Multi-Attention and Ensemble Learning for Precise Dermoscopic Diagnosis

Muhammad Amir Khan
School of Computing Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
amirkhan@uitm.edu.my (corresponding author)

Razia Manan
Faculty of Arts, Humanities and Linguistics, IIC University of Technology, Phnom Penh, Cambodia
raziaamir2007@gmail.com

Ali Khalid
School of Computing Sciences, Faculty of Computer Science and Mathematics, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
2024184591@student.uitm.edu.my

Mohammad Shahid
Department of Computer Science, IIC University of Technology, Phnom Penh, Cambodia
shahidmaddni@gmail.com

Umar Farooq Khattak
School of Information Technology, UNITAR International University, Kelana Jaya, Petaling Jaya, Malaysia
umar.farooq@unitar.my (corresponding author)

Received: 18 October 2025 | Revised: 14 November 2025 and 28 November 2025 | Accepted: 29 November 2025

 DOI: https://doi.org/10.48084/etasr.15635

ABSTRACT

Skin cancer has high incidence rates, and early diagnosis is crucial to improving survival rates. Thus, efficient diagnostic tools are imperative. This study suggests a new way to classify skin cancer by integrating a multi-attention Convolutional Neural Network (CNN) with ensemble learning for increased diagnostic precision. Utilizing the HAM10000 dataset with 10,015 dermoscopic images of seven different lesions, this study addresses significant issues such as class imbalance, noise, and inadequate feature extraction. Advanced preprocessing techniques, including contrast correction, noise removal, and hair removal, were used to standardize images and eliminate artifacts. The multi-attention mechanism enables the CNN to weight significant lesion features, such as asymmetry, irregular margins, and color variation, which are critical to distinguishing malignant from benign lesions. In addition, ensemble learning additionally ensures model stability by ensembling multiple classifiers to reduce prediction bias. Experimental results demonstrate that the proposed model compares favorably with standard deep learning techniques in terms of accuracy and sensitivity in distinguishing melanoma and non-melanoma lesions. By facilitating non-invasive and precise diagnosis, this technique has the potential to assist clinicians in early diagnosis, thereby increasing survival rates and reducing treatment costs. The proposed automated skin cancer diagnosis system holds potential for global healthcare systems, but its clinical deployment is currently limited by high computational demands and reduced generalizability to non-dermoscopic images.

Keywords-skin cancer; deep learning; convolutional neural network; multi-attention mechanism; melanoma; ensemble learning; dermoscopic images; HAM10000 dataset; early diagnosis; class imbalance

I. INTRODUCTION

Skin cancer, and melanoma in particular, poses a significant global health burden considering its ability to advance rapidly if not diagnosed early. In 2020, there were 247 skin cancer deaths in Malaysia as reported by the World Health Organization, demonstrating the urgent need for accurate diagnostic tools [1]. The high mortality of melanoma is due to its likelihood of metastasizing, and therefore, diagnosis at an early stage is critical, with survival rates as high as 99% when diagnosed early [2]. Squamous cell carcinoma and basal cell carcinoma are two other types that add to the disease load and need to be treated early to avoid complications. Traditional diagnostics rely heavily on biopsy, which is invasive, time-consuming, and can potentially postpone treatment. Dermoscopic imaging provides a non-invasive alternative, capturing close-up lesion images for analysis. However, issues such as noise (e.g., hair artifacts, shadows), low contrast, and varied lesion presentations make accurate interpretation difficult [3]. The ABCD criteria—Asymmetry, Border irregularity, Color variation, and Diameter (more than 6 mm)—govern melanoma examination, but manual inspection is often thwarted by inconspicuous features [4]. Additionally, datasets such as HAM10000, which are typically used in skin cancer research, are unbalanced, leading to biased model predictions [5].

Skin cancer Computer-Aided Diagnosis (CAD) has changed as a result of recent developments in deep learning. Convolutional Neural Networks (CNNs) are excellent at processing intricate dermoscopic images, but their effectiveness is hampered by noise and class imbalance [6]. By combining models, ensemble learning can enhance classification robustness, and attention techniques have been developed to help CNNs focus more on pertinent lesion features [7, 8]. Despite these advances, there is still a research gap in integrating various methods to simultaneously address noise, bias, and inefficiency in feature extraction. This study presents a unique architecture that blends ensemble learning and a Multi-Attention CNN (MA-CNN) to improve the accuracy of skin cancer classification. The multiattention module reduces the impact of noise and draws attention to clinically significant features, such as uneven boundaries and color heterogeneity [7]. By combining the predictions of several classifiers, ensemble learning reduces the bias caused by an unbalanced dataset and improves generalization to other types of lesions [8].

The HAM10000 dataset consists of 10,015 dermoscopic images of seven distinct types of lesions: vascular lesions, dermatofibroma, melanoma, melanocytic nevi, benign keratosis-like lesions, actinic keratosis, and basal cell carcinoma. This dataset is popular for training robust diagnostic models due to its diversity in terms of sources and modalities [5]. This study aimed to address three main challenges in skin lesion classification using this dataset. First, class imbalance poses a problem, as the skewed distribution of samples may lead to biased predictions where malignant classes dominate. To address this, class-balancing strategies, such as oversampling and weighted loss functions, are applied to ensure fair performance across all categories. Second, feature extraction is complicated by noise and artifacts that obscure vital diagnostic details. To overcome this, an MA-CNN was designed to enhance focus on key diagnostic markers, thereby improving the differentiation between malignant and benign lesions. Third, model robustness is a concern, as single CNN models often struggle to maintain consistent performance across diverse datasets. This issue was mitigated through ensemble learning, which integrates multiple classifiers to improve reliability and generalization.

The main objectives were to minimize prediction bias through class-balancing methods, to develop an MA-CNN capable of extracting richer diagnostic features from dermoscopic images, and to evaluate the proposed model's performance against existing approaches. Compared to prior work on single-model architectures, this study employed multi-attention and ensemble techniques to address real-world complexities [9]. In addition, unlike prior works that apply attention or ensemble separately [7-14], this study introduces a unified MA-CNN framework that simultaneously integrates CBAM-based multi-attention into three diverse CNN backbones (VGG16, ResNet50, EfficientNetB0) and applies Random Forest (RF) stacking as a meta-learner, achieving synergistic bias reduction and feature focus in a single end-to-end pipeline.

II. RELATED WORKS ON DEEP LEARNING MODELS FOR SKIN CANCER CLASSIFICATION

CNNs use convolutional, pooling, activation, and fully connected layers to extract hierarchical features in medical imaging for detecting edges, boundaries, and complex patterns. Dimensionality reduction by pooling reduces overfitting, while ReLU introduces non-linearity. Class imbalance and limited data are addressed through transfer learning from pre-trained models and augmented data simulating lighting and skin tone variations. The batch normalization, dropout, and attention mechanisms improve stability and focus during training.

A. Multi-Attention Mechanisms in Skin Cancer Diagnosis

Multi-attention mechanisms improve the focus on clinically significant features in dermoscopic images, such as irregular borders and color variations, enhancing the accuracy and interpretability of the classification. CAFNet [15] merged dermoscopic images, clinical images, and metadata through co-attention, achieving 76.8% average accuracy on a seven-point checklist dataset, but was limited by the dataset size and diversity across modalities. In [16], a CBAM-enhanced YOLOv4-tiny with channel and spatial attention was suitable for resource-constrained settings, although generalizability remained underexplored. MSLANet [17] combined three LANets with self-supervised learning, reporting AUCs of 93.7% and 92.4% for the ISIC 2017 and SIIM-ISIC 2020 datasets, respectively, at the cost of high computational complexity that limits real-time applications. In [18], a comparison of five transfer learning models using six attention mechanisms (a total of 105 experiments) showed that spatial/channel attention was the best option, but also computationally demanding. In [19, 20], multi-modal fusion with metadata was investigated, but suffered from heterogeneity among the data. These works demonstrate the

potential of multi-attention approaches, but they emphasize the need for more robust and generalizable models.

B. Ensemble Learning in Skin Cancer Diagnosis

Ensemble learning helps improve diagnostic accuracy and overcome overfitting. In [10], an ensemble of VGG16, CapsNet, and ResUNet reported 86% accuracy on the ISIC dataset, limited by class imbalance. In [11], combining modified CNNs, enhanced segmentation, and artificial multiple intelligence achieved 99.7% detection and 96% classification accuracy, but this approach requires further clinical validation. In [12], an ensemble of Xception, ResNeXt101, GoogLeNet, SeResNeXt101, ResNet152V2, and DenseNet201 achieved 96% ROC-AUC on the HAM10000 and ISIC datasets, but was computationally heavy. In [13], a CNN ensemble was used with an augmented HAM10000 dataset, achieveing 97.85% accuracy, but the contribution of each model was not analyzed thoroughly. In [14], a hybrid Xception-ResNet50 achieved an accuracy of 97.8% on the HAM10000 dataset using both depth and residual connections. However, this model scales poorly on smaller datasets. In [21], adaptive fine-tuned CNNs were used with two-stage transfer learning for melanoma/non-melanoma classification on HAM10000, mitigating overfitting but with high computational cost. In [22], a meta-learning ensemble of Inception, ResNet50, and DenseNet121 achieved 90% accuracy on BUSI, which was limited by dataset imbalance. In [23], a VGG-NIN hybrid for binary skin cancer classification achieved 90% accuracy on HAM10000 with fewer parameters, but was sensitive to image quality. These works show a balance of strengths through ensemble methods but raise concerns about balancing complexity with clinical use, especially as attention mechanisms come into play.

C. Transfer Learning in CNNs

CNNs perform exceptionally well in medical image analysis by learning hierarchical features taken from raw images. Despite this, ResNet and DenseNet, along with Inception, have become the go-to architectures for skin cancer diagnosis due to their very deep structures that enable the extraction of intricate patterns. However, this largely increases the risk of overfitting when dealing with small datasets, a problem that is greatly mitigated by transfer learning. In [24], Inception-v3 was trained on 129,450 images over 2,032 disorders in 747 classes, reporting very high performance that outperformed dermatologists in taxonomic classification. In [25], AlexNet with transfer learning was applied on Dermofit, reporting an accuracy of 81.8% over 10 classes by replacing fully connected layers. In [26], MobileNet was optimized on HAM10000, reaching an accuracy of 91% across seven classes, whereas in [27], MobileNetV2 was used in a mobile app and attained an accuracy of 93.9%. In [28], DenseNet169 was employed on low-resolution images (64×64) in grayscale on HAM10000, achieving an accuracy of 78.56%, while in [14], ResNet50 and Xception were fused together, achieving an accuracy of 97.8% due to enhanced feature extraction. In [29], the pre-trained EfficientNet was adapted to DeepMelaNet, employing ImageNet normalization in the form and achieving 93.40% accuracy on 10,000 images, reducing computational cost through fine-tuning.

III. DATASET DESCRIPTION

The HAM10000 dataset, or the 'Human Against Machine with 10000 training images' dataset, is a large dermatoscopic image dataset aimed at facilitating research for automatic classification of skin lesions, specifically in the detection of skin cancer [30, 31]. The collection comes from institutions in Vienna, Austria, and Queensland, Australia. All images have histopathologically confirmed diagnoses. The dataset consists of seven classes: Melanoma (MEL), Basal Cell Carcinoma (BCC), Actinic Keratosis (AKIEC), Squamous Cell Carcinoma (SCC), Benign Keratosis (BKL), Melanocytic Nevi (NV, ~67% of data), and Vascular Lesions (VASC, least frequent). Images are RGB, usually 600×450 pixels, and are diagnostic, covering features such as asymmetry, border, color, and texture. Metadata includes lesion type, age, gender, and location, although some values are missing. This dataset serves as the gold standard for multi-class classification, segmentation, and transfer learning in medical imaging. However, much care has to be taken to overcome challenges such as class imbalance and variability in acquisition conditions, such as illumination and hardware.

IV. THE PROPOSED MODEL

This study used an MA-CNN approach and ensemble learning to make the final prediction on skin lesion classification. Multi-attention enhances feature representation and diagnostic accuracy by allowing the network to focus on regions of the input images that hold knowledge. Numerous implementations of the proposed CNN model were trained using different initializations and hyperparameters. Ensemble learning methods such as stacking were used to aggregate the predictions of these models and use their diversity for enhanced diagnostic performance. The MA-CNN learns the feature representations and part proposals on each part at the same time. It accepts entire images as input and generates multiple part suggestions, combining convolution, channel grouping, and part classification sub-networks. The proposed method was evaluated using standard performance metrics, including F1-score, confusion matrix, sensitivity, specificity, and accuracy.

A. Data Preparation

HAM10000 has a .csv file with columns of the following elements: lesion_id, image_id, dx, dx_type, age, sex, and localization. It also includes two folders of images. New data frames were created to map lesion IDs to image paths, and such images were then converted into pixel arrays. The NV class contained an overwhelming 6,705 images. To avoid overfitting and biased predictions, the classes needed to be balanced by oversampling. Oversampling was preferred over SMOTE to preserve original image fidelity; focal loss was also tested but reduced convergence speed. Data augmentation (random rotations ≤20°, flips, brightness ±15%) was applied only to training splits to prevent leakage and overfitting, which was validated by stable validation loss. Data augmentation was performed using the Keras ImageDataGenerator. The dataset was then split into 80:20 training and testing after balancing, and would undergo further data augmentation during training.

B. Model Development

This study employed ResNet50, VGG16, and EfficientNetB0 CNN models pre-trained on the ImageNet dataset. These models were combined with two attention mechanisms, spatial attention and channel attention, to improve performance on skin cancer classification. These two processes, referred to as Convolutional Block Attention Modules (CBAMs), were positioned between the convolutional layer in every CNN model. The Channel Attention Module identifies which feature channels are most informative. For a given feature map F $\in R^{C\times H\times W}$, where C is the number of channels, H is the height, and W is the width, the module computes attention weights as follows:

- Global Pooling: Apply global average pooling and max pooling across spatial dimensions to obtain channel-wise descriptors:

  $$F^c_{\text{avg}} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F(c,i,j), \quad F^c_{\max} = \max_{i,j} F(c,i,j)$$

  where F^c_{avg} and $F^c_{\max}$ max are the average and max-pooled features for channel c.

- Shared MLP: Process both descriptors through a shared Multi-Layer Perceptron (MLP) with a reduction ratio r to reduce computational complexity:

  $$M_c(F) = \sigma\left(W_1\left(W_0\left(F_{\text{avg}}\right)\right) + W_1\left(W_0(F_{\max})\right)\right)$$

  where $W_0 \in R^{C/r\times C}$, $W_1 \in R^{C\times C/r}$, and σ is the sigmoid activation function. The output $M_c(F) \in R^C$ represents channel attention weights.

- Feature Refinement: Multiply the attention weights by the original feature map:

  $$F' = M_c(F) \otimes F$$

  where $\otimes$ denotes element-wise multiplication, enhancing informative channels.

The Spatial Attention Module focuses on spatially significant regions within the feature map. For the refined feature map $F' \in R^{C\times H\times W}$, the process is:

- Channel Pooling: Apply average and max pooling along the channel dimension to generate spatial descriptors:

  $$F^s_{\text{avg}} = \frac{1}{C}\sum_{c=1}^{C} F'(c,i,j),\ F^s_{\max} = \max_{c} F'(c,i,j)$$

  Concatenate these to form a descriptor $\left[F^s_{\text{avg}}, F^s_{\max}\right] \in R^{2\times H\times W}$.

- Convolutional Layer: Apply a 7×7 convolutional layer to generate a spatial attention map:

  $$M_s(F') = \sigma\left(\text{Conv}_{7\times 7}\left(\left[F^s_{\text{avg}}, F^s_{\max}\right]\right)\right)$$

  where $M_s(F') \in R^{H\times W}$ is the spatial attention map. CBAM is inserted after each residual block in ResNet50 (stages 2–4), after every 3×3 conv in VGG16, and after each MBConv in EfficientNetB0. Output probabilities from the three MA-CNNs are concatenated, $p = \left[p_{VGG}, p_{Res}, p_{Eff}\right] \in R^{3\times 7}$, and fed to an RF meta-learner (n_estimators=100, max_depth=10) trained on soft predictions

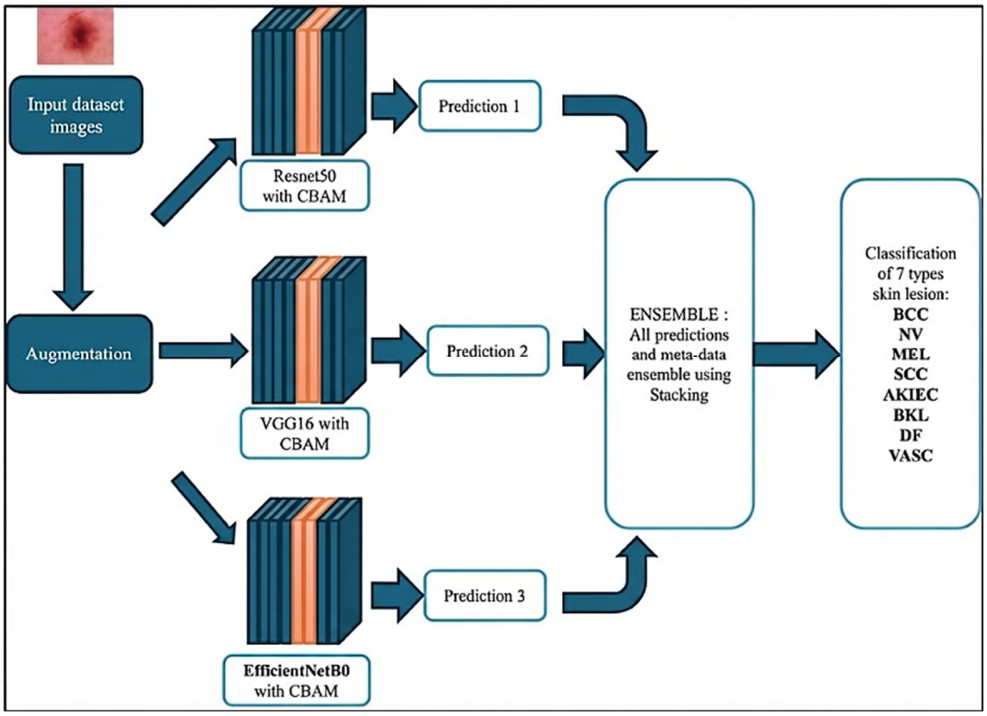

Fig. 1. MA-CNN ensemble: CBAM inserted post-conv blocks in VGG16/ResNet50/EfficientNetB0; softmax outputs are stacked and classified by an RF meta-learner.

C. MA-CNN with Ensemble Learning

To improve the prediction accuracy, state-of-the-art CNNs were combined with channel and spatial attention, CBAM, and ensembled with stacking. The system outputs image predictions with their true classes and diagnoses using Gradio. The system was implemented in Google Colab. The ResNet50 backbone consists of 50 layers, with five stages, each with residual connections between them. CBAM modules are inserted after each convolutional block in stages 2 through 4 to enhance feature extraction. The network takes an input image of size $224 \times 224 \times 3$, first convolving it with a 7×7 kernel along the spatial dimensions with 64 filters, and a stride of 2 in Stage 1. In Stage 2, three residual blocks with the same structure of $[1 \times 1\ (64), 3 \times 3\ (64), 1 \times 1\ (256)]$ are followed by a CBAM module. Stage 3 has four blocks with a structure of $[1 \times 1\ (128), 3 \times 3\ (128), 1 \times 1\ (512)]$ followed by CBAM, while Stage 4 consists of six blocks with the structure of $[1 \times 1\ (256), 3 \times 3\ (256), 1 \times 1\ (1024)]$ and a CBAM. Stage 5 consists of three blocks having a structure of $[1 \times 1\ (512), 3 \times 3\ (512), 1 \times 1\ (2048)]$. Finally, global average pooling reduces the dimensions, followed by a fully connected layer and softmax over seven classes of HAM10000 skin lesions. Figure 1 shows an overview of the proposed architecture.

V. EXPERIMENTAL RESULTS

This study utilized CNNs in conjunction with multi-attention mechanisms and ensemble learning for the classification of skin lesions from the HAM10000 dataset, deployed on an A100 GPU for stable performance. The parameters were tuned to ensure stable model performance. Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$), ReduceLROnPlateau scheduler ($factor = 0.5$, $patience = 3$), and ImageNet-pretrained weights were used across all backbones. An image size of (224,224) was optimal, as it was compatible with the pre-trained models and maintained low computational loads, compared to low-resolution options (128,128) or computationally demanding (512,512) ones. A batch size of 32 was selected as a compromise between avoiding the noise of 16 and the excessive memory usage of 64. Early stopping at 15 epochs prevented the underfitting occurring with 5 epochs and overfitting with 25 epochs. A learning rate of 0.00001 supported the multi-attention mechanism, yielding stable convergence and enhanced model accuracy.

A. Class Imbalance Handling

The HAM10000 dataset was extremely class-imbalanced, with the NV class dominating the others, leading to biased predictions when trained without balancing. The models trained on the imbalanced dataset had 66.94% accuracy for VGG16, 70.74% for ResNet50, and 81.07% for EfficientNetB0, with extremely high mismatch predictions ranging from 769 to 887. To address this, oversampling, undersampling, and data augmentation were performed, leading to a tremendous performance improvement. Balanced datasets improved accuracies to 93.72% for VGG16, 94.71% for ResNet50, and 94.10% for EfficientNetB0, with mismatch predictions reduced to 176–489. These techniques enhanced model fairness, with predictions becoming more accurate for all classes and less biased towards majority classes. Figures 2 and 3 show a comparison of ResNet50 on the imbalanced and balanced datasets, respectively.

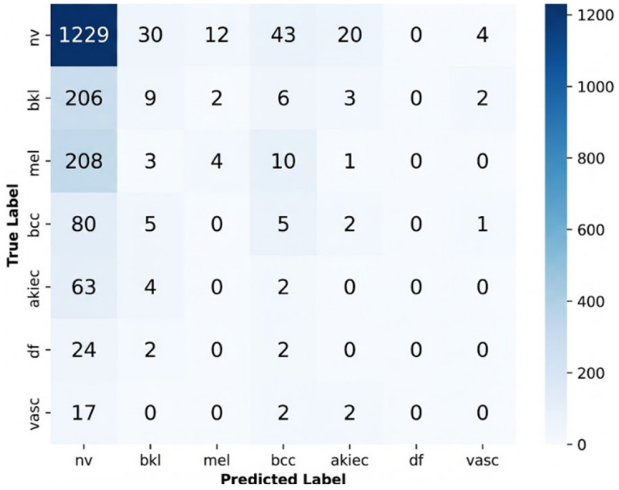

Fig. 2. Confusion matrix of ResNet50 on imbalanced HAM10000 (80:20 split), sorted by class frequency. Rows: true classes (MEL=melanoma, NV=nevi, etc.); columns: predicted.

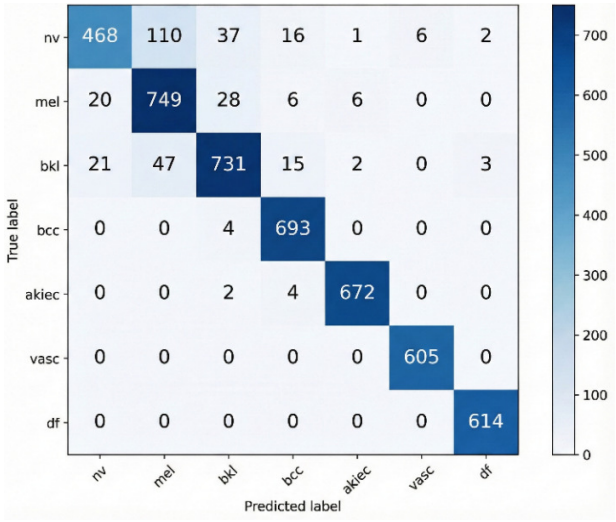

Fig. 3. Confusion matrix of ResNet50 on balanced HAM10000 (oversampled, 80:20 split).

B. Model Performance

Three MA-CNN models, namely MA-VGG16, MA-ResNet50, and MA-EfficientNetB0, were used with ensemble stacking through an RF meta-learner. Individual models achieved 95.74% (MA-VGG16), 96.17% (MA-ResNet50), and 96.38% (MA-EfficientNetB0) accuracy with good precision, recall, and F1-scores but significant misclassifications, particularly for the NV class, because of visual similarities with MEL and BKL. The ensemble model, employing stacking with RF, accomplished an impressive 99.10% accuracy, with precision (0.97–1.00), recall (0.94–1.00), and F1-scores (0.97–1.00) approximating perfection across all classes. Five-fold cross-validation yielded 99.10±0.12% accuracy (95% CI:

[99.00, 99.20]), confirming statistical robustness. Confusion matrices supported the excellent reduction in misclassifications of NV, revealing the ensemble's improved classification accuracy. ROC curves and loss plots also demonstrated that MA-ResNet50 and MA-EfficientNetB0 converged faster with less validation loss compared to MA-VGG16, whereas the ensemble model exhibited excellent generalization and stability. Figure 4 summarizes the performance across all models.

Class Label	a) MA-VGG16 (Test Acc: 93.73%)				b) MA-ResNet50 (Test Acc: 94.10%)				
	Precision (P)	Recall (R)	F1-Score (F1)	Support (Sup.)		Precision (P)	Recall (R)	F1-Score (F1)	Support (Sup.)
nv	0.94	0.73	0.82	640	nv	0.91	0.80	0.85	640
mel	0.86	0.72	0.89	809	mel	0.92	0.86	0.86	809
bkl	0.81	0.95	0.94	819	bkl	0.83	0.99	0.72	819
bcc	0.95	0.97	0.99	697	bcc	0.96	1.00	0.98	697
akiec	0.96	0.99	0.96	678	akiec	0.96	1.00	0.98	678
vasc	1.00	1.00	1.00	605	vasc	0.96	1.00	0.99	605
df	0.99	1.00	1.00	614	df	0.99	1.00	1.00	614
Averages:					**Averages:**				
accuracy	0.94	0.94	0.94	4862	accuracy	0.94	0.94	0.94	4862
macro avg	0.94	0.94	0.94	4862	macro avg	0.94	0.94	0.94	4862
weighted avg	0.94	0.94	0.94	4862	weighted avg	0.94	0.94	0.93	4862
	c) MA-EfficientNetB0 (Test Acc: 93.56%)				d) Ensemble (Random Forest) (Acc: 99.16%)				
	Precision (P)	Recall (R)	F1-Score (F1)	Support (Sup.)		Precision (P)	Recall (R)	F1-Score (F1)	Support (Sup.)
nv	0.94	0.73	0.82	640	nv	1.00	0.94	0.97	640
mel	0.92	1.00	0.86	809	mel	0.97	1.00	0.99	809
bkl	0.96	0.88	0.91	819	bkl	0.98	1.00	0.99	819
bcc	0.95	0.88	0.91	677	bcc	0.99	1.00	0.99	677
akiec	0.91	0.97	0.98	605	akiec	0.96	1.00	1.00	678
vasc	0.91	0.99	0.99	604	vasc	0.96	1.00	0.99	664
df	0.99	1.00	0.99	614	df	0.99	1.00	0.99	614
Averages:					**Averages:**				
accuracy			0.94	4862	accuracy	0.99	0.94	0.99	4862
macro avg	0.94	0.94	0.94	4862	macro avg	0.99	0.99	0.99	4862
weighted avg	0.94	0.94	0.93	4862	weighted avg	0.99	0.99	0.99	4862

Fig. 4. Detailed classification performance summary across models.

C. Comparative Analysis

All benchmarks were re-evaluated on HAM10000 with an identical 80:20 split, 224×224 input, and 5-fold Cross-Validation (CV) to ensure fair comparison. The proposed model achieved 99.10% accuracy, outperforming prior models. Benchmark models like VGG16 (80.46–91.0%), InceptionV3 (85.8–91.0%), DenseNet+MobileNet (97.7%), and state-of-the-art CNNs (96.0%) were outperformed due to the inclusion of multi-attention mechanisms and ensemble learning. The multi-attention layers enhanced feature extraction by highlighting key lesion regions, while the RF meta-learner enhanced decision-making through the fusion of predictions, promoting generalization and avoiding overfitting. This comparative analysis stresses the improvement of the suggested method in the diagnostic performance of skin cancer classification.

TABLE I. ACCURACY SUMMARY OF ALL MODELS

Models	Accuracy
VGG16	93.72%
ResNet50	94.71%
EfficientnetB0	93.56%
MA-VGG16	95.74%
MA-Resnet50	96.17%
MA-EfficientnetB0	96.38%
Proposed method	**99.10%**

D. Implementation

To demonstrate real-world application, a Graphical User Interface (GUI) was designed using the Gradio library for real-time classification of skin lesions. The web-based tool allows users to upload dermoscopic images and receive predictions, confidence scores, and diagnostic recommendations. The tool aids clinicians and researchers with an easy-to-use environment for the evaluation of skin lesions, thus enhancing the likelihood of clinical adoption and further research validation.

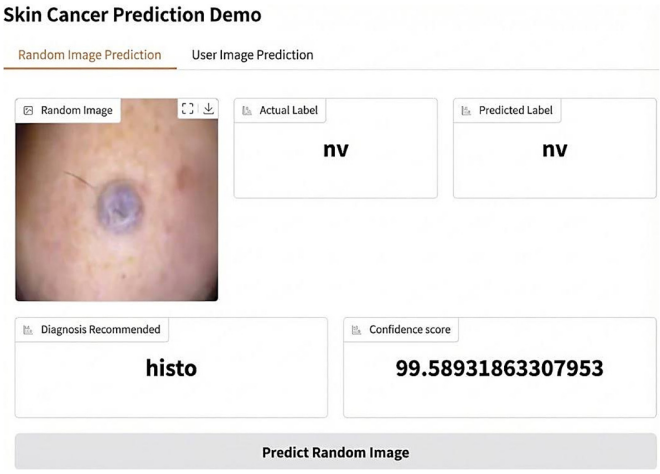

Fig. 5. A GUI webpage to demonstrate the prediction of the proposed method.

VI. CONCLUSION

This study advances skin cancer classification by addressing bias, feature extraction, and credibility. Class-balancing through oversampling, undersampling, and augmentation reduced dataset imbalance and minority-class

errors. Multi-attention mechanisms in CNNs improved focus on key lesion features, while stacking with an RF meta-learner achieved 99.10% accuracy, outperforming individual models and stabilizing NV-class predictions. The contributions of this study include balanced predictions, enhanced feature learning with visualizations, robust ensemble performance, and a generalizable framework for medical imaging. However, the proposed model involves high GPU demands, restricting low-resource use, and persistent challenges with rare lesions. In addition, this study did not focus on ethical/regulatory issues, such as those related to GDPR and HIPAA, for implementation in real healthcare settings. Despite high accuracy, clinical deployment requires GPU acceleration (A100 inference: ~0.3s/image) and Grad-CAM visualization for dermatologist trust. Future work will seek to optimize the proposed model for edge devices, integrate metadata, and improve generalizability on non-dermoscopic images.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no competing interests that could have influenced the results of this study.

ACKNOWLEDGMENT

The authors would like to acknowledge the UNITAR International University for its financial support in facilitating the publication of this paper.

DATA AVAILABILITY

The data that support the findings of this study are publicly available. The HAM10000 dataset used in this research can be accessed at [31]. No additional private datasets were used.

REFERENCES

[1] "WHO - Cancer Country Profiles Malaysia 2020," *ICCP Portal*. https://www.iccp-portal.org/resources/who-cancer-country-profiles-malaysia-2020.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 1, pp. 7–30, 2018, https://doi.org/10.3322/caac.21442.

[3] Y. S. Alsahafi, M. A. Kassem, and K. M. Hosny, "Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier," *Journal of Big Data*, vol. 10, no. 1, June 2023, Art. no. 105, https://doi.org/10.1186/s40537-023-00769-6.

[4] J. B. A. Das, D. Mishra, A. Das, M. N. Mohanty, and A. Sarangi, "Skin cancer detection using machine learning techniques with ABCD features," in *2022 2nd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON)*, Nov. 2022, pp. 1–6, https://doi.org/10.1109/ODICON54453.2022.10009956.

[5] W. Paja, J. Szkoła, K. Pancerz, J. Sarzyński, and M. żychowska, "A Preliminary Research on Automatic Identification of Melanocytic Skin Lesions from Digital Images," *Procedia Computer Science*, vol. 225, pp. 4706–4712, Jan. 2023, https://doi.org/10.1016/j.procs.2023.10.469.

[6] W. Salma and A. S. Eltrass, "Automated deep learning approach for classification of malignant melanoma and benign skin lesions," *Multimedia Tools and Applications*, vol. 81, no. 22, pp. 32643–32660, Sept. 2022, https://doi.org/10.1007/s11042-022-13081-x.

[7] S. Qian, K. Ren, W. Zhang, and H. Ning, "Skin lesion classification using CNNs with grouping of multi-scale attention and class-specific loss weighting," *Computer Methods and Programs in Biomedicine*, vol. 226, Nov. 2022, Art. no. 107166, https://doi.org/10.1016/j.cmpb.2022.107166.

[8] I. A. Alfi, M. M. Rahman, M. Shorfuzzaman, and A. Nazir, "A Non-Invasive Interpretable Diagnosis of Melanoma Skin Cancer Using Deep Learning and Ensemble Stacking of Machine Learning Models," *Diagnostics*, vol. 12, no. 3, Mar. 2022, https://doi.org/10.3390/diagnostics12030726.

[9] V. Anand, S. Gupta, D. Koundal, and K. Singh, "Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images," *Expert Systems with Applications*, vol. 213, Mar. 2023, Art. no. 119230, https://doi.org/10.1016/j.eswa.2022.119230.

[10] J. Avanija, C. Chandra Mohan Reddy, C. Sri Chandan Reddy, D. Harshavardhan Reddy, T. Narasimhulu, and N. V. Hardhik, "Skin Cancer Detection using Ensemble Learning," in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, June 2023, pp. 184–189, https://doi.org/10.1109/ICSCSS57650.2023.10169747.

[11] K. Sethanan *et al.*, "Double AMIS-ensemble deep learning for skin cancer classification," *Expert Systems with Applications*, vol. 234, Dec. 2023, Art. no. 121047, https://doi.org/10.1016/j.eswa.2023.121047.

[12] K. M. Selvaraj, S. Gnanagurusubbiah, R. R. R. Roy, J. H. J. Peter, and S. Balu, "Enhancing skin lesion classification with advanced deep learning ensemble models: a path towards accurate medical diagnostics," *Current Problems in Cancer*, vol. 49, Apr. 2024, Art. no. 101077, https://doi.org/10.1016/j.currproblcancer.2024.101077.

[13] M. M. Musthafa, T. R. Manesh, K. V. Vinoth, and S. Guluwadi, "Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification," *BMC Medical Imaging*, vol. 24, no. 1, Aug. 2024, Art. no. 201, https://doi.org/10.1186/s12880-024-01356-8.

[14] A. Panthakkan, S. M. Anzar, S. Jamal, and W. Mansoor, "Concatenated Xception-ResNet50 — A novel hybrid approach for accurate skin cancer prediction," *Computers in Biology and Medicine*, vol. 150, Nov. 2022, Art. no. 106170, https://doi.org/10.1016/j.compbiomed.2022.106170.

[15] X. He, Y. Wang, S. Zhao, and X. Chen, "Co-Attention Fusion Network for Multimodal Skin Cancer Diagnosis," *Pattern Recognition*, vol. 133, Jan. 2023, Art. no. 108990, https://doi.org/10.1016/j.patcog.2022.108990.

[16] P. Li, T. Han, Y. Ren, P. Xu, and H. Yu, "Improved YOLOv4-tiny based on attention mechanism for skin detection," *PeerJ Computer Science*, vol. 9, Mar. 2023, Art. no. e1288, https://doi.org/10.7717/peerj-cs.1288.

[17] Y. Wan, Y. Cheng, and M. Shao, "MSLANet: multi-scale long attention network for skin lesion classification," *Applied Intelligence*, vol. 53, no. 10, pp. 12580–12598, May 2023, https://doi.org/10.1007/s10489-022-03320-x.

[18] D. Halabi, "Enhancing Skin Cancer Detection and Classification: Exploring the Impact of Attention Mechanisms in Transfer Learning Models." Medicine and Pharmacology, Dec. 13, 2023, https://doi.org/10.20944/preprints202312.0943.v1.

[19] Y. Wang, J. Cai, D. C. Louie, Z. J. Wang, and T. K. Lee, "Incorporating clinical knowledge with constrained classifier chain into a multimodal deep network for melanoma detection," *Computers in Biology and Medicine*, vol. 137, Oct. 2021, Art. no. 104812, https://doi.org/10.1016/j.compbiomed.2021.104812.

[20] G. Cai and N. Lynch, "A Geometry-Sensitive Quorum Sensing Algorithm for the Best-of-N Site Selection Problem," in *Swarm Intelligence*, 2022, pp. 1–13, https://doi.org/10.1007/978-3-031-20176-9_1.

[21] M. A. Khan *et al.*, "Automatic melanoma and non-melanoma skin cancer diagnosis using advanced adaptive fine-tuned convolution neural networks," *Discover Oncology*, vol. 16, no. 1, Apr. 2025, Art. no. 645, https://doi.org/10.1007/s12672-025-02279-8.

[22] M. D. Ali *et al.*, "Breast Cancer Classification through Meta-Learning Ensemble Technique Using Convolution Neural Networks," *Diagnostics*, vol. 13, no. 13, June 2023, Art. no. 2242, https://doi.org/10.3390/diagnostics13132242.

[23] M. A. Khan *et al.*, "An Advanced Deep Learning Framework for Skin Cancer Classification," *The Review of Socionetwork Strategies*, vol. 19, no. 1, pp. 111–130, Apr. 2025, https://doi.org/10.1007/s12626-025-00181-x.

[24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.

[25] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2016, pp. 1397–1400, https://doi.org/10.1109/ISBI.2016.7493528.

[26] N. Bansal and S. Sridhar, "Skin Lesion Classification Using Ensemble Transfer Learning," in *Second International Conference on Image Processing and Capsule Networks*, 2022, pp. 557–566, https://doi.org/10.1007/978-3-030-84760-9_47.

[27] Y. C. Hum *et al.*, "The development of skin lesion detection application in smart handheld devices using deep neural networks," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 41579–41610, Dec. 2022, https://doi.org/10.1007/s11042-021-11013-9.

[28] M. D. R. H. Khan, A. H. Uddin, A. A. Nahid, and A. K. Bairagi, "Skin Cancer Detection from Low-Resolution Images Using Transfer Learning," in *Intelligent Sustainable Systems*, 2022, pp. 317–334, https://doi.org/10.1007/978-981-16-2422-3_26.

[29] M. S. Al Huda, T. E. Shrestha, A. Hossain, N. B. Sharif, M. A. Ali, and T. I. Erdei, "DeepMelaNet: Advancing Melanoma Stage Classification in Skin Cancer Diagnosis," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19627–19635, Feb. 2025, https://doi.org/10.48084/etasr.8336.

[30] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, Aug. 2018, Art. no. 180161, https://doi.org/10.1038/sdata.2018.161.

[31] P. Tschandl, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." Harvard Dataverse, Feb. 07, 2023, https://doi.org/10.7910/DVN/DBW86T.