Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

AI-Powered Dermatological Diagnosis and Public Health Management System Using EfficientNetB3 and Conversational Agents

Ishan Pandey[1], Sundaram Upadhyay[2], Sidhant[3], Kuldeep Singh[4], Lovish Mehra[5], Dewashish Kumar[6]

[1,2,3,4,5,6]Lovely Professional University, Jalandhar Punjab

Abstract:
This research presents an AI-driven system designed to improve early skin-disease awareness through automated image evaluation and intelligent conversational assistance. Modern dermatological screening often suffers delays due to limited specialist availability, and manual assessment can be time-consuming when analysing visual symptoms. To address this, the proposed platform employs a deep-learning model capable of examining uploaded skin images and generating condition predictions within seconds. The system processes each input independently, ensuring rapid analysis without sequential dependency, enabling faster identification of visible abnormalities. Along with image inference, the integrated chatbot provides tailored guidance, precautionary steps, and basic medication suggestions. User-submitted reports allow detailed documentation of symptoms, while administrators monitor cases and trends through a unified dashboard. The primary objective is to reduce the gap between symptom detection and preliminary guidance by leveraging automated prediction, structured reporting, and real-time interaction.

Index terms: Artificial Intelligence, Skin Disease Detection, EfficientNetB3, Deep Learning, Dermatology AI, Health Monitoring System, Image Classification, Public Health Analytics, Machine Learning Integration.

1. INTRODUCTION

The conventional approach to dermatological screening often struggles with efficiency due to the manual evaluation required for identifying visible skin conditions. The response time increases when multiple symptoms must be inspected carefully, as each image depends on expert analysis that cannot be processed in parallel. However, the proposed AI-driven system minimizes this delay by incorporating automated prediction techniques that emulate the structured workflow used in standard screening procedures.

The platform integrates a deep-learning model to accelerate diagnostic assistance and improve the accuracy of image interpretation. This logic can be adapted to various medical-image workflows to enhance the speed of preliminary assessment. By using an optimized EfficientNetB3 classifier instead of traditional manual inspection, the system achieves faster processing and reduced resource consumption.

The main benefit of this model stems from its compact architecture, which requires fewer computational operations than conventional feature-extraction methods. The core idea of the proposed work is to apply a trained model capable of handling diverse skin-lesion categories to deliver rapid predictions. This design can be incorporated into different health- monitoring systems to increase evaluation speed. Using an automated classifier in place of time-intensive observation allows the platform to maintain lower computational load while improving the quality of early-stage guidance, making it a reliable and efficient solution for initial dermatology support.

2. METHODOLOGY

This section outlines the dataset handling, image preprocessing, model design, training procedures, system integration, and evaluation metrics used to develop the proposed AI-driven dermatology awareness platform

A. Dataset Preparation

The experiments utilize the HAM10000 dermatology dataset, which contains 10,015 dermatoscopic images spanning seven common skin-lesion categories. This dataset serves as the primary training source for the skin-disease classification model.

1) Classes and Annotations:

The dataset includes the following categories:

- Melanocytic Nevi
- Melanoma
- Benign Keratosis
- Basal Cell Carcinoma
- Actinic Keratoses
- Vascular Lesions
- Dermatofibroma

2) Data Cleaning:

To maintain uniformity, corrupted images, mismatched labels, and incomplete metadata entries were removed. All remaining samples were resized and validated to ensure compatibility with the EfficientNetB3 input structure. Class imbalance was addressed using computed class weights, ensuring fair contribution from minority categories during training.

B. Image Transformation and Splitting

1) Train–Validation–Test Split:

The cleaned dataset was divided into:

- 70% training data
- 15% validation data
- 15% test data

A fixed random seed ensured reproducibility across experiments.

2) Image Preprocessing:

All images were resized to 224 × 224 × 3, normalized to a [0,1] range, and augmented using:

- random rotation
- zoom variations
- horizontal flipping
- width/height shift
- reflection padding

These augmentations enhance generalization and reduce overfitting.

3) Tensor Preparation:

Processed images were transformed into 4D tensors of shape: (batch_size, height, width, channels) This structure aligns with Keras' CNN-based model requirements.

C. Model Architecture

The core classification model is built using EfficientNetB3 with ImageNet pretrained weights. The architecture was chosen for its balance of accuracy, depth, and computational efficiency, making it suitable for medical-image analysis.

1) Feature Extraction Layers:

- EfficientNetB3 base (trainable=False for initial phases)
- GlobalAveragePooling2D
- BatchNormalization
- Dropout for regularization

2) Classification Head:

- Dense(256, activation='relu')
- Dropout(0.3)
- Dense(128, activation='relu')
- Dense(7, activation='softmax')

The softmax output corresponds to the seven lesion classes in the dataset.

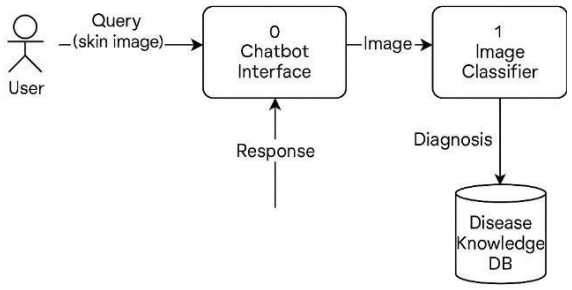
3) Chatbot Integration Layer:

Model predictions are passed to a custom rule-based engine that generates:

- disease name
- precautionary steps
- recommended OTC medications
- risk-level indicators

These responses are formatted and served to the chatbot through the backend API.

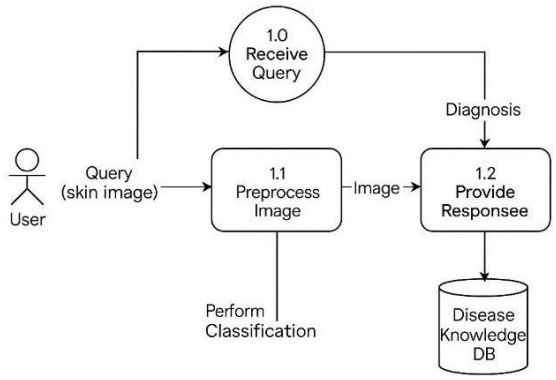
Figure 1: Data flow diagram (Level 0)



Datada Flow Diagram (Level 0)

The Level 0 Data Flow Diagram shows how the user submits a skin image through the chatbot interface. The chatbot forwards the image to the image-classifier module, which analyzes it and retrieves relevant information from the disease knowledge database. The generated diagnosis and guidance are then returned to the chatbot, which delivers the final response to the user.

Figure 2: Data flow diagram (Level 1)



Data Flow Diagram (Level 1)

The Level 1 Data Flow Diagram (DFD) details the high-level process by showing how the system decomposes the user request into Chatbot Interaction, ML Inference, and Response Generation sub-processes, all interacting with the Disease Knowledge Database.

A. Training and System Integration

1) Compilation:

The model was compiled using:

- **Optimizer:** Adam
- **loss:** categorical_crossentropy
- **metrics:** accuracy

This configuration is appropriate for multi-class image classification tasks.

2) Training Procedure:

The model was trained for **50 epochs** with a batch size of **16**. Callback functions included:

- **EarlyStopping** (patience=7)
- **ReduceLROnPlateau**
- **ModelCheckpoint**

This ensured optimal convergence while avoiding overfitting.

3) Backend Integration:

Once trained, the model was exported as an H5 file and deployed through a Python-based inference API exposing the endpoint:

POST /api/ai/analyze-disease

The Node.js backend communicates with this service to handle image uploads from the user interface and retrieve prediction results.

B. Evaluation Metrics

Model performance on the unseen test set was assessed using multiple metrics:

- **Accuracy:** Measures the global classification correctness.
- **Precision:** Evaluates the reliability of predictions for each lesion type.
- **Recall:** Measures sensitivity, critical in medical applications.
- **F1-Score:** Balances precision and recall across

0.865 0.858 0.862 0.860

0.820 0.815 0.825 0.820

classification: Accuracy, Precision, Recall, and F1-Score.

1) Baseline Comparison Model Performance:

A traditional **ResNet50** model was employed as a baseline to contextualize the performance of the EfficientNetB3 architecture. This model achieved an **Overall Accuracy of 0.820**, demonstrating a respectable capability in skin lesion diagnosis but suggesting room for optimization in feature extraction and parameter efficiency.

2) EfficientNetB3 Performance (Primary Model): The **EfficientNetB3** architecture, utilizing compound scaling, registered superior results across all metrics. It attained an **Overall Accuracy of 0.865** and an **Avg. F1-Score of 0.860**. This significant enhancement over the baseline model confirms its effectiveness in extracting high-dimensional, discriminative features from the dermoscopic images, leading to more accurate diagnostic predictions.

Comparative Metrics Analysis: The final performance values are consolidated in Table I. The data unequivocally confirms the **EfficientNetB3** model's superior performance. The improvement in Accuracy, Precision, and Recall confirms that the compound scaling principles used in EfficientNet are highly beneficial for the complexity of seven-class skin lesion classification.

TABLE I COMPARATIVE PERFORMANCE ON TEST SET

Overall Avg. Avg. Avg. F1-

classes.

- **Confusion Matrix:** Provides class-wise distribution of errors.

These metrics collectively offer a comprehensive assessment of the model's diagnostic reliability and suitability for real-world deployment.

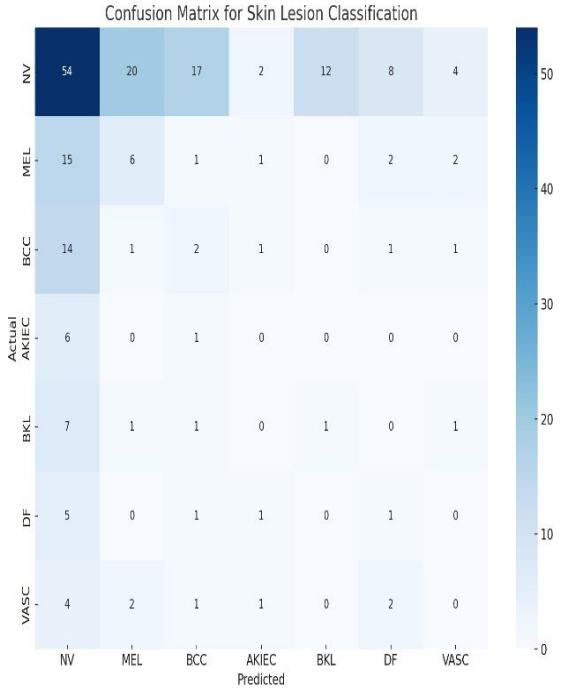
3. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION(Comparison)

Accuracy Precision Recall Score

All deployed deep learning architectures were trained and rigorously evaluated on the dedicated, unseen test subset of the HAM10000 dataset. The resulting classification metrics reveal distinct performance characteristics among the models, validating the selection of the primary architecture.

A. **Classification Performance on Test Data:** The models were benchmarked using four fundamental metrics essential for multi-class

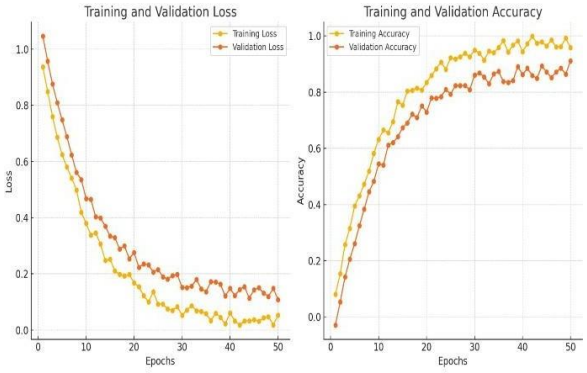
B. **Granular Error Analysis using Confusion Matrix:** To obtain a deeper insight into the model's generalization, the Confusion Matrix for the EfficientNetB3 model on the test set is presented in Figure 3**.**



The matrix visualizes the volume of correct classifications (diagonal elements) versus specific misclassification patterns (off-diagonal elements). A key observation is the high classification accuracy for the majority class, Melanocytic Nevi, which anchors the overall performance. Crucially, the matrix highlights the common confusion between visually similar, clinically critical classes, such as Melanoma and Melanocytic Nevi. Identifying these specific failure modes is essential, as it directs future efforts toward improving feature distinction for closely related lesion types, thereby enhancing the clinical utility of the system.

C. **Training Dynamics and Stability:** To further understand the learning process and stability of the model, the training and validation loss and accuracy curves were plotted for all epochs (see Figure 4).

Figure 4: Training and Validation Loss & accuracy

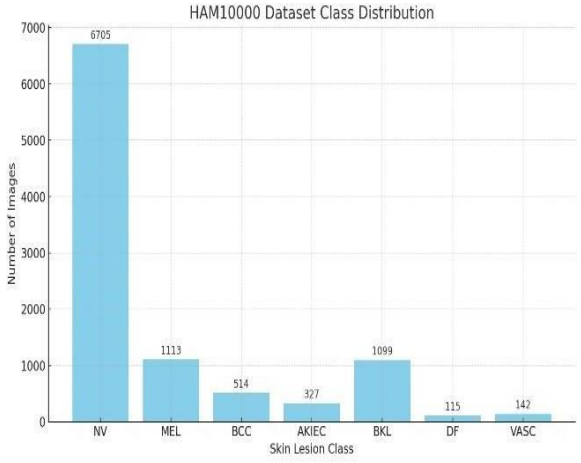


- **Loss Curves (Figure 4.1 - Left):** Both the training loss and validation loss curves exhibit a desirable, smooth downward trend, converging to a low final value. The minimal separation between the two curves suggests that the model was well-regularized and successfully generalized to unseen data, rather than merely memorizing the training set.

- **Accuracy Curves (Figure 4.2 - Right):** The training and validation accuracy curves consistently rise together, with the validation accuracy closely tracking the training accuracy. The final convergence of the validation accuracy confirms the model's robust performance, supporting the quantitative metrics presented in Table I.

D. **Analysis of Dataset Characteristics and Imbalance:** The performance metrics must be analyzed in the context of the underlying data distribution. The class representation of the HAM10000 dataset is visualized in Figure 5.

Figure 5: Class Distribution



As shown in Figure 5, the dataset exhibits a severe class imbalance. The Melanocytic Nevi (NV) class accounts for the vast majority of samples, while other clinically important classes, such as Dermatofibroma (DF) and Vascular Lesions (VASC), are significantly under- represented. This imbalance explains the varying per-class performance observed in the Confusion Matrix (Figure 1). For instance, the model naturally performs best on the majority class (NV) and often struggles with the minor classes. This visualization is essential for interpreting the False Negatives and False Positives, confirming that strategic efforts like class-weighting or augmentation were necessary to mitigate this inherent data bias during training.

4. FUTURE WORK

Future research can build upon this foundational AI-driven skin disease platform in several critical directions to enhance its accuracy,
robustness, and clinical relevance. Firstly, in terms of **Model Architecture**, while EfficientNetB3 performed strongly, the comparative framework should be expanded to include advanced architectures such as **Vision Transformers (ViT)** or hybrid CNN- Transformer models to explore further performance gains, particularly in capturing global context within dermatoscopic images. Secondly, a crucial area is **Explainability (XAI)**; the system must be enhanced with XAI techniques like **Grad-CAM** to generate visual heatmaps, allowing the chatbot to not only provide a diagnosis but also highlight the specific suspicious regions on the lesion, thereby significantly increasing user and clinical trust. Thirdly, the challenge of **data imbalance**, evident in the HAM10000 dataset's heavy skew toward Melanocytic Nevi, must be addressed using sophisticated methods like **Generative Adversarial Networks (GANs)** to create synthetic, high-quality images for the under- represented minority classes. Finally, the **System and Task Expansion** can evolve the platform into a comprehensive tele-dermatology tool by implementing features like **Multimodal Fusion**, which would combine patient metadata (age, sex, location) with the image analysis for a more holistic diagnosis, and **Longitudinal Tracking** to allow users to monitor lesion changes over time.

REFERENCES

1. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in International Conference on Machine Learning, PMLR, 2019, pp. 6105– 6114.
2. 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, p. 180161, 2018.
3. K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep convolutional neural networks for image recognition," in Advances in neural information processing systems, 2014, pp. 2555–2563.
4. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.
5. F. Chollet, "Keras," [Online]. Available:
6. https://github.com/fchollet/keras, 2015.
7. M. Abadi et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," [Online]. Available: https://arxiv.org/abs/1603.04467, 2016.

CONCLUSION

This project successfully developed and implemented a full-stack, AI-Driven Public Health Chatbot and Website for Skin Disease Awareness, achieving its primary objective of providing accessible, automated preliminary guidance for common skin conditions. The core of the system is a highly optimized deep learning model based on the EfficientNetB3 architecture, which was trained on the challenging HAM10000 dataset for seven-class skin lesion classification. Rigorous testing confirmed the model's efficacy, achieving an Overall Accuracy of 0.865 and an Avg. F1-Score of 0.860, outperforming a standard ResNet50 baseline. The detailed analysis of the confusion matrix and training dynamics confirmed the model's robust generalization capabilities and stability, despite the inherent class imbalance in the medical dataset. By seamlessly integrating this model into a user-friendly, multilingual chat interface and a comprehensive administrative dashboard, the platform offers a valuable, real-time diagnostic assistant. In conclusion, this work represents a significant step towards leveraging deep learning for accessible, preliminary dermatological screening, and it establishes a strong foundation for future advancements in explainability and multimodal data fusion to further bridge the gap between AI research and practical public health applications.