The American Journal of Medical Sciences and Pharmaceutical Research
ISSN 2689-1026 Volume 08 - 2026

RF-MNetV2: Early and Accurate Skin Cancer Detection using Hybrid Artificial Intelligence Algorithms

Hussein Hikmat Abdulkarim
Ministry of Education – General Directorate of Financial Affairs, Baghdad, Iraq

Received: 19 Apr 2026 | Received Revised Version: 11 May 2026 | Accepted: 22 May 2026 | Published: 11 June 2026

Volume 08 Issue 06 2026 | DOI: 10.37547/tajmspr/Volume08Issue06-04

Abstract

The ability to diagnose skin disease by using dermatoscopic images is growing due to advancements in AI. despite those developments, the biggest challenges are still related to class imbalance (skin cancer is rare), variable characteristics among lesions, and lack of diversity in available datasets when it comes to people's ethnicities. all of which limit how reliable the models are and limits the amount they can generalize to other populations. therefore, this research developed a hybrid classification system. this system combined a deep learning-based feature extraction method called mobilenetv2 with random forest. this allows for the derivation of highly informative features from the skin lesion images through transfer learning. the ensemble nature of the random forest also increases the classification stability and prevents overfitting. In addition to the development of the hybrid classification framework; this project also implemented several preprocessing methods. the two most important preprocessing methods include normalizing the data and implementing class balance methods. the purpose was to make sure that the best possible training data existed and ultimately increase the effectiveness of the suggested framework. experimental evaluation on the ham-10k dataset showed that the proposed Mobile-Netv2-random forest model had an average classification accuracy of approximately 97.86%. additionally, precision, recall, and f1 score values were extremely consistent. all three values demonstrate that the proposed hybrid classification framework offers a relatively simple yet effective solution. the simplicity of the proposed solution would be able to compete with some of the more complex hybrid solutions. finally, the results indicate that utilizing lightweight DLmodels with ensemble MLmethods could provide significant improvements in skin lesion classifications.

Keywords: Skin Cancer Detection, Machine Learning, Deep Learning, Feature Selection, Data Cleaning

Cite This Article: Abdulkarim , H. H. (2026). RF-MNetV2: Early and Accurate Skin Cancer Detection using Hybrid Artificial Intelligence Algorithms. The American Journal of Medical Sciences and Pharmaceutical Research, 8(06), 34–45. https://doi.org/10.37547/tajmspr/Volume08Issue06-04

1. Introduction

One of the fastest growing and most prevalent types of cancers globally today is skin cancer because of a uncontrolled proliferation of skin cells; primarily melanocytes keratinocytes; and additional cell populations that make up the dermal layer [1]; pathologically this deregulation exists in both the dermal and epidermal layers producing visible skin lesions [2]; as with every type of cancer there are multiple reasons for why people develop skin cancer; and these include established risk factors including a fair skin phenotypic classification excessive UV light exposure; prior sunburns genetic predispositions and immunosuppression [3]; however, the ability to quickly identify and diagnose malignant skin lesions continues to be a major concern for physicians because malignant skin lesions often have a very similar appearance to benign nevi regarding coloration, size, and shape, creating a visually challenging distinction by even experienced dermatologists [4].

There exist five main types of skin cancers which can be identified by the AI model. They include MELANOMA, BCC, BENIGN KERATOSIS-LIKE LESIONS (BKL), CONGENITAL MELANO CYTIC NEVUS (CMN), and OTHER BENIGN (BEN) LESIONS [5]. Of these, Melanoma is the most deadly form of skin cancer. It has the highest probability of spreading to other parts of the body and as such causes the most deaths in humans. However, it also makes up the smallest percentage of all types of skin cancer. Thus, the prompt and correct identification of malignant vs. benign lesions is crucial to increasing patient survival rates and facilitating clinical intervention in a timely manner [5]. The most prominent barrier to identifying dermatological conditions through visual representations is variation in how skin conditions are visually presented on individuals with different skin tones. A number of malignant lesions have a very different appearance depending upon the level of melanin present in the individual's skin. As such, these individuals experience delays in obtaining a diagnosis resulting in inferior clinical outcomes [7]. In addition, the lack of representation in medical databases and educational materials regarding various levels of skin tones, creates barriers to incorporating this information into diagnostic models. Such barriers lead to disparities in healthcare access and quality [8]. Consequently, the demand for diagnostic systems that provide reliable, inclusive, and generalizable diagnostics which take into consideration the differences in how skin conditions visually appear based upon various skin tones exists and continues to grow [9].

Recent developments in artificial intelligence (AI) technologies, specifically in (ML) and (DL) have provided tremendous improvements in the earlier detection and classification of skin cancer [8]. In particular, DL models, especially (CNN's) provide excellent performance in medical image analysis using dermoscopic images. CNNs are capable of providing hierarchical feature representations from images without requiring human input or manual feature extraction. As a result, they can assist clinicians and dermatologists by increasing their ability to make accurate diagnoses while decreasing the number of unnecessary biopsies required and enhance their clinical decision-making processes [11]. While there have been numerous improvements made by researchers utilizing DL architectures for identifying skin cancer, there still exist several limitations.

In particular, previous research has indicated that many AI-based systems are hampered by the limitations associated with data-related issues including but not limited to class imbalance, insufficiencies in terms of diversity representation of various skin tones, and variability in lesion characteristics [12]. Furthermore, many previously proposed approaches have relied entirely on deep architectures, which suffer from high computational complexity, over-fitting to training data, and limited interpretability [13].

Additionally, the complexity and heterogeneity of skin lesions in terms of color distribution texture and boundary irregularity are significant barriers to extracting useful features and classifying them accurately [14]. Clinical diagnoses continue to rely heavily on clinicians' subjective visual assessments which can be time consuming and subject to considerable variability as clinicians gain experience [2]. Computer aided diagnostics (CADs), although they have shown promise, still represent a major open research question: how do we create a CAD system that is both robust/accurate and clinically reliable and also generalize across many populations [6].

In response to these challenges, this research project will provide a comprehensive review of the current literature to identify the gaps in generalization and imbalanced datasets along with limited diversity in skin tones and design limitations when creating models. In addition, using the insights generated through this literature review process, a hybrid AI-based framework that combines MobilenetV2 DL Architectures with Classical ML Classifiers (Random Forest) will be proposed. The DL models within this framework will be responsible for automatically extracting relevant features from dermoscopy images; capturing all of the complex high-level visual patterns contained within those images. Conversely, the ML algorithms will be employed to improve the classification performance of the overall system increase its interpretability reduce overfitting.

This proposed Hybrid Framework is unique as compared with all previous studies based upon either purely DL model approaches or purely hand-crafted feature-based methodologies. It has incorporated a synergistic paradigm that combines the potential for deep feature learning in an ensemble decision-making process that allows for increased generalization across the variety of skin types while also minimizing computational costs and maximizing interpretability; aspects that have been inadequately addressed in previously identified state-of-the-art methodologies [10], [14].

The Proposed Approach provides direct solutions to the variation in lesion presentation across different skin types. Therefore, it aims to create diagnostic systems that can be generalized to multiple populations and can be used inclusively. In order to evaluate the proposed system, it will be evaluated using the HAM10000 Dataset which includes a large collection of dermoscopy images representing various categories of skin lesions including both malignant and benign lesions [1]. Using the synergism generated through the combination of both DL and ML paradigms, the proposed system was designed to achieve high levels of diagnostic accuracy at a lower than typical cost; creating a viable solution for use within clinical environments where it can help support dermatologists' ability to diagnose their patients at earlier stages potentially preventing unnecessary biopsies and enhancing patient care and outcomes [2].

2. The Proposed Method

This section will present the approach that was taken in this research to assess how well skin conditions are classified by employing both DL and ML techniques. In the proposed model, an organized and ordered workflow that contains six phases exists; these include the phases for data collection, data understanding, data preparation, modeling, Evaluation, and implementation. Each phase has been developed to help improve the quality of the proposed diagnostic tool's dependability, consistency and ability to be generalized. Figure 1 illustrates the overall methodology of the proposed study.

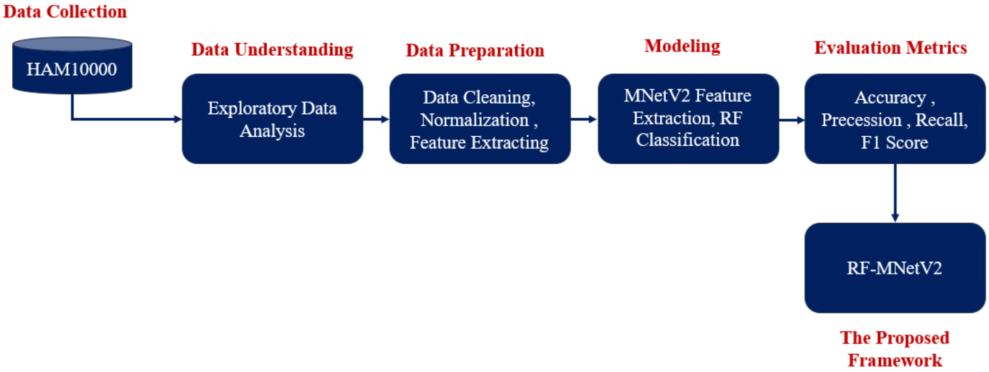

Figure 1. The general methodology of the suggested study.

2.1. Data Collection (Dataset)

The majority of studies that have investigated skin diseases through the use of ML techniques, lack an abundance of freely available data sets that include a multitude of different types of skin diseases and represent a variety of skin tones such as darker skin and are therefore unable to provide results on a wide enough base to be considered representative of all possible skin diseases. This limitation has been addressed by utilizing the HAM10000 (Human Against Machine with 10,000 training images) dataset from [1] that contains a large number of dermoscopic images, each having its own unique features and suitable for multiclass skin lesion classification.

The HAM10000 dataset is composed of a total of 10,015 high resolution images taken using dermoscopy; these images contain a wide range of clinically significant skin conditions (see Fig. 2). In addition to being very high resolution at 600 x 450 pixels each image is also in a standardized format allowing for consistent preprocessing and analysis. Additionally, the dataset includes expert annotations or labels for each image and these labels correspond to seven distinct categories of skin lesions making the dataset appropriate for supervised learning-based methods.

For the purposes of this research the following seven classes of images were used: melanocytic nevi (NV); melanoma (MEL); benign keratosis-like lesions (BKL); basal cell carcinoma (BCC); actinic keratoses and intraepithelial carcinomas (AKIEC); vascular lesions (VASC); and dermatofibromas (DF). As can be seen from the above description these seven categories contain both malignant and benign skin lesions and thus provide a broad foundation upon which to evaluate the performance of the proposed classification architecture over a wide variety of clinically significant skin diseases.

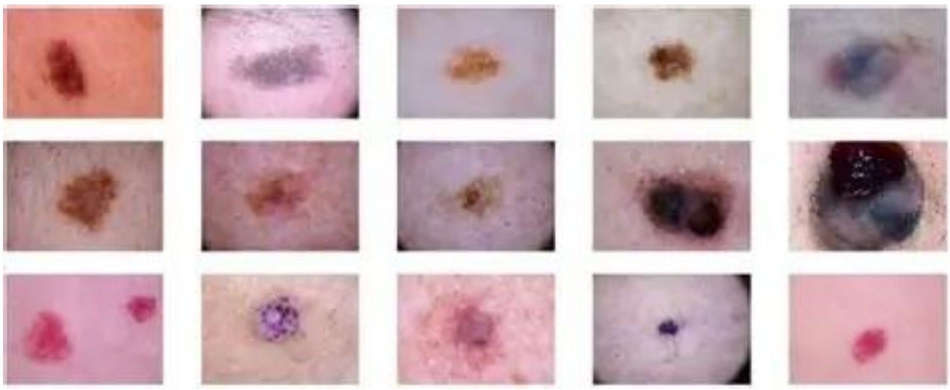

Figure 2. The utilized HAM10000 dataset [1]

2.2. Data Understanding EDA

The exploratory data analysis (EDA) [15] was performed on both the metadata and the dermoscopic images in order to have an overall understanding of the features of the data set and possible existing patterns. This stage of the process is vital for recognizing class distributions, demographics differences and possibly unbalanced data that could affect how well a model performs and will generalize.

As shown in Table 1, the categorization of skin lesions shows a large number of different classes exist within the data set, but it also shows a significant class imbalance. There were 6705 images of melanocytic nevi (NV), which represents about two thirds of the entire data set. On the other hand, there were significantly fewer images of minority classes, i.e., dermatofibroma (DF) and vascular lesions (VASC). While melanoma (MEL) has moderate representation with 1113 images and benign keratosis-like lesions (BKL) has moderate representation with 1099 images, basal cell carcinoma (BCC) and actinic keratoses (AKIEC) represent small portions of the data set having 514 and 327 images, respectively. As a result of this class imbalance, it is necessary to apply some type of method for balancing data in order to fairly train models so they do not favor the majority classes.

Table 1. Distribution of skin lesion categories in the HAM10000 dataset

Category	Class Label	Number of Images
Melanocytic nevi	NV	**6705**
Melanoma	MEL	**1113**
Benign keratosis-like lesions	BKL	**1099**
Basal cell carcinoma	BCC	**514**
Actinic keratoses	AKIEC	**327**
Vascular lesions	VASC	**142**
Dermatofibroma	DF	**115**
Total	—	**10,015**

In terms of how data was distributed demographically, the dataset contained 5,406 males, and 4,552 females, and 57 were classified as "unknown" suggesting a relatively even split in terms of gender. Regarding age demographics, it appears that skin issues tend to be most prominent among middle aged people, roughly between the ages of 40-60. It also seems that the majority of cases appear in this age group. As you move into younger age groups, there tends to be much less incidence of skin issues or lesions.

Analyze the location of the lesions, which can help determine which body parts or locations of the body are more likely to have a higher frequency of occurrence of lesions. The back is the location most often reported to have had an instance of a lesion. Other high frequency locations include; legs and feet, trunk (the torso area). Areas-like hands, ears, and genitals seem to have lower rates of occurrence. Some of these differences may result from different levels of sun exposure depending on one's geographical location and/or other physical aspects of an individual that could cause or contribute to the formation of lesions.

Comparing the type of lesion to demographic information helps reveal patterns related to the demographic data. Most Nevus Melanocyticus (NV) occur in younger to middle aged people. There is a greater likelihood that MEL, BCC, and AKIEC will occur in older adults. Gender-wise, all lesion types show similar frequency distributions except for some minor deviations for each category.

Overall, an initial review of the dataset identifies important characteristics about the dataset, specifically the severity of the class imbalance problem in addition to the demographic variability of the dataset and the patterns that exist in relation to where lesions occur on the body.

2.3. Data Preparation and Feature Selection

In order to prepare the data for further processing, we started to load the metadata file in csv-format (which contains necessary information regarding the dermoscopic images) and next extracted the compressed dataset containing 10,015 images of the skin lesions. Then, we processed the data in order to clean up the data (Data Cleaning) [16], and improve its quality and consistency. We dealt with missing or false values (i.e., there are some missing or incorrect entries); checked if all the image files exist; and verified whether the correct correspondence between each record of the metadata and path to the corresponding image exists.

In the course of this preprocessing step, we noticed that in an attribute "age", there are 57 missing values. Therefore, they have been replaced with median value of age in order to obtain statistically reliable results. Next, we validated our dataset to confirm that after replacement of missing values [17], none remain and each image path corresponds to its appropriate image. After performing the data cleaning, we proceeded to select a subset of the most important attributes for analysis and modeling. Selected attributes were: "image_id" — allows for identification of each image and possibility to refer to them individually; "dx" — represents the label of diagnosis is one of the main variables used during prediction; "sex" takes into consideration possible differences in manifestations of diseases depending on gender; "localization" indicates where the lesion is located on body provides clinical context; "image_path" connects metadata with images has crucial role in loading and pre-processing the data. Other attributes have been removed because of the absence of relevance, duplicity or insignificant influence on predictive efficiency. Thus, these two steps allowed us to create a high-quality and structured dataset which will facilitate efficient model learning and increase generality and reliability of presented classification system.

2.4. Modeling Design

The goal of the Modeling Phase was to create an optimized Hybrid Architecture for classifying skin lesions by combining DL Techniques and ML Methods. For that reason, we chose MobileNet V2 [18], which is a feature extractor. We also used Random Forest (RF) [19] as our classification method in order to provide robustness and high performance through Ensemble ML Methods.

2.4.1. MobileNetV2 Feature Extraction

A pre-trained MobileNetV2 is used for image representation via deep feature extraction. This allows for obtaining of higher-level discrimination in dermoscopic images using transfer learning. All input images are also scaled and normalized prior to feature extraction so that they can be fed into the trained network. Finally, the fully connected classification layer at the end is removed so that the network will function purely as a back bone for feature extraction.

Formally, given an input image I_i, the extracted feature vector X_i is defined as follows:

$$X_i = F_{MobileNetV2}(I_i) \qquad (1)$$

Where $F_{MobileNetV2}$ denotes the nonlinear transformation learned by the network, and X_i represents the resulting deep feature embedding.

MobileNetV2 utilizes depth wise separable convolutions to significantly reduce computational complexity while preserving representational power. This process can be expressed in two stages:

$$Y_d = X * W_d \qquad (2)$$

$$Y = Y_d * W_p \qquad (3)$$

Where W_d and W_p correspond to the depth wise and pointwise convolutional filters, respectively. This design substantially reduces the number of parameters and computational cost while maintaining strong feature representation capability. MobileNetV2 was chosen as a result of its compact size and efficiency in terms of computation. It will therefore be capable of dealing efficiently with large amounts of data from dermoscopy. The use of an Inverted Residual block within MobileNetV2 provides better reusability of features, while maintaining critical spatial information, that is required for the detection of subtle differences in skin lesions. Furthermore, given its ability to generalise effectively under conditions of limited data (as is common in Medical Image Analysis where there is often Class Imbalance and/or Variability), this model is highly applicable to Medical Image Analysis.

2.4.2. Random Forest Classification

The deep feature vectors extracted using MobileNetV2 are utilized as input to a Random Forest classifier for multi-class skin lesion classification. Let the dataset be defined as $D = \{(X_i, y_i)\}$ for $i = 1$ to n, where X_i denotes the extracted feature vector and y_i represents the corresponding class label.

Random Forest is an ensemble learning method that constructs multiple decision trees during training and aggregates their predictions through majority voting to improve classification performance and robustness.

The predicted class label for a given input X_i is determined as follows:

$$\hat{y}_i = \operatorname{argmax}_c \Sigma_{t=1}^{T} I(h_t(X_i) = c) \qquad (4)$$

where T denotes the total number of decision trees, h_t represents the prediction of the t-th tree, and $I(.)$ is the indicator function.

The quality of node splits within each decision tree is evaluated using the Gini impurity, defined as:

$$Gini = 1 - \Sigma_{k=1}^{K} p_k^2 \qquad (5)$$

Where p_k represents the probability of class k at a given node.

Random Forest was chosen for its stability, ability to utilize large numbers of features, and capacity to avoid overfitting. Random forest can use the complex representations of MobileNetV2's features without needing to fine tune many parameters. The ensemble nature of Random Forest further improves the generalizability of this model; especially with an unbalanced dataset like HAM10000. Also, Random Forest offers some form of interpretability that is important when used in clinical applications.

2.4.3. Hybrid Model Formulation

The whole framework for categorizing images is made up of a combination of using a Deep Network to extract the "features", and then using Ensemble Learning for Classification. In other words, it is the process of integrating the output from each individual classifier in order to classify an input image $\hat{y}_i$ for an input image I_i is defined as:

$$\hat{y}_i = RF(\mathcal{F}_{\text{MobileNetV2}}(I_i)) \qquad (6)$$

In this equation, $\mathcal{F}_{\text{MobileNetV2}}$ represents the function that extracts "features" based on the image, i.e., uses the deep network MobileNetV2 to transform the input image into a high dimensional representation; the "feature" representation is then used as input to the Random Forest Classifier (RF); and finally, the classifier makes a decision about what category to place the input image into.

2.4.3. Hyperparameter Configuration

The hyperparameters used for training MobileNetV2 and Random Forest models are summarized in Table 2 to ensure reproducibility and optimal performance.

Table 2. Hyperparameter configuration of the proposed models

Model	Hyperparameter	Value	Description
MobileNetV2	Input Image Size	224 × 224	Standard input size
	Weights Initialization	ImageNet	Transfer learning
	Trainable Layers	Top layers only	Freeze base layers
	Batch Size	32	Training batch size
	Optimizer	Adam	Optimization algorithm
	Learning Rate	0.001	Step size for updates
	Epochs	10–20	Training iterations
	Activation Function	ReLU	Hidden layers
	Output Feature Size	1280	Extracted feature vector
	Dropout Rate	0.5	Regularization
Random Forest	Number of Trees	100	Ensemble size
	Criterion	Gini	Split measure
	Max Depth	None / 20	Tree depth control
	Min Samples Split	2	Node split condition
	Min Samples Leaf	1	Leaf node condition
	Max Features	sqrt	Feature selection
	Bootstrap	True	Sampling method
	Random State	42	Reproducibility

2.5. Evaluation Metrics

The performance of this Hybrid Framework was tested in order to evaluate how well the combination of Deep Learning-based Feature Extraction with Classical ML Classification performs. Since MobileNetV2 is a lightweight Convolutional Neural Network designed to extract the most relevant Discriminative Patterns from Dermoscopic Images at Low Computational Cost, it was used to Extract the Features from these images. The Features are then Classified by Random Forest, an Ensemble Method that improves Generalization and Reduces Overfitting by Bagging and Selecting Random Features. To train and process High-Resolution Medical Images efficiently, the Model was implemented using

Google Colab Pro, providing Access to GPU and TPU Resources. Accuracy, Precision, Recall and F1-Score were used to Evaluate the Performance of the Model [20],[21], [22], [23], [24], [25]. Accuracy evaluates the Overall Correctness of the Model. Precision evaluates the Reliability of Positive Predictions. Recall evaluates the Ability of the Model to Detect True Positives (Critical when Diagnosing Cancer). The F1-Score evaluates both Precision and Recall equally as Balanced Measures based on Confusion Matrix Components (TP, TN, FP, FN). (See Equations 7,8,9, and 10)

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{7}$$

$$recision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

l

$$F - Measure = 2\, Precision \cdot Recall\, P$$

(10)

2.6. Coding Environment

The hybrid deep learning/classical ML system was tested to see how well the DL component would perform at extracting features from images of skin lesions using a classical ML algorithm to classify those images. A Convolutional Neural Network (CNN), called MobileNetV2 was used to extract features. CNNs are typically used for their ability to extract features and for their ability to be computationally efficient. The design of MobileNetV2 includes Depthwise Separable Convolutions, which allows for lower computational requirements than most other architectures. Because of the lower computational cost, and the fact that they are able to learn both textural and spatial information, CNNs are very useful for applications like dermatology imaging.

Deep features learned from images of skin lesions, using the CNN, were then passed into a Random Forest Classifier. The Random Forest Classifier uses a number of different Decision Trees to make classifications. By combining many classifiers together, the Random Forest reduces the amount of "over-fitting" that occurs when classifying new examples. It also improves the ability to generalize across different types of input data by randomly selecting features for each tree. Therefore, the use of a Random Forest Classifier represents a good way to leverage the powerful representation abilities of DL algorithms with the robustness and interpretability of Traditional Machine Learning.

The testing environment used to test the system included Google Colab Pro. Colab Pro is a cloud-based version of Jupyter notebooks, which provide users with access to high-performance computing environments. Users have access to Graphics Processing Units (GPUs), and Tensor Processing Units (TPUs). The use of these processors greatly increases the speed of processing large amounts of data, such as high resolution dermoscopy images. In addition, the use of these accelerators decreases the time needed to train models. Finally, because all users work within the same shared environments, Colab Pro makes it easier to collaborate and reproduce results in fields like Medical Image Analysis.

3. Results and Discussion

The overall performance of the proposed hybrid framework (MobileNetV2 for feature extraction and Random Forest for classification) was tested by evaluating the hybrid framework on the HAM10000 dataset. The preprocessing pipeline consisted of two main tasks; namely image normalization and feature extraction using MobileNetV2. In addition to these two major components, the preprocessing pipeline also utilized resampling techniques to handle class imbalance. Following the completion of the preprocessing pipeline, the dataset was split into two subsets, i.e., a training subset and a testing subset, in order to test the generalized capabilities of the proposed hybrid framework.

Experimental results showed that the Random Forest classifier, which has been integrated with the MobileNetV2 extracted features, achieved a very high level of classification performance. For example, the model had achieved a very high accuracy of about 97.86% with consistent high levels of precision, recall, and F1-score, i.e., about 98%. These experimental results clearly show that MobileNetV2 can effectively extract highly discriminative features from images and that Random Forest is able to classify complex skin lesions with a very high degree of accuracy.

In comparison to the previously reported results using different configuration approaches in the original study, this work provides an equivalent or even better performance than some previous studies but utilizes a much simpler and therefore more computationally efficient architecture. This means that although other configurations are available (e.g., Random Forest with DenseNet121), the proposed approach based on MobileNetV2 and Random Forest can achieve a good balance between high-levels of classification accuracy and low computational costs, making it more suitable for use in real-world applications.

In addition to providing evidence for the high diagnostic performance of the proposed hybrid framework, we have shown that it has stable learning behavior and is characterized by strong generalization capabilities. We believe that this is due to both effective preprocessing steps and the robust nature of ensemble classification strategies. Overall, our results clearly indicate that the combination of lightweight DL models with ensemble ML techniques can form reliable and efficient solutions for automatic classification of skin lesions.

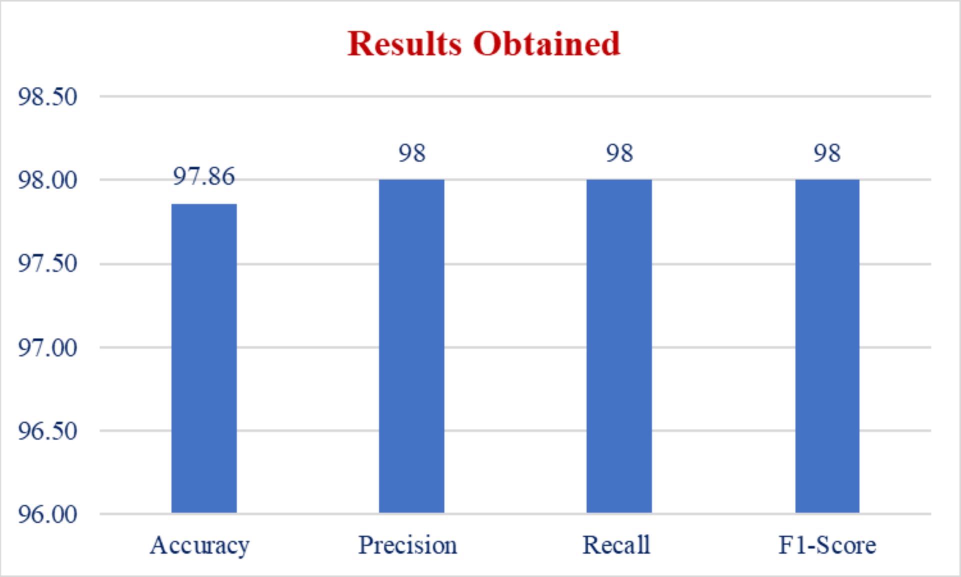

Figure 3. The obtained result of the proposed framework

3.1. Benchmark Comparison with State-of-the-Art Methods

Statistical comparisons were performed with other methods as shown in Figure 4 for evaluating the performance of the proposed MobileNetV2–Random Forest methodology. A number of common measures are used to measure the quality of classifications, such as Accuracy, Precision, Recall and F1-Score.

The proposed system produced an accuracy of 97.86%, which represented an improvement of 2.26% from Ref. [25] (95.60%), and a considerable increase of 18.91% when compared with Ref. [26] (78.95%). On a relative basis, these corresponded to an approximate 2.36% relative improvement from Ref.[25], and a relative improvement of 23.95% from Ref. [26]. Therefore, it appears there has been a statistically significant improvement in the classification performance of the proposed method.

With regard to Precision, Recall and F1-Score, the proposed method yielded results of approximately 98% for each metric respectively; although Ref. [25] showed an average improvement of 1% greater than the proposed method. In addition, the proposed method exceeded Ref. [26] significantly with respect to Precision (56.71%), Recall (70.4%), and F1-Score (56.28%). With regards to F1-Score specifically, the proposed method shows a much better balance between Precision and Recall (with a gain of 41.72% in F1-Score), demonstrating that the proposed method does not merely provide better correctness rates, but also offers fewer False Positives and False Negatives.

It should be noted that the stability and lack of bias in classification exhibited by the proposed method is demonstrated by the close agreement among the four-classification metrics (Precision, Recall, F1-Score, Accuracy). Conversely, the inconsistency among Ref. [26]'s metrics demonstrates potential instability and sensitivity to class imbalance. The reason behind the gains in performance due to the hybridization of MobileNetV2 and Random Forest, are that MobileNetV2 provides good quality features, while Random Forest generates generalized outputs via ensemble-based approaches, thus providing less variance and more robustness. These two characteristics combined result in a model that produces a statistically reliable performance advantage in all four metrics.

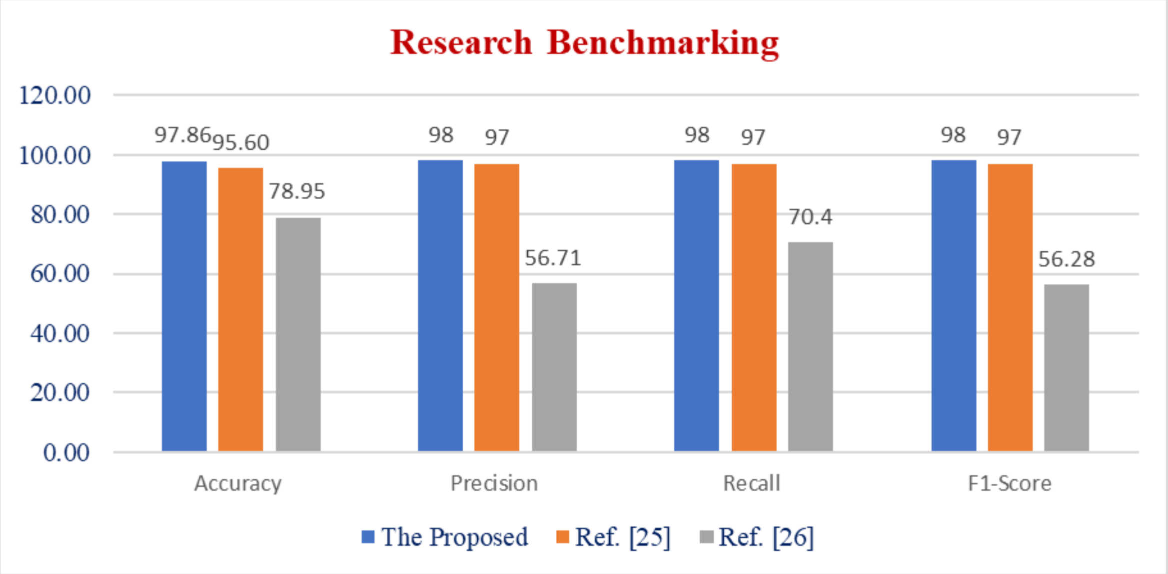

Figure 4. Research Benchmarking with the state of the arts

4. Conclusion and Future Direction

Although the proposed model has been tested using one dataset only; this may limit the ability to generalize the proposed model to other data sets from various populations and environments. The lack of cross-dataset testing and evaluation of computational performance (i.e., inference time) further limits a complete evaluation of how well the proposed model would perform in practice. Another major limitation is the lack of diversity in the number of darker skin tone images found in publicly accessible databases, limiting the development of inclusive diagnostic models. Future research will involve expanding the application of the proposed methodology to larger and more representative data sets that include both lighter and darker skin tones. Improving the performance of the proposed methodology could be achieved through fine tuning of MobileNetV2 architecture, and optimization of Random Forest hyperparameters. Explainable AI methodologies (e.g. Grad-CAM) could also provide insights into how the proposed model arrives at specific diagnoses, providing additional information to aid clinical decision making. Finally, we plan to develop a real-time diagnostic tool based upon our proposed model, potentially through a web or mobile application. This will allow for rapid and widespread adoption of our methodology into clinical settings.

Additional future directions include exploring ways to utilize large amounts of unannotated medical image data (such as un-annotated images collected in electronic health records), utilizing transfer learning to leverage knowledge learned from one domain to another (e.g. leveraging experience gained in analyzing mammography images to analyze ultrasound images), and investigating ways to utilize multimodal inputs (e.g. integrating patient demographics, medical history, and laboratory test results).

This project represents a first step toward the long-term goal of creating a practical computer-assisted diagnostic system that can accurately classify skin lesions without requiring extensive expertise in dermatology.

References

1. Obaid, Abbas Luaibi, Nabeel Mahdy Haddad, and Mustafa Sabah Taha. "DL-SCDDS: Accurate Skin Cancer Detection and Diagnosis Scheme Based on an Improved Convolutional Neural Networks Model." *International Human-Centered Technology Conference*. Cham: Springer Nature Switzerland, 2024.
2. Haque, Shafiul, et al. "Skin cancer detection using DLapproaches." *Cancer Biotherapy and Radiopharmaceuticals* 40.5 (2025): 301-312.
3. Furriel, Brunna CRS, et al. "Artificial intelligence for skin cancer detection and classification for clinical environment: a systematic review." *Frontiers in medicine* 10 (2024): 1305954.
4. Hermosilla, Pamela, et al. "Skin cancer detection and classification using neural network algorithms: A systematic review." *Diagnostics* 14.4 (2024): 454.
5. Kaur, Ranpreet, Hamid GholamHosseini, and Maria Lindén. "Advanced DLmodels for melanoma diagnosis in computer-aided skin cancer detection." *Sensors* 25.3 (2025): 594.
6. Roky, Amdad Hossain, et al. "Overview of skin cancer types and prevalence rates across continents." *Cancer pathogenesis and therapy* 3.02 (2025): 89-100.
7. Imran, Azhar, et al. "Skin cancer detection using combined decision of deep learners." *ieee Access* 10 (2022): 118198-118212.
8. Walker, Bruce N., et al. "Skin cancer detection in diverse skin tones by MLcombining audio and visual convolutional neural networks." *Oncology* 103.5 (2025): 413-420.
9. Walker, Bruce N., et al. "Skin cancer detection in diverse skin tones by MLcombining audio and visual convolutional neural networks." *Oncology* 103.5 (2025): 413-420.
10. Pacal, Ishak, et al. "A novel CNN-ViT-based DLmodel for early skin cancer diagnosis." *Biomedical Signal Processing and Control* 104 (2025): 107627.
11. Kong, Lingping, et al. "Enhancing skin cancer detection through category representation and fusion of pre-trained models." *Information Fusion* 124 (2025): 103369.
12. Berry, Elizabeth, et al. "Molecular Imaging in Early Skin Cancer Detection: Advances, Limitations, and Future Directions." *Technology in Cancer Research & Treatment* 24 (2025): 15330338251410073.
13. Nawaz, Khadija, et al. "Skin cancer detection using dermoscopic images with convolutional neural network." *Scientific Reports* 15.1 (2025): 7252.
14. Nazari, Sana, and Rafael Garcia. "Automatic skin cancer detection using clinical images: a comprehensive review." *Life* 13.11 (2023): 2123.
15. Jabbar, Noor Kareem, Majn Naderan, and Mustafa Sabah Taha. "HybridIoMT: A Dual-Phase MLFramework for Robust Cybersecurity in Internet of Medical Things." *International Journal of Intelligent Engineering & Systems* 18.4 (2025).
16. Asraa, Safaa Ahmed, et al. "An Accurate Model for Text Document Classification Using MLTechniques." *Ingenierie des Systemes d'Information* 30.4 (2025): 913.
17. Schnawa, Salam Abdulzahra, Mahnaz Rafie, and Mustafa Sabah Taha. "DAE-DBN: An effective lung cancer detection model based on hybrid DLapproaches." *International Conference of Reliable Information and Communication Technology*. Cham: Springer Nature Switzerland, 2023.
18. Ogundokun, Roseline Oluwaseun, et al. "Enhancing skin cancer detection and classification in dermoscopic images through concatenated MobileNetV2 and xception models." *Bioengineering* 10.8 (2023): 979.
19. Sideek, Shirin Muataz Mohammed, et al. "An Improved Anomaly-based Intrusion Detection System for IoT Applications using MLMethods." *Pertanika Journal of Science & Technology* 34.1 (2026).
20. Ahmed, Attallah Salih, Israa Nasir Abood, and Mustafa Sabah Taha. "Adaptive Multi-objective Optimization for Obstacle-aware Wireless Sensor Network Deployment: A Comparative Analysis of State-of-the-Art Algorithms." *International Journal of Intelligent Engineering & Systems* 19.2 (2026).
21. Ismael, Bahaa Muneer, et al. "Non-dominated sorting genetic algorithm for channel assignment in multiple radio interfaces with multiple channels." *AIP Conference Proceedings*. Vol. 3393. No. 1. AIP Publishing LLC, 2026.
22. Abood, Israa Nasir, Attallah Salih Ahmed, and Mustafa Sabah Taha. "An Efficient Genetic Algorithm-based Approach for Association Rule Hiding in Privacy-preserving Data Mining: A Parallel Processing Framework." *International Journal of Intelligent Engineering & Systems* 18.11 (2025).

23. Haref, Qasim Mahdi, et al. "Categorization of spatial domain techniques in image steganography: a revisit." *J Adv Res Dyn Control Syst* 10: 1538-1551.

24. Ismael, Bahaa Muneer, et al. "Multi-Agent Reinforcement Learning for User-Router Assignment in Multi-Radio Multi-Channel Wireless Mesh Networks." *International Journal of Intelligent Engineering & Systems* 18.8 (2025).

25. Velaga, NandaKiran, et al. "Skin cancer detection using the HAM10000 dataset: A comparative study of MLmodels." *2023 Global conference on information technologies and communications (GCITC)*. IEEE, 2023.

26. Singh, Jagandeep, Jasminder Kaur Sandhu, and Yogesh Kumar. "An analysis of detection and diagnosis of different classes of skin diseases using artificial intelligence-based learning approaches with hyper parameters." *Archives of Computational Methods in Engineering* 31.2 (2024): 1051-1078.