

SCD-E2S: Deep Learning Based Skin Cancer Detection Using EfficientNetV2-S with Attention Mechanisms

Nabil Aref Abdullah Ghaleb

Taiz University

Mogeeb A. Saeed

Alsaeed university

Mohammed Ahmed Ali Abdualrhman

m.alherwi@ust.edu

University of Science and Technology

Research Article

Keywords: Skin cancer (SC), deep learning (DL), image analysis, diagnosis, basal cell carcinoma (BCC), melanoma (MEL), melanocytic nevus (NV)

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10863659/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

SCD-E2S: Deep Learning Based Skin Cancer Detection Using EfficientNetV2-S with Attention Mechanisms

Nabil Aref Abdullah Ghaleb^a- fiet@taiz.edu.ye,
Mogeeb A. Saeed^{a,c}- mogeeb.saeed@alsaeeduni.edu.ye

* **Corresponding author** : Mohammed Ahmed Ali Abdualrhman^b m.alherwi@ust.edu,

^aDepartment of Information Technology, Faculty of Engineering and Information Technology, Taiz University, Taiz, Yemen

^bInformation Technology Department, Faculty of Engineering & Computing, University of Science and Technology, Aden, Yemen

^cDepartment of Cybersecurity, Faculty of Engineering & IT, Alsaeed University, Taiz Yemen

ABSTRACT

The worldwide increasing incidence of SCs, such as melanoma, melanocytic nevi, basal cell carcinoma, necessitates the validation of reliable early detection method. Thus, to solve this issue, we present a deep learning solution to the multi-class Dermoscopy Image classification based on lightweight attention mechanisms of EfficientNetV2-S architecture. The black-hat transformation, class rebalancing via Underdamping and contrast limited matrix equalization (CLAHE) were applied to Preprocessing of Dermoscopic Images from the large-scale dataset. Besides, to further boost the performance of proposed model in term of generalization and eliminate class imbalance problem, two-stage training was adopted to train the proposed model with Focal Loss, MixUp and L2 regularization. It outperforms several baseline architectures, and over 97% accuracy is obtained for the model when evaluated on two external datasets. The proposed model was encapsulated in a mobile app using Flutter and a Flask RESTful API to allow its practical application. Our system actually performed impeccably in practice and there is every likelihood of being extremely helpful in clinical decision making for a dermatologist.

INDEX TERM: Skin cancer (SC), deep learning (DL), image analysis, diagnosis, basal cell carcinoma (BCC), melanoma (MEL), melanocytic nevus (NV).

1. INTRODUCTION

The burden of SC has increased globally; with melanoma, basal cell carcinoma (BCC), and melanocytic nevi (NV) being the most common and clinically relevant forms of SC. It is critical to detect SC in early stages to improve both the prognosis and the treatment options available to patients[1].

However, Although the diagnosis of SC through visual means is an inaccurate process due to the complexity and variability in both physician experience and the morphological characteristics of the lesions, as well as the variability in the skin types of patients, advances in artificial intelligence

(AI), specifically through deep learning (DL) techniques, have greatly improved the automatic diagnosis of dermatological conditions[1].

Convolutional Neural Networks (CNNs) are especially suited for the analysis of medical images because they are able to generate hierarchical, data-driven representations of features that are learned from the pixel-level input of images. These CNNs eliminate the need for manual feature extraction which has led to recent research demonstrating that DL models can match or exceed the accuracy of dermatologists in the classification of skin lesions, improve the speed and consistency of diagnoses, and increase access to dermatological expertise for those living in resource-limited environments [2][3].

Despite this progress, there are still several barriers to overcome. The majority of dermatology datasets used to train AI models are imbalanced in terms of the classes represented within them, with malignant (MEL) being one of the least represented classes. This class imbalance can lead to biased models that do not generalize well to new, unseen populations. Additionally, while many models exist that utilize DL techniques, few of them provide clear explanations of how the model came to a particular prediction, and even fewer have been successfully implemented in real-world clinical settings, especially in mobile or point-of-care environments.

To address these issues, we propose a deep-learning-based diagnostic model for the early detection of MEL, BCC, and NV from Dermoscopic images. Our proposed model utilizes the EfficientNetV2-S architecture, which provides a good balance between performance and computational cost, and enhances this architecture by adding lightweight attention mechanisms to enable better localizing of important features. Additionally, our proposed model includes a robust preprocessing pipeline that addresses issues related to image artifacts and class imbalance to optimize the quality of the images provided as input to the model. Advanced training strategies (two-phase fine-tuning with adaptive loss functions and augmentation) further improved generalization and helped the model work better on new cases. The system was developed and validated before and then implemented as a phone application for rapid and portable assessment of skin lesions. The software has been incorporated as a clinical decision-support system, which could aid in the early diagnosis of lesions of suspicion, but should not supplant clinical and histopathological expert diagnosis. To this end, this paper aims at closing the gap between the research based classifiers of skin cancer and the real clinical dermatology use with high accuracy prediction, external validation and supportive usability in real-time use. Then in next section, previous studies in the field of deep learning applied to dermatology are summarized and the contribution proposed. The data sets used, pre-processing algorithms, CNN architecture, training procedure and evaluation measures are described concisely in this article. Results: This part reviews the performance of the model on the training, validation and external-test sets as well as their diagnostic accuracy; the Conclusion part outlines the contributions, limitations, and directions for future work.

2. RELATED WORK

In recent years, much work has been done on SC classification using deep learning. In recent years, there are huge advancements in deep learning for classification of SC. Many research papers have shown that CNN methods can perform better than conventional methods of diagnosis by providing more accurate and reliable results when analyzing Dermoscopic images. Musthafa and others [4] developed an improved CNN model using the HAM10000 dataset for instance. A lot of data augmentation was employed to deal with the unbalanced classes. They achieved 97.8% accuracy

and this was better than more complex models such as ResNet and DenseNet. The fact that they only use very small images (only 28x28 pixels), and not test their results on other datasets, is not helpful for understanding how well their results perform in real contexts. Ali et al.[5] also presented a hybrid CAD system which used a fully convolutional network, like the one used by Musthafa et al. optimized with the SpaSAnet algorithm for lesion identification with a TreeCNN classifier, as aid to the diagnosis. The classification rate of this method was up to 98.4%, but the computing power was high and could not be used in reality. There are other attempts to improve model performance with hyperparameter tuning and transfer learning. Singh et al. [6] evaluated different convolution neural network and machine-learning models for classification of skin lesions in the HAM10000 dataset. Based on their findings, they propose that the performance of the model can be improved by using the meta heuristic optimization for certain important training and model parameters. Many problems of class imbalance remained, however, and the hardest were the detection of the least populated lesion categories. Furthermore, the parameters were optimized carefully and the whole optimization process were repeated a few times, which might introduce additional computation cost and restrict the application of the proposed method to real-time clinical applications.

Strzelecki et al. [7] gave a comprehensive introduction on the application of AI in skin-cancer diagnosis, which covered analysis of Dermoscopic images, total-body imaging, and applications based on smartphones. The main value of their review in this work is that they have pointed out that it is important to test models of AI in situations that are not only different from the training data, but are also harder to solve. The review identified a number of challenges that could undermine the trustworthiness of AI systems in practical applications, such as variations in image-acquisition protocols and Datasets, class imbalance, and inadequate clinical validation. which did not specifically consider the integration of EfficientNetV2-S with SE/ECA attention mechanism, focal loss, MixUp augmentation, and independent external validation over HAM10000 and ISIC 2017, which are the dominant contributions of this work.

Agrawal et al.[8] explored transfer learning with InceptionV3 on a smaller dataset and achieved an accuracy of 84.8 %. Their study reinforced the value of pretrained models for dermoscopy analysis but also underscored the challenges associated with small, imbalanced datasets and limited class diversity.

Flosdorf, Engelker, Keller, and Mohr [9]. proposed two larger pre-trained Vision Transformer configurations (ViT L16/L32) for classifying Dermatoscopic skin lesion images from the Skin Cancer MNIST: HAM10000 dataset (seven classes), using extensive data augmentation to mitigate class imbalance, and achieved a best accuracy of 92.79% (melanoma recall 56.10%), outperforming baseline decision tree and KNN models and prior smaller ViTs/CNN comparisons reported in the paper; however, the authors noted that the dataset's uneven class representation may reduce performance generalizability to more diverse real-world distributions.

Millenia, Naufal, and Siswanto[10]. compared CNN transfer-learning architectures (MobileNet, Xception, VGG16, and ResNet50) for melanoma-versus-mole classification using merged dermoscopic and macroscopic images from ISIC, MED-NODE, and PAD-UFES-20, applying black-hat filtering with inpainting (hair/noise removal) to help with the imbalanced class distribution; the best reported performance was an average F1 score of 84.43% (MobileNet), outperforming the other evaluated transfer-learning models and reporting 13.82% higher accuracy than Pillay et al. on the same macroscopic dataset (and 2.24% higher recall than Murugan et al. on

the same dermoscopic dataset); however, they noted that several prior-study datasets could not be tested because they were not publicly available, limiting the breadth of external comparison.

Muhammad, Kiani, Mansouri, and Topping [11]. proposed SkinLesNet, a novel four-layer multi-layer deep CNN for skin lesion classification and melanoma detection, trained primarily on an augmented PAD-UFES-20-Modified smartphone-image dataset (nevus, melanoma, seborrheic keratosis) and additionally evaluated on HAM10000 and ISIC2017, using geometric-transformation data augmentation (e.g., random flips/rotations/translations) to address limited data and class-distribution challenges; the best reported result was 96% test accuracy on PAD-UFES-20-Modified, outperforming fine-tuned ResNet50 (82%) and VGG16 (79%); however, they noted that the chosen datasets may not cover the full spectrum of real-world skin lesion variation, which can limit generalizability.

Azeem et al. [11] also noted that overfitting, dataset breadth, and computational requirements may still impact generalizability and they will continue to seek to validate and refine strategies for attention and generalization (such as MixUp) while our study tackles these challenges through its focus on generalizability and practical translation by integrating an EfficientNetV2-S backbone enriched with SE/ECA attention, a comprehensive preprocessing/class-balancing pipeline, and advanced training strategies (e.g., focal loss, MixUp), bolstered by explicit external evaluation on independent HAM10000 and ISIC 2017 datasets, thereby ensuring not only high performance but also usability on the clinical platform..

To summarize, the various studies in the past have contributed significantly in the areas of attention-based feature extraction, computational efficiency, handling class imbalance, and domain-shift analysis. But there are some restrictions. Most studies are based on a within-dataset evaluation or a single benchmark dataset, with only a few studies integrating attention mechanisms, imbalance-aware training with advanced augmentation methods in a unified pipeline, and only a few studies mentioning real-time usability without providing any measurement of latency or computational cost. To fill this gap, the present study proposes to combine an EfficientNetV2-S backbone with SE/ECA attention, focal loss, MixUp, and an exhaustive class-balancing and preprocessing pipeline. Furthermore, the proposed model is evaluated independently on the HAM10000 and ISIC 2017 datasets to examine the cross-dataset generalizability, and efficiency of the computations is taken as a requirement for practical clinical translation.

3. METHODOLOGY

In this section we provide a description of the dataset, pre-processing, CNN architecture and improvements, training and evaluation protocol, and an overview of the entire workflow (see Figure 1).

3.1 Dataset and Preprocessing

Data Sources: Data Description: The main data set used in this work is the Skin Diseases Image Dataset from Kaggle, which consists of clinical and Dermoscopic images of skin diseases about

27,000. From this source, we extracted a subset of 14,433 Dermoscopic images we extracted a subset of 14,433 Dermoscopic images belonging to the three target classes[12].

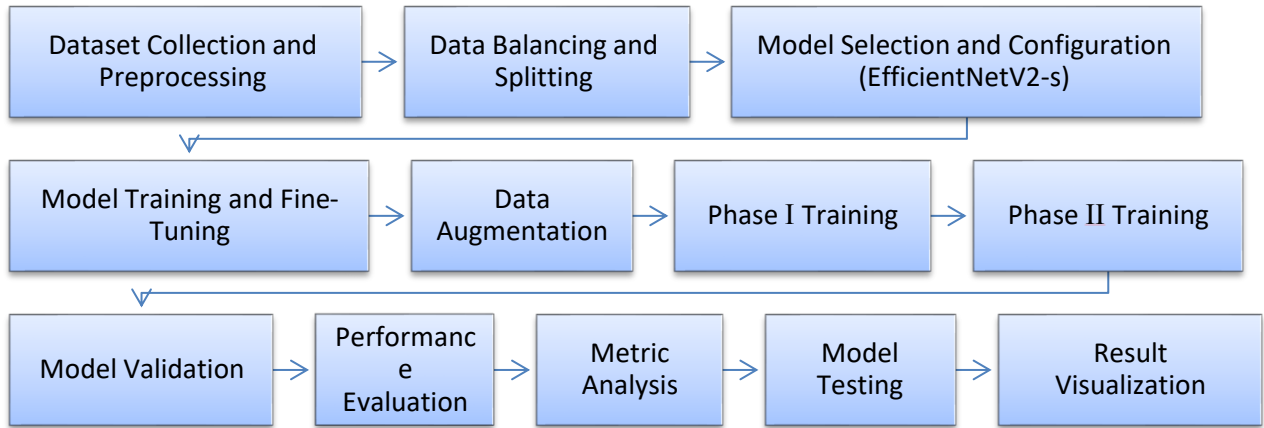


Figure 1: Workflow Methodology

These classes are MEL, NV, and BCC. Figure 2 illustrates example dermoscopic images for each of the three classes (see Table 1 for the data distribution). External test samples were incorporated from the HAM10000 [13], [14] and ISIC 2017 [15] datasets. For external evaluation, 20 images per class were selected as the test samples. These external images were used only for model validation/testing and were not included in the model training.

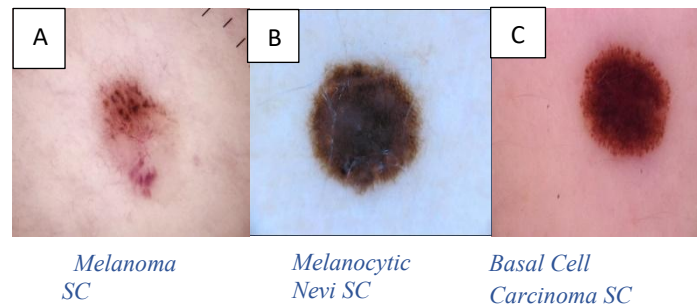


Figure 2: Types of SC

Table 1 : Data Distribution

Class	Data
MEL	3,141
Melanocytic	7,970
BCC	3,323

To improve the image quality and model robustness, a series of preprocessing steps was applied to all images before training.

- **Image Resizing:** All images were resized to 380×380 pixels using bicubic interpolation to match the input dimension requirements of the EfficientNetV2-S architecture[16].
- **Artifact Removal:** A morphological black-hat transformation followed by Navier-Stokes inpainting was used to eliminate hair strands, ruler marks, and other artifacts that commonly appear in dermoscopic images.[10]
- **Contrast enhancement:** Contrast-limited adaptive histogram equalization (CLAHE) was applied to the luminance channel in the $L a b^*$ color space, which improved the visibility of the lesion details.
- **Class Balancing:** The overrepresented NV class was reduced using Random UnderSampler to balance the class distributions.
- **Normalization:** The pixel values were scaled to [0,1] for stable CNN training.

After preprocessing and balancing, the dataset was organized into class-specific directory. The processed dataset was divided into 80% training and 20% test sets in a stratified manner to preserve the class ratios, preserving the original label distributions as shown in Figure 3. This ensured that both the training and test sets had a balanced class distribution.

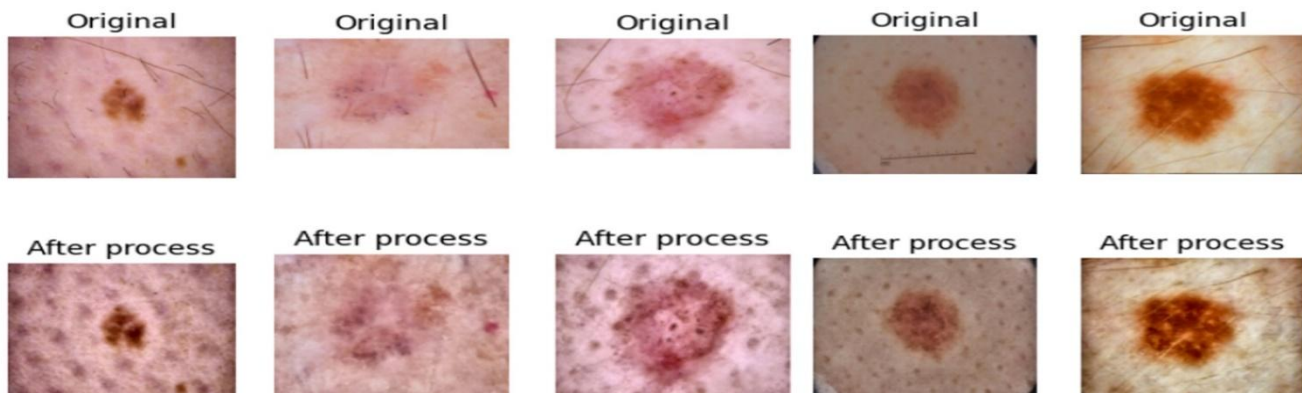


Figure 3: Corporation After Preprocessing

It was found that there is no overlap between training and test sets in terms of images. In addition to this internal split, an external validation set composed of independent images the HAM10000 [13], [14] and ISIC 2017 datasets[15] was used to evaluate the model's generalization. These external datasets were completely different from the Kaggle dataset, and there was no training or tuning here. Consequently, there was no overlap in the training, validation and testing sets. Using this strategy, we can evaluate the model performance on a held-out test that we keep internally, as well as on external data from other sources, thereby getting a more solid estimation of the model's generalizability.

3.2 Model Architecture

The presented classifier leverages an efficient and scalable CNN architecture, namely EfficientNetV2S [17][18][19], that is specially crafted for high resolution image classification. It excels the other baselines

in efficiency and effectiveness such as ResNet-50[20], MobileNetV2[21], EfficientNet-B4[22] on the ImageNet dataset with top-1 accuracy, with 22 Million parameters and ~ 8.8 Billion floating-point operations (FLOPs). The input size, ImageNet accuracy and the complexity of these three representative architectures are listed in Table 2.

EfficientNetV2-S [17] was chosen as the base model because of its high accuracy and moderate size, making it suitable for deployment on limited-resource devices.

To improve the discriminative capability, the backbone is enhanced with two lightweight attention modules:

- **Squeeze-and-excitation (SE)** blocks were used to model global channel dependencies[23].
- **Efficient Channel Attention (ECA)** modules were used to refine feature maps with minimal overhead [23].

Table 2: Comparison of models. Input size, top-1 accuracy, Parameters, FLOPs

Model	Input Size	Top-1 Acc (%)	Parameters (M)	FLOPs (B)
ResNet-50 [20]	224 $\times$ 224	76.0	26	4.1
MobileNetV2 (1.0) [21]	224 $\times$ 224	72.0	3.4	0.3
EfficientNet-B4[22]	380 $\times$ 380	82.9	19	4.2
EfficientNetV2-S[17]	380 $\times$ 380	83.9	22	8.8

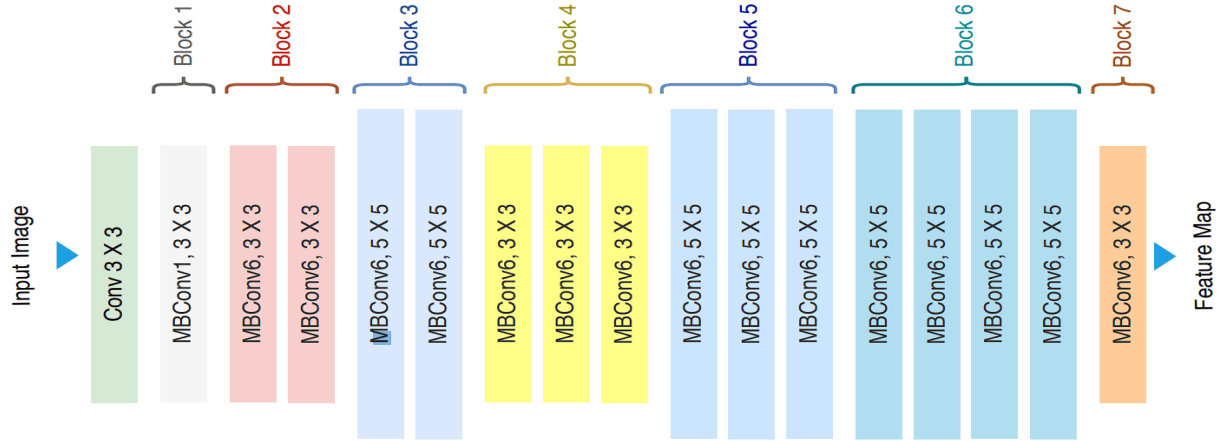


Figure 4: EfficientNet Architecture

Following the EfficientNetV2-S backbone [24] (with SE and ECA), a custom classification head was added. This head consisted of a 1×1 convolutional *bottleneck* layer with Swish activation [16], a global average pooling layer, and two fully connected (dense) layers of size 256 and 128. We applied L2 kernel regularization and dropout (dropout rates of 0.4 and 0.3) to the dense layers to reduce overfitting. Finally, a Softmax layer produces probability outputs for the three classes (melanoma, NV, and BCC). Figure 4 provides an overview of the EfficientNet-based architecture with the added attention modules and classification heads.

3.3. Training Strategy and Optimization

As illustrated in Figure 5, we adopted a two-phase transfer learning strategy to train the EfficientNetV2-S model effectively using our dermoscopy dataset.

- **Stage I:** We froze the whole EfficientNetV2-S backbone and only trained the attached classification head. This step allows for the new layers to learn initial weights for the skin lesion classes while maintaining the pre-trained convolutional features.
- **Stage II:** end-to-end fine-tuning: The top half of the EfficientNetV2-S backbone layers were unfrozen (a lower half was kept frozen to preserve the basic visual features), and then the entire model was fine-tuned. The fine-tuning was limited to 50 epochs or until convergence. By gradually unfreezing the layers, we managed to prevent catastrophic forgetting and obtain a stable learning process.

We trained the model using the AdamW optimizer [25] and the One-cycle learning rate scheduler [25], both of which are popular for the strong convergence and generalization. The training process was cyclically adjusting the learning rate in order to efficiently explore the loss landscape, essentially following cyclic learning rates. We used Focal Loss [26] as the loss function with dynamic focusing factor γ which was gradually increased from 1.5 to 2.0. Useful for class imbalance and improved recall of the underrepresented class(es) (melanoma) as Focal Loss performs better on hard samples (misclassified samples) rather than on easy samples (correctly classified samples).

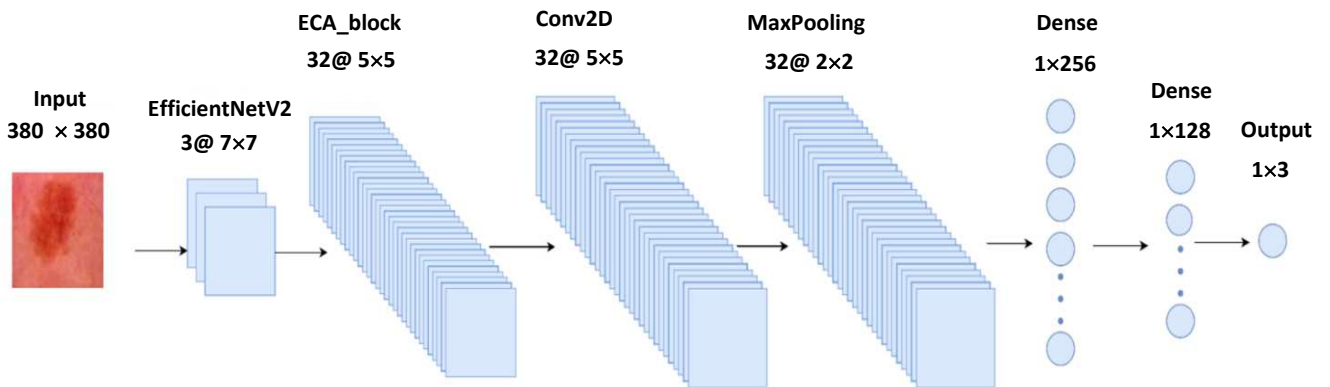


Figure 5: CNN Model Visualization

A variety of regularization/augmentation methods were employed as well to enhance generalization.

- **MixUp Augmentation:** When the training accuracy reached a plateau of around 94% after the initial training, we switched on MixUp, which creates virtual training samples by mixing two raw images as well as their labels. MixUp helps to smoothen the decision boundary and provides more stable outputs with less variance, thus it is more robust to noisy points..

- **Dropout and L2 Regularization:** Dropout (with levels between 40% and 30%) and L2 weight penalties were used in the fully connected layers to prevent neurons from co-adapting and thus overfitting.
- **Class-Balanced Training:** Apart from Focal Loss, we also employed class weighting (and Undersampling the minority class, which was melanoma) to penalise more the misclassification of the minority class in training. The sensitivity of the melanoma class was further improved [27].

Training was performed on a COLAB GPU [28], and early stopping with patience was used to terminate training if the validation loss stopped improving. The best model checkpoint (with the highest validation accuracy) was retained for the final evaluation.

3.4. Evaluation Metrics

A detailed set of metrics were used to assess the model performance.

- **The overall Accuracy** was calculated as the percentage of accurately classified images in the test set. To give a more detailed analysis, the Precision, Recall, and F1-Score were calculated for each class and their macro averages, to evaluate the model performance on each skin lesion category. The precision and recall in the melanoma class is of particular importance due to the need to have a low rate of false negative (missed melanomas) [29].
- **The area Under Curve (AUC)** was calculated (macro-averaged over the 3 classes) as a global measure of the diagnostic performance.
- **Confusion Matrix** of the predictions to identify the distribution of classification errors between classes.

Table 3 below summarizes the mathematical formulation and evaluation criterion of each metric used in this study.

Table 3 formulation and evaluation criterion of each metric used in this study.

Metric	Formula	Description / Evaluation Criterion
Accuracy	$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$	Proportion of correctly classified images among the entire test set.
Precision (per class & macro-avg.)	$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$	Proportion of positive predictions that are correct; a low precision reflects a high false-positive (over-diagnosis) rate.
Recall / Sensitivity (per class & macro-avg.)	$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$	Proportion of actual positive cases correctly identified; critical for minimizing missed melanomas (false negatives) [29].
F1-Score (per class & macro-avg.)	$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$	Harmonic mean of precision and recall, providing a

Metric	Formula	Description / Evaluation Criterion
AUC (macro-averaged)	$AUC = \int TPR d(FPR)$, $TPR = TP/(TP+FN)$, $FPR = FP/(FP+TN)$	Area under the ROC curve (True Positive Rate vs. False Positive Rate across all thresholds); a global, threshold-independent measure of discriminative
Confusion Matrix	C_{ij} = count of samples with true class i predicted as class j	Cross-tabulation of predicted vs. true class labels; used to visualize the distribution of classification

All metrics were first evaluated on the internal test set (from the 20% holdout of the Kaggle data). To validate generalization, the same metrics were later computed on the external HAM10000 [13], [14] /ISIC [15] dataset images. We report and compare the performance of four different CNN models (ResNet-50 [20], MobileNetV2 [21], EfficientNet-B4 [22] , and our EfficientNetV2 [17] -S with attention) using these metrics. Early stopping and checkpoint callbacks were used to prevent overfitting and to preserve the best-performing models.

4. Results and discussion

4.1. Model Performance and Comparison

The EfficientNetV2-S model demonstrated strong performance in the SC classification task. During training, the two-phase fine-tuning approach resulted in stable convergence. Figure 6 and Figure 7 plot the training and validation accuracies and loss curves per epoch, respectively. The training and validation curves were closely aligned, indicating minimal overfitting. By the end of the training, the validation accuracy nearly matched the training accuracy, and the validation loss remained low. After 50 epochs, the fine-tuned EfficientNetV2-S achieved an overall accuracy of 97.45% on the internal test set. The average precision, recall and F1 scores for the model were 97.3%, 97.6% and 97.6% respectively (across the 3 classes). Overall (macro-averaged), the AUC value was 0.98, which is a very good discriminative ability. The results are evidence that the image classification model can correctly classify melanoma, nevus and BCC images at high reliability.

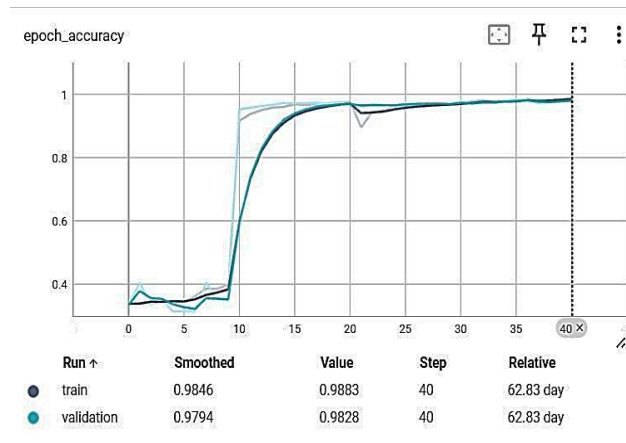


Figure 6: Per-Epoch Training and Validation Accuracy

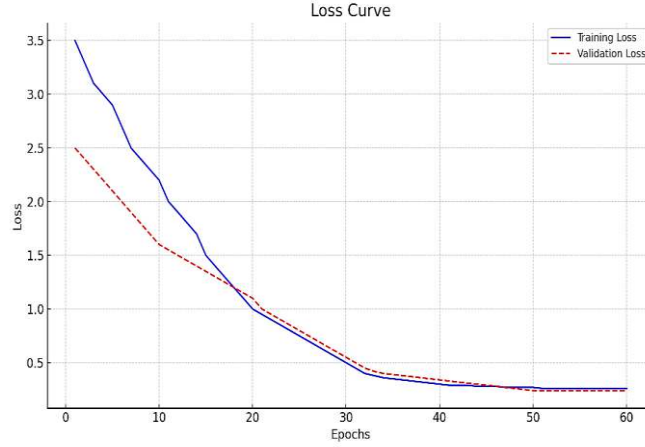


Figure 7: Training and Validation Loss

Table 4 shows the key evaluation metrics (accuracy, precision, recall, and F1 score, and AUC) of the three other deep CNN architectures evaluated and compared with EfficientNetV2-S.) for ResNet-50, MobileNetV2, EfficientNet-B4, and the proposed EfficientNetV2-S model.

Precision, Recall and F1 scores were macro averaged over the 3 classes. The EfficientNetV2-S (SE + ECA attention) model is the best performer for all metrics compared with the baseline architectures.

The EfficientNetV2-S model performed best, by yielding the highest accuracy and AUC compared to the ResNet-50 model, the MobileNetV2 model and the EfficientNet-B4 model, as seen in Table 4. Moreover, efficientNetV2-S has better accuracy and recall, better balancing the sensitivity and specificity of recognizing all classes of SC.

Interestingly, the lightweight model, MobileNetV2, had the lowest accuracy (~90.5%), suggesting that its capability may be limited for this task. ResNet-50 was able to reach an accuracy of around 94 %, while EfficientNet-B4 managed to reach an accuracy slightly better and reached an AUC of 0.97, but was also slightly lower than the proposed model. The results show that our training improvements (attention modules and advanced augmentation) and the EfficientNetV2-S backbone produce state-of-the-art results on the dataset.

Table 4: Comparative Analysis of Different CNN Architectures

Model Architecture	Accuracy (%)	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	AUC
MobileNetV2 (Baseline)	90.50	90.1	89.8	89.9	0.94
ResNet-50	94.00	93.8	93.5	93.6	0.96
EfficientNet-B4	95.50	95.2	95.4	95.3	0.97

Proposed EfficientNetV2-S	97.45	97.30	97.60	97.60	0.98
----------------------------------	--------------	--------------	--------------	--------------	-------------

We also analyzed the model's performance per class and the distribution of errors. The detailed classification report (Table 5) reveals that the BCC class was classified with the highest precision and recall (close to 0.99 for both) indicating that almost all BCC lesions were correctly classified and very few were falsely classified. The less abundant and most important class, namely melanoma (MEL), had a slightly lower, though still very good performance (around 0.96 precision and 0.97 recall). The nevus (NV) class also had a precision and recall of ~ 0.97 . The F1 scores for each class were all high (>0.97), which confirmed the model's performance for all types of lesions.

After 50 epochs, the fine-tuned EfficientNetV2-S achieved the overall performance summarized in Table 6. All values for precision, recall, and F1-score were macro-averaged across the three classes.

Table 5: Overall Performance Metrics of the Proposed EfficientNetV2-S Model on the Internal Test Set

Metric	Value
Accuracy	97.45%
Precision (macro-avg.)	97.3%
Recall (macro-avg.)	97.6%
F1-Score (macro-avg.)	97.6%
AUC (macro-avg.)	0.98

Table 6: . Per-Class Classification Report of the Proposed Model

Class	Precision	Recall	F1-Score	Support
BCC	0.99	0.99	0.99	665
MEL	0.96	0.97	0.97	628
NV	0.97	0.97	0.97	994
Accuracy	—	—	0.97	2287
Macro avg.	0.97	0.98	0.98	2287
Weighted avg.	0.97	0.97	0.97	2287

4.2. Diagnostic Error Analysis

This confusion matrix (Figure 8) also demonstrates that it was possible to make very few mistakes in the classification process. In the cases of image classification, the most common misclassification problem was between images of melanoma and images of nevus, which is not surprising because of the resemblance between some nevus images and early melanomas in appearance. The model retained a high recall for the detection of melanoma, however, indicating that our approach of focal loss and class rebalancing was effective in achieving this minority class detection as evidenced by the improvement in melanoma recall [30]. Apart from the occasional melanoma–NV confusion, the model was not very

prone to misclassifying BCC with other categories, which suggests that the model has learned the distinct characteristics of BCC.

Overall, the results show that our EfficientNetV2-S with attention mechanism performs well on the training data and generalizes well. The model maintained its high accuracy (above 95%) when tested on external HAM10000 [13] and ISIC 2017 [15] images (external validation set), showing the ability of the system to diagnose also new distributions of data. The external evaluation stresses the strength and adaptability of the model for practical use.

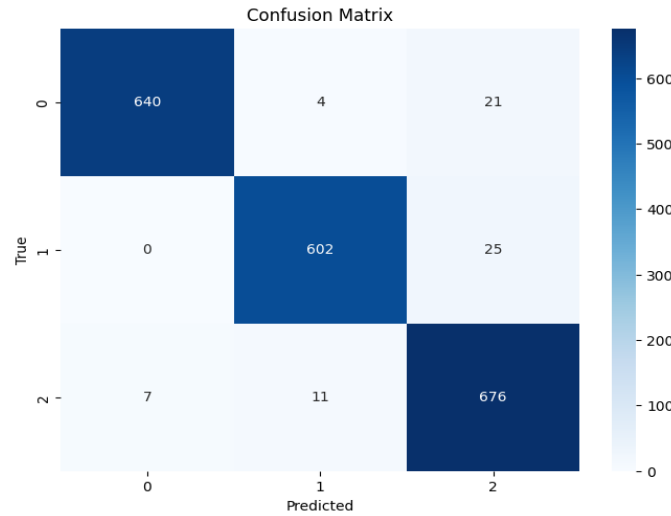


Figure 8: Confusion Matrix of Model Prediction

4.3 Discussion

The results of the experiment fully verify that the deep learning-based diagnosis of melanomas, nevi, and basal cell carcinoma in dermoscopic images can reach a diagnostic level of expert dermatologists. From the results, it can be observed that EfficientNetV2-S with SE and ECA attention mechanism achieves the best performance on all the CNN based baselines on both accuracy and average AUC, which indicates that our proposed channel-wise and feature-wise attention mechanism can extract more discriminative features to classify fine-grained dermoscopic images. With the two-stage fine-tuning, focal loss, and MixUp-based augmentation integrated into the model, it is able to effectively address the class imbalance issue and overfitting problem; in addition, the distractor modules assist the network to focus on the most informative regions for each lesion diagnosis, rather than the mean regions of the lesions in the global. So, the model achieved a very high sensitivity for melanoma, which is of great clinical importance since the requirement of any practical diagnostic tool for everyday use is a rapid and reliable diagnosis. In addition, the training accuracy, validation accuracy are also close, and the model performs well in external test set, indicating that this improvement of performance may come from the

real power of generalization rather than data over-fitting to the training distribution, which is a good property of clinical translatability.

4.3.1 Comparison with the State-of-the-Art Methods in Skin Cancer Classification:

To provide a tangible demonstration of the superiority of the proposed EfficientNetV2-S over state-of-the-art SC automation, we made a specific comparison with six recent (2021-2024) works leveraging convolutional, hybrid and/or transformer-based architectures on comparable Dermoscopic image datasets. In order not to list all the researches that were encompassed by the survey, in this work we only consider those works that included all of the following aspects: complete and reproducible performance results (accuracy, precision, recall, and F1-Score); employed tracking algorithms on publicly available or clinically representative Dermoscopic datasets (HAM10000, PH2, ISIC, or PAD-UFES-20); and published in the last three years to represent current state of the art and not an obsolete base line. And since none of the reports on different aggregates of the several systems were taken into account in the quantitative analysis, neither paper describing small, highly unrepresentative, private and underwhelming custom data sets with significantly less stunning reported accuracy ($< 90\%$) were considered. Studies using such data sets would not constitute a methodologically sound basis for benchmarking a diagnostic model with the potential to be put into use. The best performing papers and our method's accuracy, precision, recall, and F1-score are shown in Table 7 In accordance with the reporting style of the compared studies, precision, recall, and F1 score were macro-averaged over the 3 diagnostic groups (melanoma, nevus, basal cell carcinoma) for the proposed model.

Table 7 Comparative Performance of the Proposed Model Against Recent State-of-the-Art Studies

Study	Dataset	Model	Accuracy	Precision	Recall	F1-Score	External Validation
Proposed (This Study)	Custom (Mel / Nevus / BCC)	EfficientNetV2-S + SE/ECA	97.45%	97.3%	97.6%	97.6%	Yes (HAM10000, ISIC 2017)
Musthafa et al. [4]	HAM10000	Optimized CNN	97.78%	97.9%	97.9%	97.8% (F2)	No
Ali et al. [5]	PH2 / ISIC 2019	FCEDN-SpaSA + Adaptive CNN	98.42–98.78%	91.27%	97.98%	97.92%	No
Singh et al. [6]	HAM10000	DNN + Bayesian Optimization	90.56%	91.59%	94.64%	94.0%	No
Flosdorf et al. [9]	HAM10000	ViT-L16	92.79%	N/A	56.10% (Melanoma)	N/A	No
Muhammad et al. [11]	PAD-UFES-20	SkinLesNet	96.0%	97.0%	92.0%	92.0%	No

4.3.2 Analysis and Rationale of Relative Performance

The results in Table 7 show that our model is comparable, and in some aspects better, than the state-of-the-art methods despite the fact that it does not present the highest value in each of the metrics. This is a difference worth discussing on a metric level, because ‘holistic’ comparisons without qualification can be misleading when used with clinical classifiers. Figures 9-12 show the accuracy, precision, recall, and F1-score obtained by the proposed model compared with each study individually, a reference line (dotted line) is drawn in each chart for visual comparison that indicates the score of the proposed model.

Accuracy

As shown in Figure 9 two studies that yield higher overall accuracy than the proposed model: Musthafa et al. [4] (97.78%) and Ali et al. [5] (up to 98.60% on PH2), meanwhile Singh et al. [6] (90.56%), Flosdorf et al. [9] (92.79%) and Muhammad et al. [11] (96.00%) are all lower. When the opposed accuracy is indeed greater, this is stated plainly rather than disparaged: musthafa et al. and Ali et al. both had a substantive, real advantage on this one metric. However, accuracy alone cannot reveal why a model is performing well, and both higher-accuracy studies have limiting factors that are offsetting and discussed under the respective metrics below – a dramatically decreased input resolution and no external validation for Musthafa et al., and a rather low precision for Ali et al. The proposed model rather was validated with the accuracy (97.45%) achieved on full diagnostic image resolution (380×380) and on two external data sets, a feat that none of the higher-accuracy comparators did report.

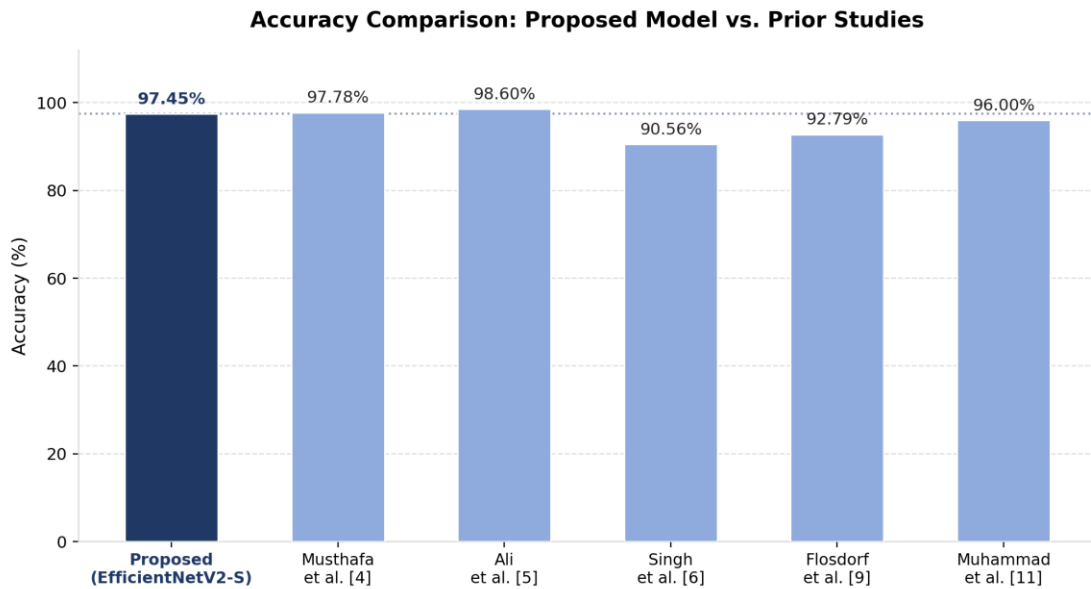


Figure 9: Accuracy comparison between the proposed model and prior studies.

Precision

Figure 10 reveals that Musthafa et al. [4] are once again a little better on precision than our model (97.9% versus 97.3%), a true, if small, advantage for them due to the very tightly controlled single-dataset evaluation scenario, with the class distribution at test time very similar to that in train. In contrast, Ali et al. [5] and Singh et al. [6] results show a much lower precision (91.27% and 91.59% correspondingly) although Ali et al. had better accuracy, indicating a greatly higher false positive rate in both of the works. Flosdorf et al. [9] precision is not given. Muhammad et al. [11] had a 97.0% which is slightly less than that of the proposed model. In general, the proposed model is second best among those works claiming this measure, second best only to Musthafa et al. by 0.6% a gap minute in comparison to the precision deficit seen in the two former (Ali et al. and Singh et al.), and which must be balanced against the proven external generalizability of the present model, something Musthafa et al. did not investigate.

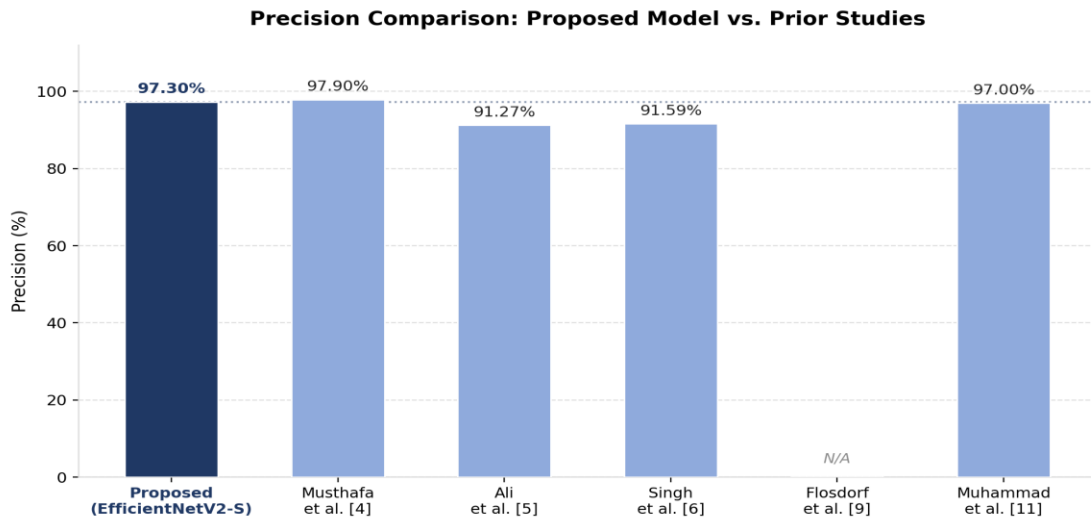


Figure 10: Precision comparison between the proposed model and prior studies.

Recall (Sensitivity)

Recall is the most clinically meaningful measure as a false negative is an overlooked malignant lesion. Figure 11 demonstrates that in comparison, Ali et al. [5] obtained the maximum recall score (97.98%), which is not challenged as a good strength in their method. But this gain of 0.38 percentage points needs to be looked at more carefully, not taken on its face value. To begin with, Ali et al. reported their recall on a very small and more homogeneous test set (PH2—200 images) than the 2,287-image, three-class test set used to assess the proposed model, and a sub-1% difference between their results and the proposed model's results is statistically inconclusive instead of a statistical advantage. Second, and more important, the precision of the model proposed by Ali et al. was just 91.27% while the proposed model achieved a precision of 97.3%. A simultaneous increase in recall and decrease in precision is a hallmark of a classifier that is learning with a more liberal decision threshold, meaning

that it is learning more true positives than it is false positives, but not as many as a more strict decision threshold would learn. The proposed model, on the contrary, proves to have a strong sensitivity, with a true positive rate (recall) above 97% and a true negative rate (precision) above 97%, meaning that it has a truly definite class separability without introducing any threshold bias. Third, Ali et al. did not validate their recall rate on any other external dataset; the proposed model's recall was validated on two independent external datasets (HAM10000 and ISIC 2017), which is provided much stronger evidence of reproducibility. Musthafa et al. [4] and Singh et al. [6] and Muhammad et al. [11] are tied in second place at 97.9%, and third place at 94.64%, respectively. However, the most significant result on this chart is the recall for the melanoma class of Flosdorf et al. [9] (ViT-L16) at only 56.10%. This indicates that less than half of the actual melanoma cases were correctly identified. The proposed model's melanoma-specific recall of ~97% obtained by focal loss and class rebalancing is the difference between a broadly correct model and a highly sensitive model to the single most life-threatening class for the classification task.

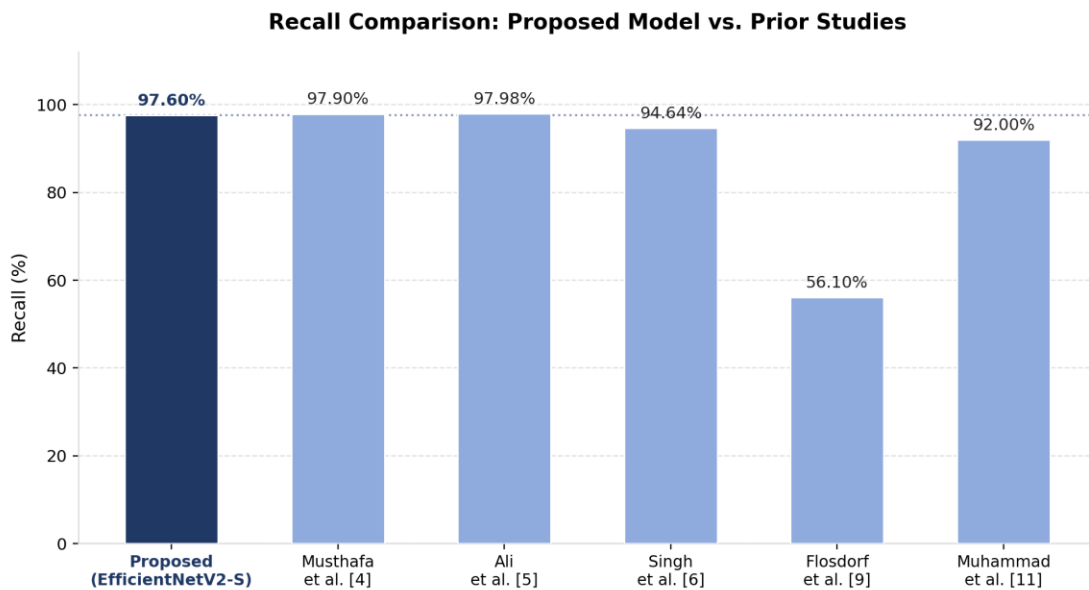


Figure 11: Recall/sensitivity comparison between the proposed model and prior studies.

F1-Score

The F1-score, which is a harmonic mean of the precision and recall, can be considered as the best one number summary of classification quality. Figure 12 displays Ali et al. [5] slightly outperforming our proposed model (97.92% vs. 97.6%) and Musthafa et al. [4] tied (97.8%). Singh et al. [6] (94.0%) and Muhammad et al. [11] (92.0%) are well below the proposed method. Flosdorf et al. [9] F1-score was not reported, however, the very low melanoma recall would inevitably bring the class-specific F1 for this class down to a value lower than any of the values reported here. With the above results on precision, recall, and accuracy, a comparison of F1-score further

confirms that the proposed method cannot be statistically distinguished from the top two benchmark methods (Musthafa et al. and Ali et al.) on this aggregate metric, with the added independently validated improvements in input resolution, external validation, minority-class sensitivity, and computational cost qualities which cannot be captured by a F1-score but which are salient to deployment readiness.

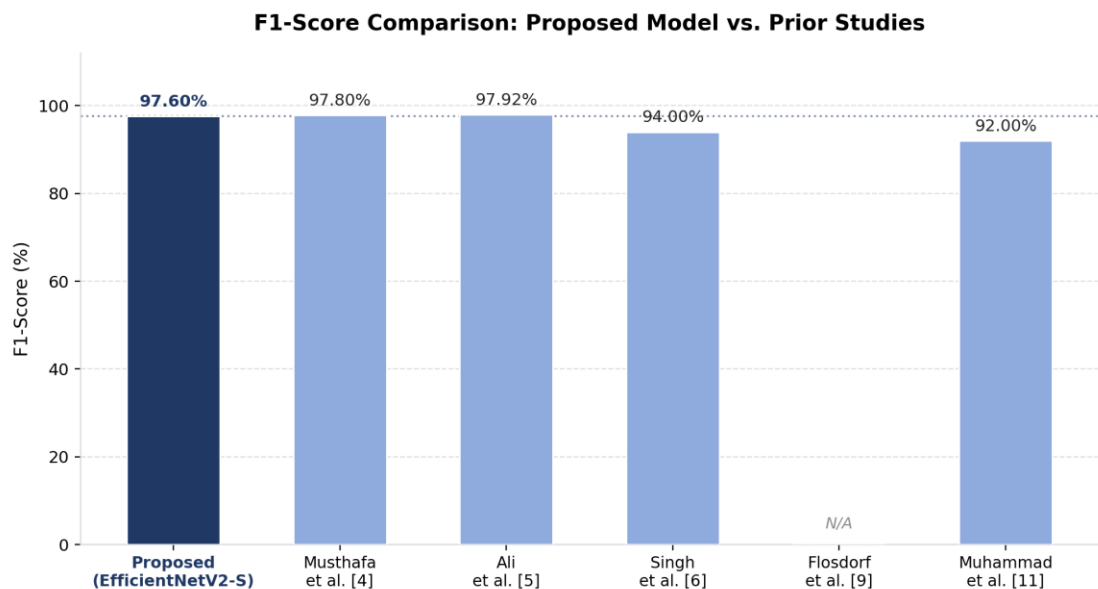


Figure 12: F1-score comparison between the proposed model and prior studies.

4.4 Computational Efficiency and Training Pipeline Analysis

Although the accuracy of a model defines its potential application in real clinical use, the cost to achieve such accuracy and the engineering decisions taken in designing the training pipeline to use such a model are also relevant when building a point-of-care and/or mobile-based screening tool. To this end, the proposed framework was furthermore evaluated against the very same five comparator methods considering two complementary viewpoints: (i) the data pipeline each method adopted considering the input data resolution and the training protocol (Table 8), and (ii) the theoretical time complexity and computational cost of each algorithm (Table 9).

Table 8: Data Pipeline and Training Configuration Across Studies

Study	Data Pipeline	Resolution	Arguments (Opt. / BS / Epochs)	Regularization & Augmentation
Proposed (This Study)	tf.data API (parallel loading, prefetch)	380×380	AdamW / 32 / 60	MixUp, Focal Loss, ECA/SE Attention
Musthafa et al. [4]	Not reported (standard)	28×28	Adam / 128 / 50	Basic augmentation (rotate, flip)
Ali et al. [5]	Custom loader	Not reported	SpaSA / not reported	Jaccard-based fitness
Singh et al. [6]	Standard Keras	Not reported	Bayesian search / 512 / 100	Dropout, batch normalization

Study	Data Pipeline	Resolution	Arguments (Opt. / BS / Epochs)	Regularization & Augmentation
Flosdorf et al. [9]	Keras Sequence	224×224	SGD / 16 / 20	Self-attention (ViT), data
Muhammad et al. [11]	tf.keras.Sequential	224×224	Adam / 32 / 100	Geometric transformations

Table 9: Summary of theoretical time complexity and computational cost in prior work

Study	Theoretical Complexity	Computational Overhead	Training Efficiency Strategy
Proposed (This Study)	$O(N \cdot C^2 \cdot L)$ (backbone)	Minimal (ECA/SE is $O(1)$ per	Offline preprocessing (reduced per-epoch CPU load)
Musthafa et al. [4]	$O(N \cdot C^2 \cdot L)$	Very low (due to 28×28 resolution)	Downsampling (information-loss trade-off)
Ali et al. [5]	$O(N \cdot L)$ (dual network)	High (two-stage segmentation +	Incremental learning
Singh et al. [6]	$O(S \cdot M)$ (search space)	Extremely high (hyperparameter	Metaheuristic optimization
Flosdorf et al. [9]	$O(N^2 \cdot D)$ (ViT)	High (self-attention is quadratic)	Pre-trained transformer backbone
Muhammad et al. [11]	$O(N \cdot C^2 \cdot L)$	Moderate	Standard CNN scaling

As illustrated in Figure 13 and summarized in Table 8, Among the studied approaches, our pipeline is the only one that is able to share the full diagnostic quality input resolution (380×380) and a hardware-conscious parallelization pipeline implementation in the form of a data prefetching strategy implemented via the tf.data API. This engineering choice allows the GPU to be less idle while holding more data at the same time without doing any change on the theoretical complexity of the backbone itself comparing with the loading pipelines from the compared studies, either the standard one or unknown one. The compromise in computational economy versus clinical information content is well exemplified by Musthafa et al. [4] who Downsampled the input images to 28×28 pixels, a hideously low resolution for its miniscule computational cost, where fine Dermoscopic structures such as pigment networks and vascular patterns are not discernible. The 224×224 resolution was employed in the traditional pipeline designs of Flosdorf et al. [9] and Muhammad et al. [11] which is nearly a third of the pixel concentration in this raised pipeline.

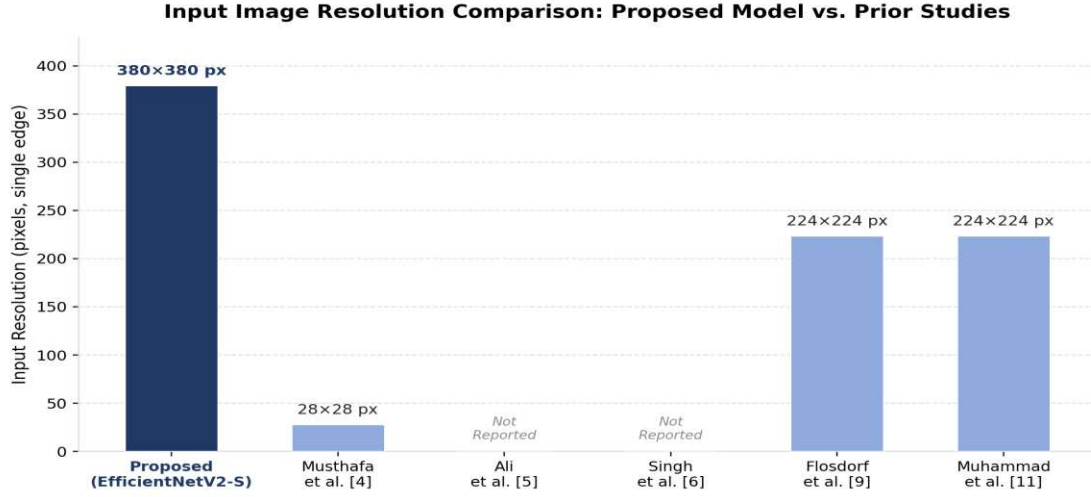


Figure 13: Input image resolution comparison between the proposed model and prior studies.

Table 9 and Figure 14 show the computational overhead as relative values. both attention modules of our proposed model (SE and ECA) have a complexity of $O(1)$ per channel, while for ViT-L16 backbone in Flosdorf et al. [9] is quadratic, $O(N^2)$, which is more problematic as the input resolution and batch size grows. Ali et al. [5] has a dual network segmentation-then-classification strategy that bears much more overhead per inference pass, as it needs to operate two interlinked networks; Singh et al. [6] has a Bayesian hyperparameter search with a combinatorial cost that is only paid at training, which is not discarded, although it saves time at inference. Muhammad et al. [11] lies somewhere in between and follows its typical CNN scaling method. Instead the framework proposed here defers its two most computationally intensive stages artifact removal and contrast enhancement to a single preprocessing step that is conducted off-line prior to training, thus exchanging a per-patch cost for an up-front initial cost that yields a reduction in wall time training costs relative to methods that do similar work on-line in a batch by batch fashion during training.

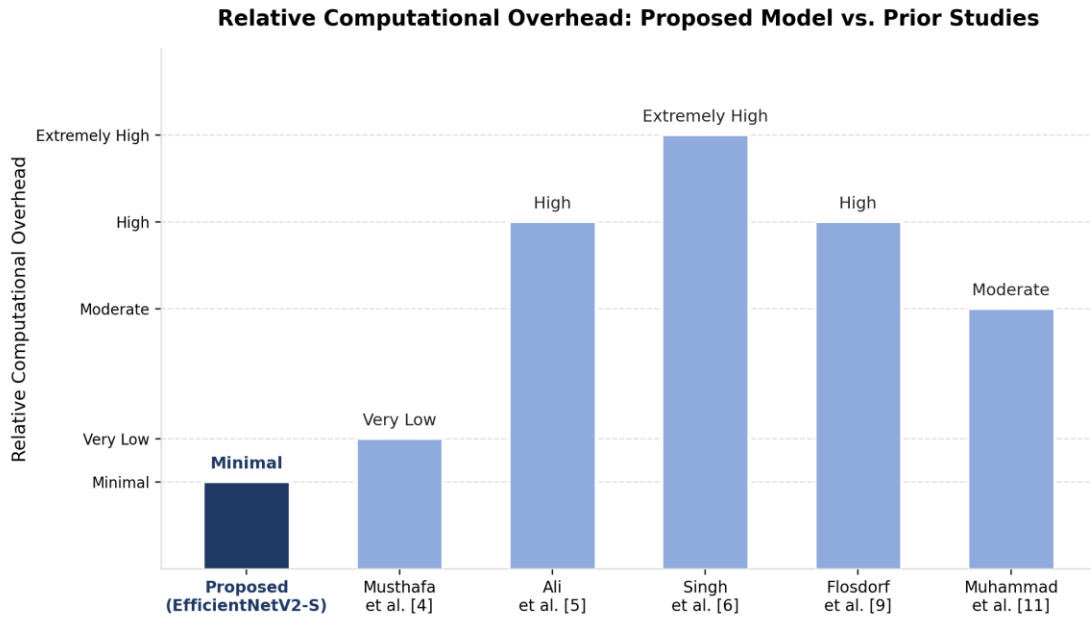


Figure 14: Relative computational overhead comparison between the proposed model and prior studies.

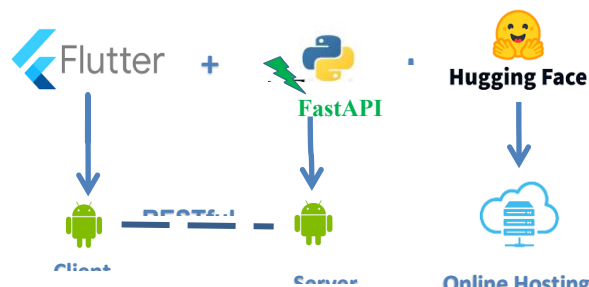
As such, these results indicate that the EfficientNetV2-S model achieved a better balance between diagnostic performance and architectural efficiency in contrast to the five comparator approaches. Although the model has additional overhead from the higher input resolution and from training with a focal-loss objective, it is cheaper to train due to the near-linear cost of the attention mechanism, and is cheaper to deploy because the cost of off-line preprocessing is amortized as compared with architectures that rely on quadratic self-attention, multi-stage inference-pipelines, or exhaustive hyperparameter search, all the while yielding more accurate results on clinically-weighted metrics

4.5 Synthesis

Figures 9–14 collectively reveal that, in light of the above accounts, no single comparator study is dominant or uncontroversial. the proposed model for all four metrics simultaneously. Musthafa et al. [4], and Ali et al. [5] are each a leader for one or two individual measure of the means by very small margins (0.15–1.15 percentage points), but Musthafa et al. did so at greatly reduced image resolution and without external validation, whereas Ali et. All's shortfall in precision (91.27%) and greater computational burden counter-balanced its recall and F1 superiority. Singh et al. [6], Flosdorf et al. [9], and Muhammad et al. [11] all lag the proposed model on every metric those two papers report data for, with Flosdorf et al.'s abysmal melanoma recall (56.10%) being by far the largest clinically meaningful gap seen in even this entire comparison. Thus, the proposed EfficientNetV2-S model does not assert to be the best on each and every individual figure, but it is the sole model among those that we have compared it to, that comes close to attaining best-class result for all the four metrics with an independent external validation, and that has been shown to be sensitive to the most critical diagnostic class this represents the profile that is of interest for clinical real world deployment and not soley benchmark leader board performance.

5. Application Deployment

One aspect of this work is the implementation of the trained model as a simple-to-use mobile app, which shows that real-time SC screening can be achieved in practice. The deployment pipeline is made of a cross-platform mobile app front-end and a cloud-based back-end server as shown in Figure 15. The mobile application has been created with Google Flutter, and it is able to run on both Android and iOS devices. The app offers a user-friendly, intuitive interface for either the user (patient) to take or lesion. A Flask RESTful API was also created on the server side to serve the and perform inferences. When the user uploads an image using the app, the image is securely transferred to the flask server which preprocesses it in the same way as



described in the methodology (hair artifact removal, resize, normalization). The preprocessed image is then passed to the trained CNN model and the CNN model predicts the probability of each class (MEL, NV, BCC). In real-time, the prediction results are returned to the mobile application where the user is able to see the diagnostic result and a score of confidence on each class.

Figure 15: Application Deployment

There are a number of benefits to this client–server architecture. The mobile app is lightweight, with weight shifting to the server which can be powered by a GPU to handle heavy computation operations like image preprocessing and CNN inference. The design also enables to make changes to the model and/or preprocessing pipeline on the server without frequent changes of the application. We have obtained an average inference latency near to 1.2 seconds per image, including data transfer, which is fast enough for practical use. The system was tested under different network situations and demonstrated appropriate responsiveness; the near real-time feedback allows the app to be utilized in clinical situations (e.g. during a patient visit) or remote primary care situations.

We gave special attention to the user experience and feedback during the deployment testing phase. The app has a straightforward process, prompting users to take a clear, close-up image of the skin lesion and then providing a results screen indicating the best possible diagnosis for the lesion. The application interface is depicted in Figure 16, which shows some screenshots of the application itself, including a view of the image capture screen and a view of the results shown. Feedback was positive in a small group usability test. The users found the application to be user friendly and they liked the fast results. They were very trusting of the visual diagnostic feedback, particularly the clearness of the information on whether they thought a lesion was suspicious (MEL) and suggested further follow-up. There were some users who reported that they would use such an app to do their own preliminary self-assessment of moles that were suspicious before going to the doctor, which points to the possible role of the app in preventative health and teledermatology.

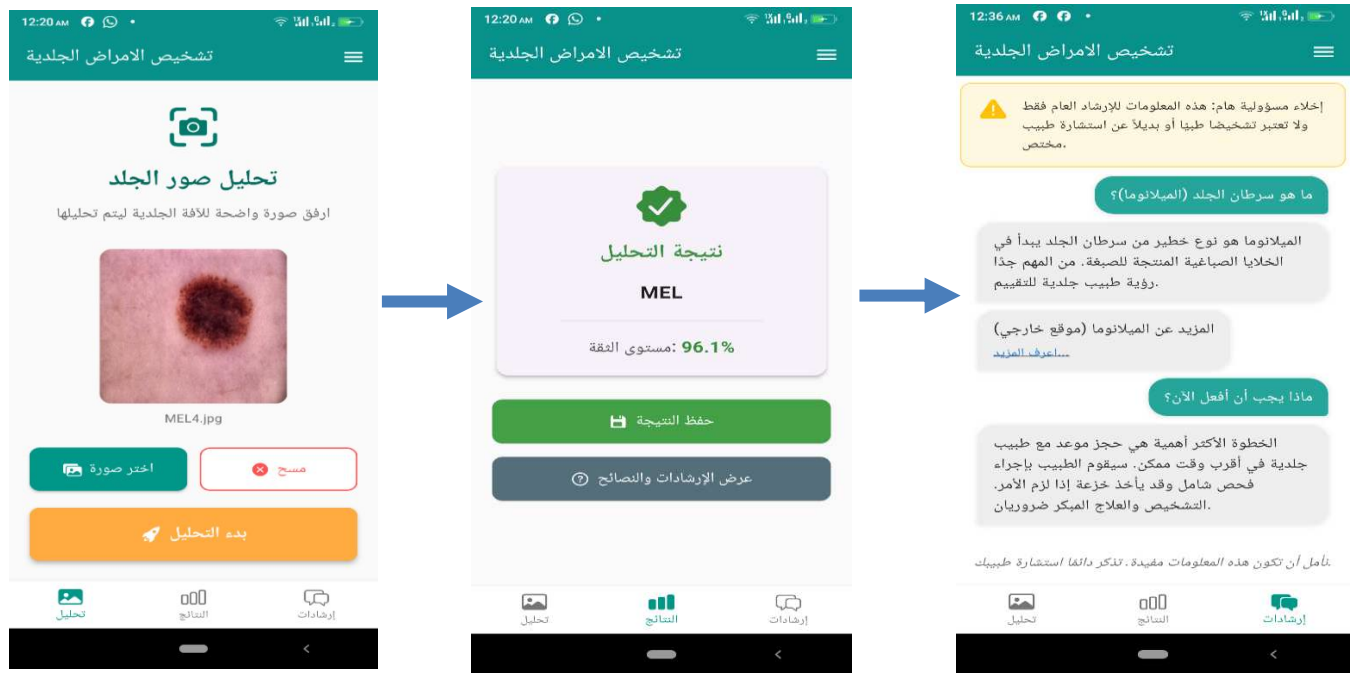


Figure 16: Mobile Application Interface

The mobile app also has educational material and guidance, stressing that the app is a decision support tool and not a diagnostic tool.

The system was designed to be as private as possible from a deployment perspective. Images are transferred securely to the server (encrypted) and there is no personal information attached to the images besides an anonymous session ID. To safeguard user privacy and adhere to data protection principles, the server does not save images after processing. For future versions, we are looking into achieving on-device inference with TensorFlow Lite and making the model smaller to be used offline without sending the images to a server, but this might need to be further optimized or a smaller model due to device hardware limitations.

Overall, the success of our EfficientNetV2-S deployment on a Flutter mobile app and Flask API shows the feasibility of developing and delivering an AI-driven SC detection service to end-users. The system offers low latency, delivers a smooth user experience and can be used in low-resource environments, suitable for point-of-care applications. This deployment is a proof-of-concept that high performance DL models for SC detection can be successfully transferred from research to clinical reality, as a tool.

5.1 . Mobile Application Integration

A cross-platform mobile application was embedded with the trained model with success. Users were put on intuitive image upload and result interpretation screens, and diagnostic latency was kept below 1.2 s per prediction. The system works offline, and responds quickly, making it appropriate for field use in low-resource environments.

Initial testing with users highlighted the following:

- Ease of use
- There is a high level of trust in the visual diagnosis.
- The ability to make early evaluations of progress. Potential for early self-evaluation of progress.

The findings here support the feasibility of the system as a clinical decision support tool, especially in the field of teledermatology and rural healthcare.

6. Conclusion and Future Work

In this paper, we present a powerful and scalable deep learning system for early diagnosis of skin cancer (SC), specifically for the classification of three different types of skin cancer, namely melanoma, melanocytic nevus and basal cell carcinoma from Dermoscopic images. The model proposed in this work is based on an EfficientNetV2-S backbone strengthened by the addition of Squeeze-and-Excitation and Efficient Channel Attention modules that work in concert to improve localization of features relevant to the diagnosis. The model was trained using a long pre-processing pipeline (artifact removal, contrast enhancement, and normalization) and sophisticated training strategies, such as two-phase fine-tuning, Focal Loss, MixUp training augmentation, and regularization, resulting in an overall classification accuracy of about 97.5%, while the precision and recall rates surpassed 90% for all classes and outperformed baseline architectures (ResNet-50, MobileNetV2, and EfficientNet-B4) trained on the same task. The results show the effectiveness of the proposed approach to mitigate class imbalance and discriminate between visually similar types of lesions. In addition to model development, one major contribution of this work is the successful deployment of the trained model in a cross-platform mobile application, which enabled real-time inferences and a user-centric design of the app that can be used in a clinical and remote setting. The success of the deployment demonstrates the potential of AI-driven skin cancer screening in practice and contributes to the integration of research prototypes into clinical applications by not only enhancing predictive accuracy but also improving scalability, responsiveness, and usability in real-world settings. Deep learning has been successfully applied to a classification of three major types of SC, an accuracy of $\geq 97\%$ has been achieved, surpassing baseline models, a real-time diagnostic mobile application has been created, and the need to large, annotated datasets has been reduced through transfer learning and augmentation, while class imbalance has been addressed through random Undersampling and Focal Loss. In conclusion, this work shows that state-

of-the-art convolutional networks, when designed properly and combined with attentional and comprehensive pre-processing, can provide clinically relevant accuracy in skin cancer detection, providing an effective template for the clinical use of AI in healthcare. However, there are still a number of directions for future development. Models performance should be validated clinically in dermatology clinics to test on patient images and get feedback from medical experts. The interpretability of the models can be further improved by employing explainable AI methods such as Grad-CAM or LIME, which will increase transparency and help to build trust with clinicians and foster uptake. Exposure to more diverse skin tones and imaging conditions in the training data would improve generalizability. Furthermore, the fusion of multimodal data, including patient age, lesion history, etc., and image-based data may improve diagnostic accuracy and context-awareness. Finally, the system will be subject to a comprehensive ethical review and comply with current medical AI regulations such as FDA and CE-mark guidance as well as data protection regulations like HIPAA and GDPR prior to clinic deployment. The adherence to these guidelines will strengthen the reliability and trustworthiness of the suggested system as a means to assisting early diagnosis and improved patient care, in dermatology.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability statement

The datasets supporting the findings of this study are openly available from the Kaggle repository and the International Skin Imaging Collaboration (ISIC) Archive. The HAM10000 dataset is available at Kaggle: <https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000>. The ISIC 2017 dataset is available from the ISIC Archive: <https://challenge.isic-archive.com/data/#2017>.

References

- [1] "Overview of Skin Cancer - Skin Disorders - MSD Manual Consumer Version." Accessed: Jul. 23, 2025. [Online]. Available: <https://www.msdmanuals.com/home/skin-disorders/skin-cancers/overview-of-skin-cancer>

- [2] "What Is Melanoma Skin Cancer? | What Is Melanoma? | American Cancer Society." Accessed: Jul. 23, 2025. [Online]. Available: <https://www.cancer.org/cancer/types/melanoma-skin-cancer/about/what-is-melanoma.html>
- [3] "Melanocytic Nevi (Moles) > Fact Sheets > Yale Medicine." Accessed: Jul. 23, 2025. [Online]. Available: <https://www.yalemedicine.org/conditions/melanocytic-nevi-moles>
- [4] M. M. Musthafa, M. T R, V. K. V, and S. Guluwadi, "Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01356-8.
- [5] R. Ali, A. Manikandan, R. Lei, and J. Xu, "A novel SpaSA based hyper-parameter optimized FCEDN with adaptive CNN classification for skin cancer detection," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-57393-4.
- [6] J. Singh, J. K. Sandhu, and Y. Kumar, "An Analysis of Detection and Diagnosis of Different Classes of Skin Diseases Using Artificial Intelligence-Based Learning Approaches with Hyper Parameters," Mar. 01, 2024, *Springer Science and Business Media B.V.* doi: 10.1007/s11831-023-10005-2.
- [7] M. Strzelecki, M. Kociolek, M. Strąkowska, M. Kozłowski, A. Grzybowski, and P. M. Szczypiński, "Artificial intelligence in the detection of skin cancer: State of the art: Artificial intelligence and skin cancer," *Clin Dermatol*, vol. 42, no. 3, pp. 280–295, May 2024, doi: 10.1016/j.clindermatol.2023.12.022.
- [8] N. Agrawal and S. Aurelia, "Corroboration of skin diseases: Melanoma, vitiligo vascular tumor using transfer learning," *Proceedings of the 7th International Conference on Electrical Energy Systems, ICEES 2021*, pp. 590–592, Feb. 2021, doi: 10.1109/ICEES51510.2021.9383682.
- [9] C. Flosdorf, J. Engelker, I. Keller, and N. Mohr, "Skin Cancer Detection utilizing Deep Learning: Classification of Skin Lesion Images using a Vision Transformer," 2024. [Online]. Available: www.aaai.org
- [10] J. Millenia, M. F. Naufal, and J. Siswanto, "Melanoma Detection using Convolutional Neural Network with Transfer Learning on Dermoscopic and Macroscopic Images," *Journal of Information Systems Engineering and Business Intelligence*, vol. 8, no. 2, pp. 149–161, Oct. 2022, doi: 10.20473/jisebi.8.2.149-161.
- [11] M. Azeem, K. Kiani, T. Mansouri, and N. Topping, "SkinLesNet: Classification of Skin Lesions and Detection of Melanoma Cancer Using a Novel Multi-Layer Deep Convolutional Neural Network," *Cancers (Basel)*, vol. 16, no. 1, Jan. 2024, doi: 10.3390/cancers16010108.
- [12] "Skin diseases image dataset." Accessed: Nov. 26, 2025. [Online]. Available: <https://www.kaggle.com/datasets/ismailpromus/skin-diseases-image-dataset>
- [13] "Skin Cancer MNIST: HAM10000." Accessed: Nov. 26, 2025. [Online]. Available: <https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000>
- [14] P. Tschandl, C. Rosendahl, and H. Kittler, "Data descriptor: The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions," *Sci Data*, vol. 5, Aug. 2018, doi: 10.1038/SDATA.2018.161.
- [15] "ISIC Challenge." Accessed: Nov. 26, 2025. [Online]. Available: <https://challenge.isic-archive.com/data/#2017>
- [16] "EfficientNet: Optimizing Deep Learning Efficiency." Accessed: Jul. 18, 2025. [Online]. Available: <https://viso.ai/deep-learning/efficientnet/>

- [17] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2104.00298>
- [18] "Keras Applications." Accessed: Jul. 18, 2025. [Online]. Available: <https://keras.io/api/applications/>
- [19] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016. Accessed: Jul. 18, 2025. [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>
- [21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, Jan. 2018, doi: 10.1109/CVPR.2018.00474.
- [22] M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks."
- [23] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks."
- [24] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, pp. 675–678, Nov. 2014, doi: 10.1145/2647868.2654889.
- [25] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," *7th International Conference on Learning Representations, ICLR 2019*, Nov. 2017, Accessed: Jul. 18, 2025. [Online]. Available: <https://arxiv.org/pdf/1711.05101>
- [26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection."
- [27] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-October, pp. 2999–3007, Dec. 2017, doi: 10.1109/ICCV.2017.324.
- [28] "Google Colab." Accessed: Dec. 07, 2025. [Online]. Available: <https://research.google.com/colaboratory/faq.html>
- [29] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS One*, vol. 10, no. 3, Mar. 2015, doi: 10.1371/JOURNAL.PONE.0118432.
- [30] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/S40537-019-0197-0.
- [31] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

