Fuzzy Gompertz-Based Deep Ensemble with Explainable AI for Skin Lesion Classification

Mostofa Rakib Raihan[1], Kamrul Hasan Talukder[2]

[1]Researcher, Computer Science and Engineering Discipline, Khulna University, Khulna-9208, Bangladesh
[2]Supervisor, Computer Science and Engineering Discipline, Khulna University, Khulna-9208, Bangladesh

Abstract
The skin cancer presents a formidable issue that requires prompt and precise diagnosis to ensure effective treatment. Analysis of medical imagery has been considerably enhanced by deep learning, particularly in the classification of skin disease. Deep ensemble approaches offer a compelling opportunity to further improve diagnostic accuracy. This research proposes an ensemble approach based on transfer learning techniques to achieve more precise outcomes. An ensemble model is created using ResNet50V2, DenseNet121 and MobileNetV2 for classifying skin lesions. Data augmentation methods were employed to enhance model accuracy by mitigating class imbalance. The final predictions are generated using the Gompertz function, which produces a fuzzy ranking of the base classifier models. The ensemble model shows an outstanding performance accuracy of 97.00% on HAM10000 dataset. The model's predictions were validated through Grad-CAM visualizations, revealing its focus on relevant lesion areas. These findings underscore that artificial intelligence-driven medical diagnostics can provide dependable and interpretable assistance for physicians, particularly in areas with reduced access to professional diagnostic tools.

Keywords: Skin lesion classification, Transfer Learning, Ensemble learning, Gompertz function, Fuzzy ranking, Explainable AI, Medical Image Analysis.

1. Introduction

Skin is considered to be the single largest organ in human, covering nearly 20 square feet and playing a crucial role in regulating body temperature, protecting vital organs from external harm and pathogens, and facilitating the sensations of touch, heat, and cold. Skin cancer is the unusual growth of skin tissue. This form of cancer is the most prevalent worldwide, including three primary types: “squamous cell carcinoma” and “basal cell carcinoma”. Studies [1] have shown that only in the United States, over 3.5 million instances are diagnosed yearly, more than any other type of cancers. Every minutes a new skin cancer case is reported.

Skin lesions, whether benign or malignant, represent a significant global public health challenge. Some skin conditions may exhibit symptoms months after their onset, allowing the disease to progress and develop more severely before detection. Detecting and classifying skin disease is a very challenging task in the medical industry. Sometimes it is really tough to detect the exact type, because of the complicated texture of human skin. Dermatologists also face challenges in diagnosing skin diseases and requires costly laboratory tests to precisely identify the nature and stage of the disease. Advancements in laser and photonics technology have improved the quick and accurate diagnosis of skin diseases, but these methods remain costly and time-consuming. Detecting skin cancer at an early phase is crucial, as it can increase the five-year survival rate by approximately 14% [2]. In recent years, there has been a substantial rise in the widespread use of computer-aided diagnosis (CAD), particularly in the classification of skin lesions that use deep learning techniques. High quality visualization of dermatoscopic images is a must for this process. Deep Convolutional Neural Networks can classify these images, segment data, and make predictions [3] [4]. AI has shown its potential in overcoming certain challenges associated with conventional diagnostics [5–8], particularly in rapidly identifying skin lesions [9, 10]. AI can detect patterns and features in medical images that may be missed by humans, improving accuracy and speed [11], especially when specialized testing is unavailable.

This study utilizes three CNN architectures along with fuzzy ensemble techniques to improve performance by combining predictions from multiple algorithms. Ensemble methods enhance efficiency by merging the strengths of base models. Explainable AI aims to clarify how image features influence decisions, thereby increasing trust in the AI system and providing understandable visual explanations for users [12]. The objective of the research is to build a tool for diagnosis as well as interpretable and accurate, thereby allowing dermatologists to confidently trust the prediction. To tackle the issue of unbalanced dataset, data augmentation method is used. By creating a more balanced dataset, different skin types can be classified accurately. The main points are summarized below:

- Medical skin lesion images are resized to pixel 224*224 for better memory usage and speed things up. This helps make the model more efficient and usable in real-life situations.
- Classic data enhancement techniques are used to deal with the imbalance between classes. This not only helps prevent overfitting but also makes the model tougher.
- For classifying skin lesions effectively an ensemble model is created using ResNet50V2, DenseNet121, and MobileNetV2. The overall classification performance is enhanced by integrating the characteristics of these transfer learning models.
- A confusion matrix and a variety of evaluation metrics are employed for performance evaluation & comprehensive understanding.
- Grad-CAM (Gradient-weighted Class Activation Mapping) is integrated into deep learning to facilitate the clarification of the predictions generated by this intricate model.

Rest of this paper is laid out like this: Section 2 gives a quick look at related research, Section 3 outlines our ensemble framework based on CNN for classifying skin diseases and explains how it works, and Section 4 shares our results and comparisons. Finally, we wrap it all up with our conclusions and talk about what future steps we can take.

2. Literature Review

The World Health Organization (WHO) [14] anticipates that, “13.1 million individuals will die from to cancer worldwide by 2030, with the majority of these deaths occurring in the United States, where skin cancer is the most prevalent”. Skin cancer is a common human disease that frequently spreads to other parts of the body because of its rapid penetration and abnormal cell growth [15]. Various approaches have been tried in recent years. Presents and practices, with a focus on the health field classify skin lesions.

Sonmez et al. [16] implemented convolutional neural networks for the accurate classification and detection of skin cancer utilizing dermoscopic images. They acknowledged that the issues posed by imbalanced data in this datasets can only be addressed through comprehensive data augmentation and sophisticated preprocessing changes. They achieved an accuracy of 80.79% on MNIST HAM10,000 dataset for the classification task. Sandhua et al. [17], illustrated the potential of CNNs for developing a classification method of skin disease. In their study, they employed ResNet 50 & MobileNetV2 for classification task. The both models achieved peak accuracies of 96% and 89% respectively. Chowdhury et al. [18] employed customized CNN to classify seven (07) categories of skin cancer on HAM10,000 dataset [19]. This approach has attained accuracy of 82.7% and precision of 78%. Class activation mapping (CAM) was used for validating the accuracy. Nunnari et al. [20] showed the impact of GradCAM [13] on ISIC 2019 dataset. The accuracy achieved was 72.20% and 76.70%, respectively with VGG16 and ResNet-50 as classification models. Natasha et al. [22] proposed an explainable AI based deep learning approach on ISIC-2019 dataset. The model identifies a total of eight (08) types of skin lesions. Classification accuracy of 94.47%. LIME -The local interpretable model-agnostic explanations framework is employed to further analyze these predictions in order to produce visual explanations that supports the general explanation. Jasil and Ulagamuthalvi [23] employed three transfer learning method VGG16, VGG19 and InceptionV3 for identifying skin cancers. On ISIC 2018, accuracy rates were achieved respectively on 74%, 77% and 76% with InceptionV3, VGG16, and VGG19 models.

Alhudhaif [24] utilized six transfer learning networks on ISIC dataset to evaluate the classification. Data augmentation was applied to address the data imbalances to improve accuracy. Real-world problems have become more complicated day by day. This has prompted the development of novel and more sophisticated algorithms, transfer learning techniques, ensemble learning frameworks, and more accurate models.

Rahman et al. [26] established an ensemble method for skin cancer classification to improve dermoscopic image diagnosis. The research applied five pre-trained deep learning models: ResNeXt, SE-ResNeXt, ResNet, Xception, and DenseNet. Trained on a merged dataset which includes HAM10000 and ISIC 2019 makes it a total of 18,233 dermoscopic samples. Advanced pre-processing techniques such as augmentation, noise filtering & stratified sampling were implemented to balance the dataset. Weighted ensemble showed a macro-average recall of 94%, which was higher than all single-model classifiers.

Thwin and Park [21] proposed a deep transfer learning based ensemble framework for classifying skin lesions. They combined different pretrained CNN models, like ResNet50, VGG16, and InceptionV3, into a single model using a weighted blending approach. They achieved 96% accuracy on the balanced HAM10000 dataset.

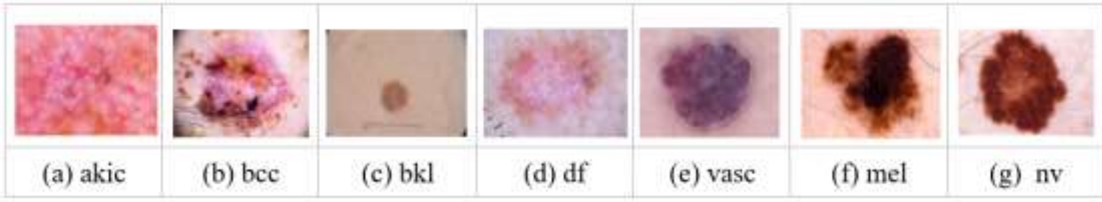
A CNN methodology for the early classification of skin cancer was proposed by Hayat and Indraswari [27]. On the ISIC 2017 dataset, they conducted an analysis of five CNN architectures: InceptionV3, ResNet50, EfficientNetB0, NASNetMobile and MobileNetV2. While MobileNetV2 attained the highest individual model accuracy of 69.3%, they used bagging as ensemble approach to combine all models boosted performance to 80.6%.

Natha and RajaRajeswari [28] have proposed an ensemble model using Max Voting technique over five classical classifiers (Random Forest, CatBoost, AdaBoost, Extra Trees, Gradient Boosting) with the improvement of the genetic algorithm based feature selection. Their approach achieves 95.80% accuracy on skin cancer classification.Liu et al. [29] achieved 86.70% accuracy on HAM10000 dataset. They used MobileNetV2, ResNet18, and VGG11 as base classifiers and stacking as ensemble approach. This is prevalent that, the reliability & classification accuracy can be increased by employing ensemble models.

3. Dataset

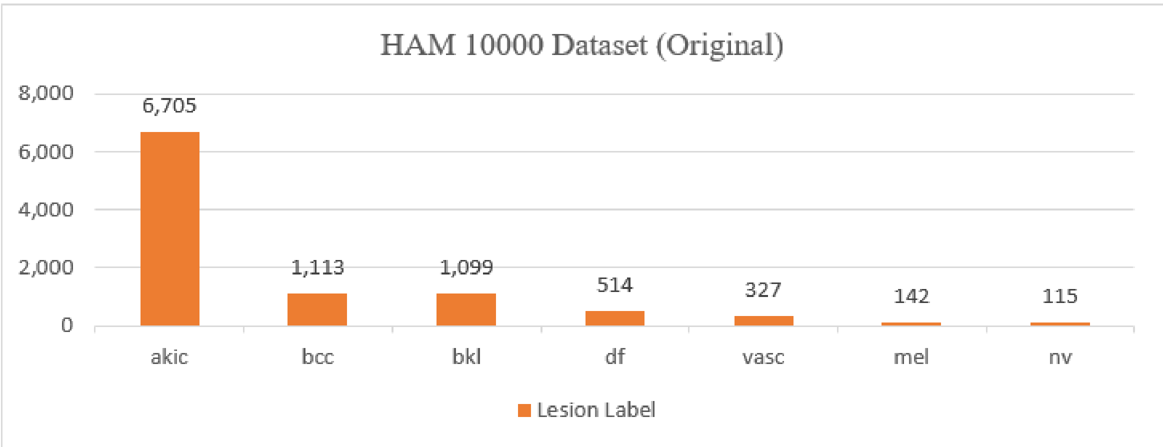
The HAM10000 [19] is a large, easily accessible repository of dermatoscopic samples that includes 10,015 images. It is designed to enable comprehensive examination of diverse skin disorders. The photos are classified into seven unique categories: "Actinic Keratoses," "Basal Cell Carcinoma," "Benign Keratosis," "Dermatofibroma," "Melanoma," "Vascular Lesions," and "Melanocytic Nevi." Figure 1 presents the sample images corresponding to each type of skin lesion.

Figure 1: Seven type of sample distribution of HAM10000



A major issue within the dataset is the clear class imbalance, with "Melanocytic Nevi" representing the most dominant group consisting almost 67% of the total dataset. This disparity highlights the necessity for strong class balancing techniques in managing uneven class distributions. Figure 2 provides an in-depth analysis of the distribution of picture samples of HAM10000 dataset for the seven skin cancer categories. Nonetheless, sufficient balanced data is required to train a deep learning model.

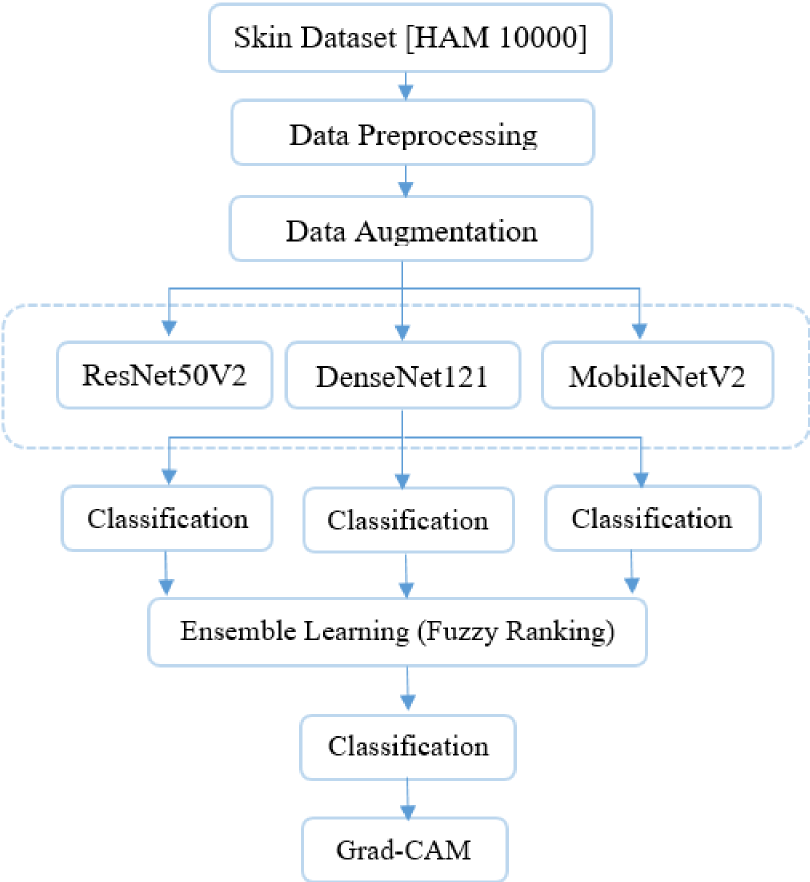
Figure 2: Number of Images for Each Class [19]



4. Proposed Methodology

To develop the model for classifying skin diseases, a structured methodology is adopted comprising several essential stages. Initially, the input images are preprocessed to improve quality and eliminate noise. Proposed convolutional neural network (CNN) models are trained separately on these preprocessed images for feature extraction. Figure 3 illustrates a brief overview of the proposed methods.

Figure 3: Proposed Methodology

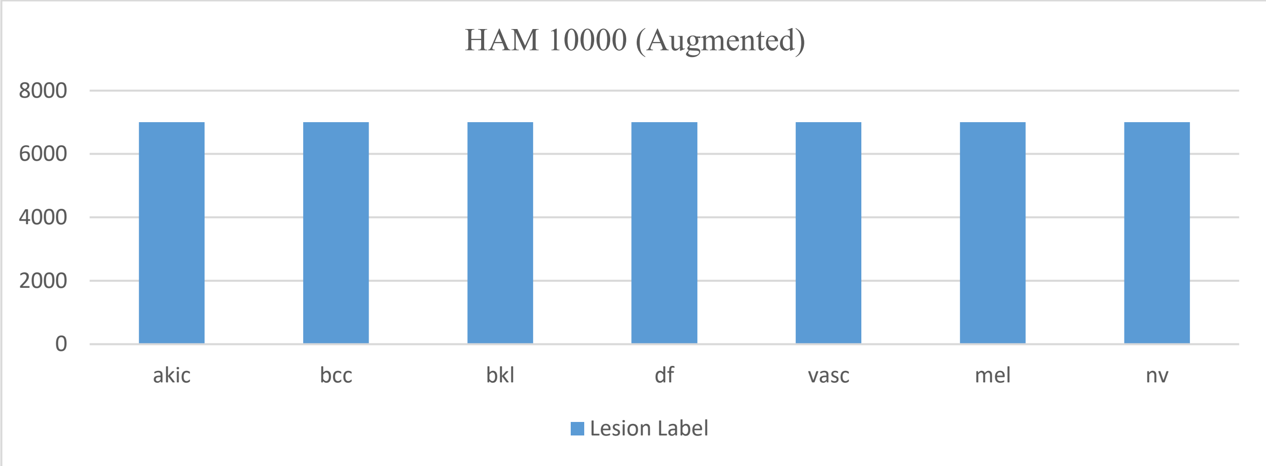


Afterward, the individual models outputs are integrated by using a fuzzy ensemble approach, which leverages their unique strengths to boost their overall classification performance. This strategy enables us to effectively address the complexity of skin disease categories and deal with the issue of data imbalance. The subsequent sections provide a detailed explanation of each phase, including data preprocessing, model training, ensemble integration and explainability.

4.1 Image Preprocessing

The dataset is significantly imbalanced, with 67% of the sample belongs to one "Melanocytic Nevi" class. To compensate for this imbalance, standard preprocessing techniques such as scaling and data augmentation were used. This dataset's original samples were 600 x 450 pixels. To ensure uniformity in input sizes, they were scaled to 224 × 224 pixels. Zooming, shearing, rotation, and flipping were among the data augmentation techniques used to improve the image quantity and modify the distribution of classes. In all, 49,000 samples from seven classes were acquired. Various data preprocessing techniques, including standardization, normalization and noise removal were implemented to enhance the quality of these images for model training.

Figure 4: Image Distribution after Augmentation



Standardization adjusted the pixel values of every image to a mean of zero and a standard deviation of 1, enhancing the model's learning process. Normalization rescaled the pixel values ranging 0 to 1, minimizing variations caused by differences in brightness and contrast. To further enhance image quality, noise reduction techniques like Gaussian blurring were used to suppress unwanted noise and improve visual clarity. Finally, the dataset was randomly divided, with 80% for training and the rest 20% reserved for validation.

4.2 Classification Models

Transfer learning is a method that uses pre-trained CNN architectures for solving new but related tasks [25]. Three such models: ResNet50v2, DenseNet121, and MobileNetV2, which was pre-trained on ImageNet, a large-scale image datasets. These architectures are well-known for their powerful ability to extract high-level patterns from input images. Through combination of several foundational networks, ResNet50v2 DenseNet121, and MobileNetV2, we intent to exploit their unique strengths and increase the classification accuracy. All of them contribute unique features and views to be finally used together to make predictions. By using transfer learning, we can leverage complex, pre-learned representations for skin cancer classification, even when annotated data is limited. These models are capable of capturing both low-level and high-level representations of the images by extracting features from a variety of layers. The extracted features effectively encode key attributes of the dermatoscopic images, including shape, texture, spatial details, and other relevant patterns. The fuzzy ranking-based approach can then use these features to classify the dataset, providing meaningful representations.

4.2.1 Resnet50V2

ResNet50V2 [30] is a commonly used deep convolutional neural network structure for image processing tasks. It is especially successful in medical image applications including skin disease classification. This model extends the original ResNet50 with a pre-activation design, with batch normalization and ReLU being applied prior to convolution layers. This improvement leads to better gradient flow and training stability in deeper networks, especially when training on complex high-resolution dermatological images. In skin disease classification, ResNet50V2 can be adjusted using dermatoscopic image datasets to accurately classify skin cancer. Its depth, when combined with residual learning, allows it to capture complex features and patterns in skin textures, making it an effective method for the early detection and diagnosis of dermatological conditions.

4.2.2 DenseNet121

DenseNet121 [31] is a deep CNN model with 121 layers that has dense connectivity pattern, wherein each layer takes inputs from all its respective previous layers and passes on its feature maps to all its following layers. This architecture forces nodes collectively to meet at a given feature before they can classify it, which encourages efficient feature reuse, alleviates the vanishing gradient issue and results in enhanced parameter efficiency. DenseNet121 has been shown to perform well in skin disease classification in multi-class cases. It's learned to recognize small details in a variety of skin lesions, and this is especially useful for looking at medical images, from other conditions of the skin such as melanocytic nevi and melanoma. After transfer learning, DenseNet121 has good classification performance and serves as a commonly used baseline for classifying dermoscopic images.

4.2.3 MobileNetV2

MobileNetV2 [32] is a lightweight CNN specifically tailored for mobile and embedded apps. It ensures the trade-off between time and accuracy of classification. It employs proprietary building blocks called inverted residual blocks and linear bottlenecks that make it possible for the model to maintain its capacity to comprehend data, whilst demanding fewer resources and processes. It is very useful for deploying deep learning models to embedded devices. As used on the HAM10000 dataset, MobileNetV2 offered an excellent performance in skin lesions identification with significantly reduced computer costs than that of more elaborate models. Even with the small size, it is capable of gathering substantial imaging features from the high-resolution dermatoscopic images, achieving promising classification performances in many skin disease categories. Due to its efficiency, MobileNetV2 is suitable for real-time diagnostic applications, such as smartphone-based skin analysis systems.

Table 1: Hyperparameter used for ResNet50v2, DenseNet121 & MobileNetV2

Hyperparameter	Value
Batch Size	16
Optimizer	Adam
Learning Rate	1e-5
Loss Function	Categorical Cross Entropy
Epochs	32, with early stopping
Activation Function	ReLU

The values of hyperparameters were selected according to the common practices in implementation of deep learning for image classification. A summary of the hyperparameters is visualized in Table 1. Maintaining the same configuration for each model allowed for a fair and reliable comparison of their effectiveness. Fine-tuning was more or less fine-tuned foranity for the ResNet50V2, EfficientNetB4, and DenseNet169 models for the skin disease classification. First, we froze their lower layers to keep the generalized feature representations obtained during pretraining. Twofold hidden layers were fine-tuned to capture more lesion-dependent properties. Two new layers were additionally applied to each base model, consisting of a GlobalAveragePooling2D layer to reduce the features dimensionality, and a dense multi-layer with six nodes output through softmax function to predict several classes. This approach mixed both the feature learning and model generalization, and was employed to avoid overfitting issues as well as for maintaining computational efficiency. Throughout all models, joint parameters were used for training to ensure consistent and optimal performance.

4.3 Ensemble Method

Ensemble learning is an effective method that integrates the outputs of different models to produce improved outcomes compared to employing one individual model. In this work, a fuzzy ranking-based ensemble approach incorporating the Gompertz function is employed. This technique benefits from assigning adaptive weights to each classifier based on their confidence scores, thereby enhancing the accuracy of final predictions. The Gompertz function plays a key role by dynamically integrating the decision outputs of the base models. This results in improved classification accuracy without the need to manually adjust weights for different test datasets, as the process adapts automatically. The Gompertz function [33] is defined as follows:

$$f(t) = pe^{\{-e^{\{q-wt\}}\}} \tag{1}$$

In the Gompertz function, the parameter q controls the horizontal shift along the x-axis, p represents the upper asymptote, e denotes Euler's number, and w serves as a scaling factor along the y-axis. For the purpose of skin disease diagnosis, a re-parameterized Gompertz function [34] is applied to calculate the fuzzy ranking of each classifier. Given a test set from the image dataset, each image will have X prediction scores, where X represents the number of ensemble models. In this study, we employed three pre-trained transfer learning models, so X = 3. Therefore, for each image, we obtain a set of decision scores from the classifiers, denoted as $\{DS_1, DS_2, \ldots DS_X\}$. If Z is the number of distinct classes, then:

$$\sum_{z=1}^{Z} DS(n)_z = 1 \tag{2}$$

Where z=1, 2...Z and n=1, 2, 3...X.

As described in Equation (5), the final decision score for a data instance is obtained by multiplying the Fuzzy Rank Score (FRSc) and the Classifier Confidence Score (CCSc), and then selecting the minimum value among all classes as the final prediction.

$$FRSc = \sum_{i=1}^{M} \begin{cases} R_C^i, & if\ R \in k^i \\ P_C^R & otherwise \end{cases} \tag{3}$$

$$CCSc = \frac{1}{M} \sum_{i=1}^{M} \begin{cases} CF_C^i, & if\ R \in k^i \\ P_C^{CF} & otherwise \end{cases} \tag{4}$$

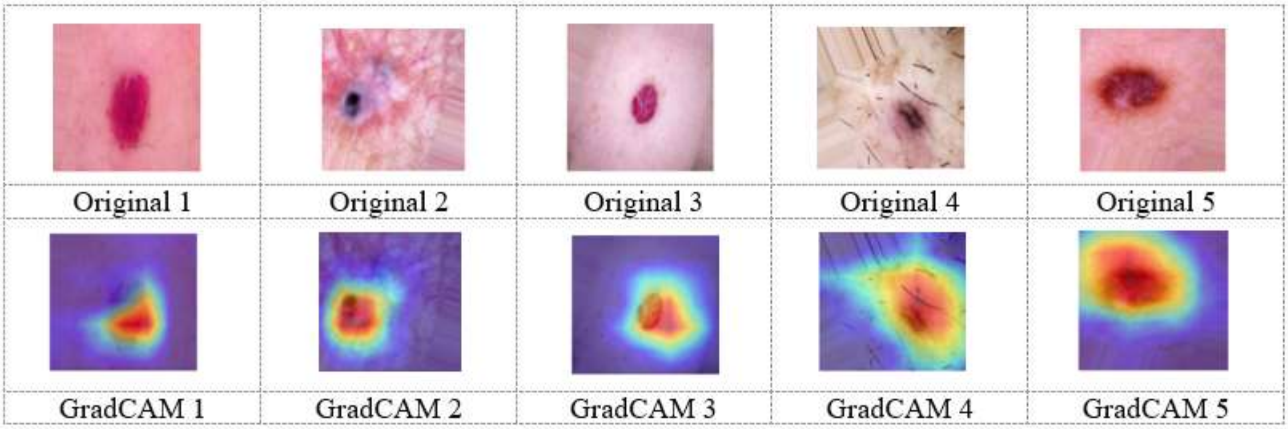
$$class\ (P) = \min \{FSc \times CSc\} \tag{5}$$

Where c=1, 2, 3…. C

4.4 Explainable AI

It is essential for healthcare professionals to have insight into the rationale behind the predictions made by classification models, especially in sensitive domains like medical diagnosis. Despite the impressive classification performance of models like ResNet50V2, DenseNet121, and MobileNetV2, they are often known for their "black box" nature, which lacks interpretability and clinical trust. To mitigate the issue, "Gradient-weighted Class Activation Mapping (Grad-CAM)" [13] is utilized as explanation tool. Grad-CAM generates a dense localised map that highlights the image regions that are most significant to the model's prediction by utilising the gradients of a specific target class that are fed into the final convolutional layer. The methodology includes the computation of the gradients of the loss function in relation to the convolutional feature maps, the application of the ReLU activation to both the gradients and the feature maps to obtain guided gradients, and the subsequent averaging of these gradients to determine the importance weights. Class activation map is generated using the weight to calculate a weighted sum of the corresponding feature maps and produces a class activation map. The generated heatmap is normalised, resized, and placed on the original image with a blending factor of 0.8, thereby enabling the visual interpretation of the model's decision-making process.

Figure 5: GradCAM visualization of various classes



5. Result Analysis and Visualization

The performance of the proposed model is evaluated using a range of recognised classification metrics when the training procedure is completed. An in-depth evaluation of the model's effectiveness is conducted using accuracy, precision, recall, and F1-score, particularly when dealing with unbalanced datasets such as HAM10000. These metrics offer important information about how well the model can deal with class imbalances and classify instances.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{7}$$

$$\text{F1} - \text{score} = \frac{2\times\text{precision}\times\text{recall}}{\text{precision+recall}} \tag{8}$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FN+TN+FP}} \tag{9}$$

where,

- “TP (True Positive): correctly classified positive samples”
- “FP (False Positive): negative samples incorrectly classified as positive”
- “FN (False Negative): positive samples incorrectly classified as negative”
- “TN (True Negative): correctly classified negative samples”

5.1 Performance Evaluation

Standard classification metrics were used to assess the effectiveness of the suggested Fuzzy Ranking Fusion model, including precision, recall, F1-score, and total accuracy. The model's performance was evaluated using the HAM10000 dataset, which has seven different skin lesion categories. For evaluation purpose, 1,400 validation samples were used for each class, for a total of 9,800 instances. With an overall accuracy of 97.00%, the model showed strong and reliable performance across all lesion categories. Table 2 provides a comprehensive analysis of the per-class evaluation metrics.

Table 2: Detailed per-class metrics

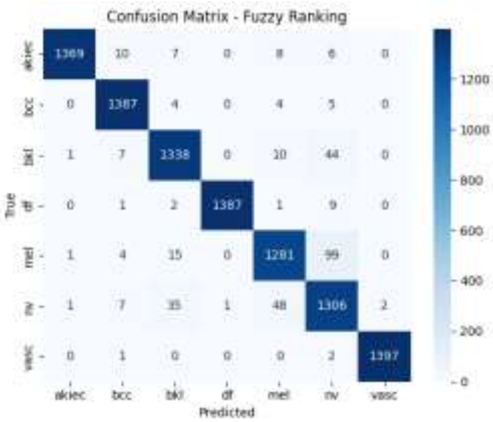
Class	Precision	Recall	F1-score
akiec	1.00	0.98	0.99
bcc	0.98	0.99	0.98
bkl	0.96	0.96	0.96
df	1.00	0.99	0.99
mel	0.95	0.92	0.93
nv	0.89	0.93	0.91
vasc	1.00	1.00	1.00

Both the macro-average and weighted-average F1-scores attained a value of 0.97, indicating that the proposed model maintained consistent performance across all classes, regardless of class imbalance

5.2 Confusion Matrix Analysis

The confusion matrix (Figure 6) offers a detailed view of the model's performance on a per-class basis, where the high values along the diagonal represent accurate classifications. The model demonstrates minimal misclassification across the majority of classes. Notably, Melanoma (mel) is occasionally confused with Melanocytic Nevi (nv), with 99 such instances, reflecting a common challenge due to their visual similarity. Additionally, Benign Keratosis-like lesions (bkl) exhibit some confusion with both nv and mel, indicating overlapping or borderline visual characteristics. Despite these minor errors, the confusion matrix confirms the model's robust discriminative ability across all skin lesion categories.

Figure 6: Confusion Matrix



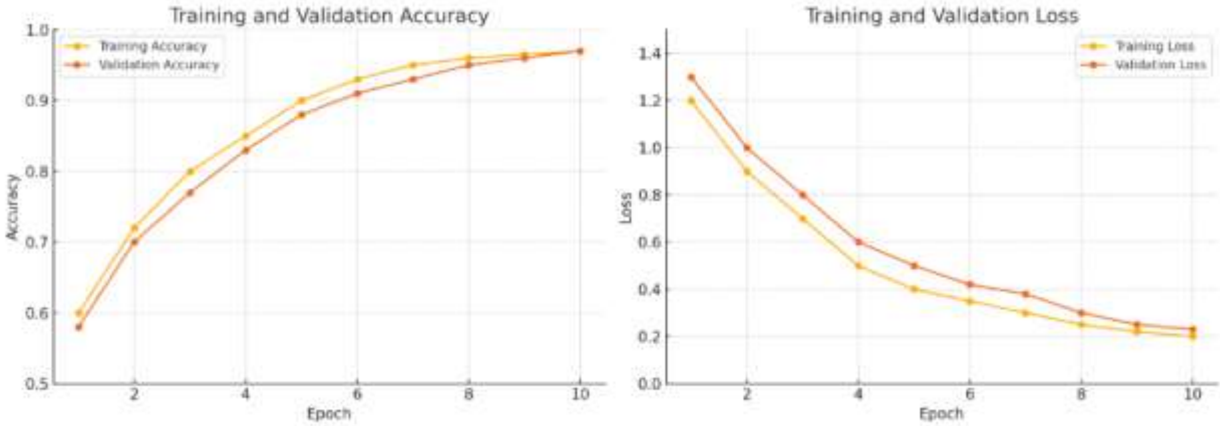
5.3 Training and Validation Trends

Figure 7 presents the training and validation accuracy and loss curves over 10 epochs. Both metrics exhibit consistent improvement with minimal overfitting.

- Accuracy Curve**:** Validation accuracy rises steadily, closely tracking training accuracy, and converges near 97% by epoch 10.
- Loss Curve**:** Training and validation loss both decrease smoothly, with validation loss maintaining proximity to training loss, further confirming effective generalization.

These trends suggest that the model architecture, along with the Fuzzy Ranking fusion strategy, facilitates both high learning capacity and generalizability.

Fig 7: Training and validation accuracy **and** loss **curves**



5.4 Result Comparison

Table 3 provides a comparative overview of the proposed model's accuracy alongside several existing models, using the HAM10000 and ISIC 2018 datasets. It lists models such as Xception, InceptionResNetV2, MobileNetV2, ensemble CNN-SVM, Random Forest, VM, AlexNet, MLPN, VGG16, ResNet50, DenseNet201, DenseNet121, InceptionV3, ResNet50V2, and others, along with their respective accuracy scores. Remarkably, the proposed model achieved the highest accuracy of 97.00% on the HAM10000 dataset, outperforming all other models in the comparison. This strong performance highlights the model's effectiveness in accurately classifying skin lesions.

Table 3: Result Comparison

References	Dataset	Methods Used	Ensemble approach	Accuracy	Year	xAI
[40]	HAM10000	Xception, InceptionResNetV2 and MobileNetV2	Fuzzy logic	95.14%	2025	Yes
[36]	HAM10000	Ensemble CNN-SVM	Weighted average	92.00	2025	No
[37]	HAM10000 and ISIC 2018	Random Forest, MLPN, SVM	Max voting	94.70	2025	No
[38]	HAM10000	VGG16, Inception-V3, and ResNet-50	Weighted average	96.00	2024	No
[35]	ISIC 2018	Ensemble learning, DenseNet-201, Lasso	-	87.72	2024	No
[29]	HAM10000	MobileNetV2, ResNet18, and VGG11	Stacking	86.70	2024	No
[28]	HAM10000	Random Forest, CatBoost, AdaBoost, Extra Trees, and Gradient Boosting	Max voting	95.80	2024	No

Proposed Method	HAM10000	ResNet50v2, DenseNet121,MobileNetV2	Fuzzy Ranking	97.00%	-	Yes

6. Conclusion

This study integrates the prediction abilities of ResNet50V2, DenseNet121, and MobileNetV2 to provide a novel deep learning ensemble strategy for skin lesion classification. The model successfully combines the results from the several classifiers by using fuzzy ranking based on the Gompertz function, producing predictions that are more reliable and accurate. Data augmentation approaches were used to address the prevalent problem of class imbalance in medical imaging, which greatly enhanced the model's performance. The suggested ensemble's remarkable 97.00% accuracy on the HAM10000 dataset demonstrated its potential for practical clinical use. Furthermore, by demonstrating that the model concentrates on the actual lesion areas when making decisions, Grad-CAM visualizations provide interpretability, which is crucial for building confidence among medical practitioners. It is still difficult to differentiate between superficially identical diseases like Melanoma and Melanocytic Nevi, which highlights the need for bigger datasets and continuous model improvement. Overall, the study shows that explainable AI, fuzzy logic, and deep learning can be used to create more transparent and dependable diagnostic systems. These technologies can be particularly helpful in situations where dermatologists are hard to reach, providing efficient assistance with early detection and treatment planning. In the future, adding advanced algorithms to the ensemble and putting it to the test in actual clinical settings may improve its usefulness and efficacy even more.

References

1. Esteva A, Kuprel B, Thrun S (2015) Deep networks for earlystage skin disease and skin cancer classification. project report WHO, "World Health Organization: Cancer," 2018. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer.
2. A.-R.A. Ali, T.M. Deserno, A systematic review of automated melanomadetection in dermatoscopic images and its ground truth data, in: Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment, Vol. 8318, International Society for Optics and Photonics, 2012, p. 83181I..
3. J.L.G. Arroyo, B.G. Zapirain, Automated detection of melanoma in dermoscopic images, in: Computer vision techniques for the diagnosis of skin cancer, Springer, 2014, pp. 139–192
4. A. Madooei, M.S. Drew, Incorporating colour information for computer-aided diagnosis of melanoma from dermoscopy images: A retrospective survey and critical analysis, International journal of biomedical imaging 2016
5. Noviandy, T. R., Nainggolan, S. I., Raihan, R., Firmansyah, I., and Idroes, R. (2023). Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach, Infolitika Journal of Data Science, Vol. 1, No. 2, 48–55. doi:10.60084/ijds.v1i2.123.
6. Mascarenhas, M., Afonso, J., Ribeiro, T., Andrade, P., Cardoso, H., and Macedo, G. (2023). The Promise of Artificial Intelligence in Digestive Healthcare and the Bioethics Challenges It Presents, Medicina, Vol. 59, No. 4, 790. doi:10.3390/medicina59040790.
7. Noviandy, T. R., Idroes, G. M., Syukri, M., and Idroes, R. (2024). Interpretable Machine Learning for Chronic Kidney Disease Diagnosis: A Gaussian Processes Approach, Indonesian Journal of Case Reports, Vol. 2, No. 1, 24–32. doi:10.60084/ijcr.v2i1.204.

8. Hidayat, T., Hadinata, E., Damanik, I. S., Vikki, Z., and Irvanizam, I. (2023). Implementation of Hybrid CNN-XGBoost Method for Leukemia Detection Problem, Infolitika Journal of Data Science, Vol. 1, No. 1, 15–21. doi:10.60084/ijds.v1i1.87.
9. Li, Z., Koban, K. C., Schenck, T. L., Giunta, R. E., Li, Q., and Sun, Y. (2022). Artificial Intelligence in Dermatology Image Analysis: Current Developments and Future Trends, Journal of Clinical Medicine, Vol. 11, No. 22, 6826. doi:10.3390/jcm11226826.
10. Yeasmin, M. N., Al Amin, M., Joti, T. J., Aung, Z., and Azim, M. A. (2024). Advances of AI in image-based computer-aided diagnosis: A review, Array, Vol. 23, 100357. doi:10.1016/j.array.2024.100357.
11. Suganyadevi, S., Seethalakshmi, V., and Balasamy, K. (2022). A Review on Deep Learning in Medical Image Analysis, International Journal of Multimedia Information Retrieval, Vol. 11, No. 1, 19–38. doi:10.1007/s13735-021-00218-1.
12. Holzinger, A., Biemann, C., Pattichis, C. S., and Kell, D. B. (2017). What Do We Need to Build Explainable AI Systems for the Medical Domain?, ArXiv Preprint ArXiv:1712.09923.
13. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, 2017 IEEE nternational Conference on Computer Vision (ICCV), IEEE, 618 626. doi:10.1109/ICCV.2017.74.
14. M. Manandhar, S. Hawkes, K. Buse, E. Nosrati, and V. Magar, ‘‘Gender, health and the 2030 agenda for sustainable development,’’ Bull. World Health Org., vol. 96, no. 9, pp. 644–653, Sep. 2018.
15. R. Erol, Skin Cancer Malignancy Classification With Transfer Learning. Conway, AR, USA: Univ. Central Arkansas, 2018.
16. A. F. Sonmez, S. Cakar, F. Cerezci, M. Kotan, ˙I. Deliba so˘glu, and G. Cit, “Deep Learning-Based Classification of Dermoscopic Images for Skin Lesions,” Sakarya University Journal of Computer and Information Sciences, vol. 6, no. 2, pp. 114–122, 2023.
17. M. Sandhua, A. Hussain, D. Al-Jumeily, B. M. Mahmmod, and S. H. Abdulhussain, “Deep learning-based skin cancer identification,” in 2023 15th International Conference on Developments in eSystems Engineering (DeSE). IEEE, 2023, pp. 404–410.
18. T. Chowdhury, A. R. S. Bajwa, T. Chakraborti, J. Rittscher, and U. Pal, ``Exploring the correlation between deep learned and clinical features in melanoma detection," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Cham, Switzerland: Springer, 2021, pp. 3_17.
19. P. Tschandl, C. Rosendahl, and H. Kittler, ``The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, pp. 1_9, Dec. 2018.
20. F. Nunnari, M. A. Kadir, and D. Sonntag, ``On the overlap between grad-CAM saliency maps and explainable visual features in skin cancer images," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extrac- tion*. Cham, Switzerland: Springer, 2021, pp. 241_253.
21. Thwin, S.M.; Park, H.S. Skin Lesion Classification Using a Deep Ensemble Model. Appl. Sci. 2024, 14, 5599. [CrossRef]
22. Nigar, Natasha, et al. "A deep learning approach based on explainable artificial intelligence for skin lesion classification." *IEEE Access* 10 (2022): 113715-113725.
23. S. P. G. Jasil and V. Ulagamuthalvi, ‘‘Deep learning architecture using transfer learning for classification of skin lesions,’’ J. Ambient Intell. Humanized Comput., vol. 21, pp. 1–8, Mar. 2021.
24. A.Alhudhaif,B.Almaslukh,A.O.Aseeri,O.Guler,andK.Polat,‘‘Anovel nonlinear automated multi-class skin lesion detection system using soft attention based convolutional neural networks,’’ Chaos, Solitons Fractals, vol. 170, May 2023, Art. no. 113409.

25. Torrey, Lisa, and Jude Shavlik. "Transfer learning." In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, pp. 242-264. IGI global, 2010.
26. Z. Rahman, M. S. Hossain, M. R. Islam, M. M. Hasan, and R. A. Hridhee, ''An approach for multiclass skin lesion classification based on ensemble learning,'' Informat. Med. Unlocked, vol. 25, Sep. 2021, Art. no. 100659.
27. Hayat, S.N.; Indraswari, R. Skin Cancer Detection Approach Using Convolutional Neural Network Artificial Intelligence. Int. J. Inform. Inf. Syst. 2024, 7, 46–54. [CrossRef]
28. Natha, P.; RajaRajeswari, P. Advancing Skin Cancer Prediction Using Ensemble Models. Computers 2024, 13, 157. [CrossRef]
29. Liu, Xiaoyi, Zhou Yu, Lianghao Tan, Yafeng Yan, and Ge Shi. "Enhancing skin lesion diagnosis with ensemble learning." In *2024 4th International Conference on Computer Science and Blockchain (CCSB)*, pp. 94-98. IEEE, 2024.
30. He,K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. arXiv 2015, arXiv:1512.03385.
31. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. arXiv 2016, arXiv:1608.06993.
32. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation. arXiv 2018, arXiv:1801.04381.
33. Trappey, C.V.; Wu, H.Y. An evaluation of the time-varying extended logistic, simple logistic, and Gompertz models for forecasting short product lifecycles. Adv. Eng. Inform. 2008, 22, 421–430. [CrossRef]
34. Tjørve, Kathleen MC, and Even Tjørve. "The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family." *PloS one* 12, no. 6 (2017): e0178691.
35. Hosseinzadeh M, Hussain D, Zeki Mahmood FM, A Alenizi F, Varzeghani AN, Asghari P, Darwesh A, Malik MH, Lee SW. A model for skin cancer using combination of ensemble learning and deep learning. PLoS One. 2024 May 31;19(5):e0301275. doi: 10.1371/journal.pone.0301275. PMID: 38820401; PMCID: PMC11142560.
36. Asaduzzaman, Abu, Christian C. Thompson, Fadi N. Sibai, and Md J. Uddin. "Application of ensemble learning models in computer-aided diagnosis of skin diseases." *Neural Computing and Applications* (2025): 1-17.
37. Natha, Priya, Sivarama Prasad Tera, Ravikumar Chinthaginjala, Safia Obaidur Rab, C. Venkata Narasimhulu, and Tae Hoon Kim. "Boosting skin cancer diagnosis accuracy with ensemble approach." *Scientific Reports* 15, no. 1 (2025): 1290.
38. Thwin, Su Myat, and Hyun-Seok Park. "Skin lesion classification using a deep ensemble model." *Applied Sciences* 14, no. 13 (2024): 5599.
39. Hossain, Md Mamun, Md Moazzem Hossain, Most Binoee Arefin, Fahima Akhtar, and John Blake. "Combining state-of-the-art pre-trained deep learning models: A noble approach for skin cancer detection using max voting ensemble." *Diagnostics* 14, no. 1 (2023): 89.
40. Halder, Arindam, Anogh Dalal, Sanghita Gharami, Marcin Wozniak, Muhammad Fazal Ijaz, and Pawan Kumar Singh. "A fuzzy rank-based deep ensemble methodology for multi-class skin cancer classification." *Scientific Reports* 15, no. 1 (2025): 6268.