural networks. *Diagnostics* **13**, 171 (2023).
27. Haenssle, H. A. et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* **29**, 1836–1842 (2018).
28. Karthik, R., Menaka, R. & Siddharth, M. V. Classification of breast cancer from histopathology images using an ensemble of deep multiscale networks. *Biocybern. Biomed. Eng.* **42**, 963–976 (2022).
29. Nagaraj, P., Subhashini, S. J. A review on detection of lung cancer using ensemble of classifiers with CNN. In: *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*. IEEE, pp. 815–820.
30. Attallah, O. A deep learning-based diagnostic tool for identifying various diseases via facial images. *Digit. Health* **8**, 20552076221124430 (2022).
31. Buciu I, Gacsadi A. Gabor wavelet based features for medical image analysis and classification. In: *2009 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies*. IEEE, 2009, pp. 1–4
32. Celebi, M. E. et al. A methodological approach to the classification of dermoscopy images. *Comput. Med. Imaging Graph.* **31**, 362–373 (2007).
33. Serte, S. & Demirel, H. Gabor wavelet-based deep learning for skin lesion classification. *Comput. Biol. Med.* **113**, 103423 (2019).
34. Attallah, O. Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning. *Comput. Biol. Med.* **178**, 108798 (2024).
35. Kassani, S. H. & Kassani, P. H. A comparative study of deep learning architectures on melanoma detection. *Tissue and Cell* **58**, 76–83 (2019).
36. Alam, T. M. et al. An efficient deep learning-based skin cancer classifier for an imbalanced dataset. *Diagnostics* **12**, 2115 (2022).
37. Sethanan, K. et al. Double AMIS-ensemble deep learning for skin cancer classification. *Expert Syst. Appl.* **234**, 121047 (2023).
38. Alenezi, F., Armghan, A. & Polat, K. Wavelet transform based deep residual neural network and ReLU based Extreme learning machine for skin lesion classification. *Expert Syst. Appl.* **213**, 119064 (2023).
39. Gururaj, H. L., Manju, N., Nagarjun, A., et al. DeepSkin: A Deep learning approach for skin cancer classification. *IEEE Access*, <https://ieeexplore.ieee.org/abstract/document/10122533/> (2023, accessed 21 November 2023).
40. Bozkurt, F. Skin lesion classification on dermatoscopic images using effective data augmentation and pre-trained deep learning approach. *Multimed. Tools Appl.* **82**, 18985–19003 (2023).
41. Khan, M. A. et al. Pixels to classes: Intelligent learning framework for multiclass skin lesion localization and classification. *Comput. Electr. Eng.* **90**, 106956 (2021).
42. Khan, M. A. et al. Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. *Pattern Recogn. Lett.* **143**, 58–66 (2021).
43. Zhang, L. et al. A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm. *Biomed. Signal Process. Control* **90**, 105858 (2024).
44. Monica, K. M. et al. Melanoma skin cancer detection using mask-RCNN with modified GRU model. *Front Physiol* **14**, 1324042 (2024).
45. Akilandasowmya, G. et al. Skin cancer diagnosis: Leveraging deep hidden features and ensemble classifiers for early detection and classification. *Biomed. Signal Process. Control* **88**, 105306 (2024).
46. Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **5**, 1–9 (2018).
47. Qin, S. et al. Facial expression recognition based on Gabor wavelet transform and 2-channel CNN. *Int. J. Wavelets Multiresolut. Inf. Process.* **18**, 2050003 (2020).
48. Huh, M., Agrawal, P. & Efros, A. A. What makes ImageNet good for transfer learning? *arXiv preprint arXiv:160808614*.
49. Remeseiro, B. & Bolon-Canedo, V. A review of feature selection methods in medical applications. *Comput. Biol. Med.* **112**, 103375 (2019).
50. Cai, J. et al. Feature selection in machine learning: A new perspective. *Neurocomputing* **300**, 70–79 (2018).
51. Attallah, O. Tomato leaf disease classification via compact convolutional neural networks with transfer learning and feature selection. *Horticulturae* **9**, 149 (2023).
52. Attallah, O. MonDial-CAD: Monkeypox diagnosis via selected hybrid CNNs unified with feature selection and ensemble learning. *Digit. Health* **9**, 20552076231180056 (2023).
53. Bashir, S., Khattak, I. U., Khan, A., et al. A novel feature selection method for classification of medical data using filters, wrappers, and embedded approaches. *Complexity*; 2022, <https://www.hindawi.com/journals/complexity/2022/8190814/> (2022, accessed 29 February 2024).
54. Ding, C. & Peng, H. Minimum redundancy feature selection from microarray gene expression data. *J. Bioinform. Comput. Biol.* **03**, 185–205 (2005).
55. Wang, G., Lauri, F. & El Hassani, A. H. Feature selection by mRMR method for heart disease diagnosis. *IEEE Access* **10**, 100786–100796 (2022).
56. Ramirez-Gallego, S. et al. Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data: FAST-mRMR algorithm for big data. *Int. J. Intell. Syst.* **32**, 134–152 (2017).
57. Peng, H., Long, F. & Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1226–1238 (2005).
58. Alam, M. J. et al. S2C-DeLeNet: A parameter transfer based segmentation-classification integration for detecting skin cancer lesions from dermoscopic images. *Comput. Biol. Med.* **150**, 106148 (2022).
59. Nguyen, V. D., Bui, N. D. & Do, H. K. Skin lesion classification on imbalanced data using deep learning with soft attention. *Sensors* **22**, 7530 (2022).
60. Mushtaq, S. & Singh, O. A deep learning based architecture for multi-class skin cancer classification. *Multimed. Tools Appl.* <https://doi.org/10.1007/s11042-024-19817-1> (2024).
61. Khan, M. A. et al. Intelligent fusion-assisted skin lesion localization and classification for smart healthcare. *Neural Comput. Appl.* **36**, 37–52 (2024).

# Author contributions

Omneya Attallah: Conceptualization, Methodology, Software, Writing—original draft, Writing—review & editing, Validation, Visualization, Investigation, Data curation.

## Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). The author carried out this research without any kind of funding.

## Competing interests

The author declares no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-69954-8>.

**Correspondence** and requests for materials should be addressed to O.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024