Privacy-Preserving Federated Hybrid Ensemble (EfficientNetB3-InceptionV3) for Multi-Class Skin Lesion Classification

Presented by

Amit Kumar Ghosh

ID : 221-15-4650

Department : CSE

Daffodil International University

Outline

- Introduction
- Problem Statement
- Objective
- Background Study
- Gap Analysis
- Methodology
- Results & Analysis
- Novelty of the Work
- Conclusion
- References

Introduction

- **Dermatological Health Burden:** Skin cancer is very common, and harmless-looking spots like moles or brown patches can sometimes hide dangerous melanoma, so doctors need tools that can reliably tell many lesion types apart at an early stage.
- **AI-Assisted Dermoscopy:** Modern deep learning, especially CNN-based models, can now match dermatologists on datasets like HAM10000 and help automatically classify different kinds of skin lesions from dermoscopic images.
- **Privacy-Preserving Collaboration:** In reality, skin images are stored in many hospitals and cannot easily be shared because of strict privacy rules, so federated learning is used to train a shared model while keeping all images inside each institution.
- **Hybrid Federated Ensemble:** This project builds a federated hybrid model that combines EfficientNetB3 and InceptionV3 with an attention module, training on three uneven (non-IID) clients and using class balancing to handle the highly imbalanced HAM10000 dataset.
- **Clinical Impact and Performance:** The system reaches over 93% test accuracy without moving raw patient images and applies Grad-CAM to show which lesion regions influenced the prediction, making it suitable for clinician-in-the-loop use in real dermatology practice.

Problem Statement

- **Centralized Systems:** Privacy-utility tradeoff; vulnerability to data breaches
- **Severe Class Imbalance:** HAM10000 exhibits 58:1 ratio (majority vs minority classes)
- **Non-IID Data Distribution:** Real-world hospital data is heterogeneous and skewed
- **Limited Explainability:** Most federated methods lack interpretability mechanisms
- **Multi-Class Complexity:** 7-category classification (Melanoma, Nevi, Keratosis, BCC, AK, Dermatofibroma, Vascular)
- **Practical Deployment:** Need for clinician-in-the-loop decision support

Objective

- Build a hybrid EfficientNetB3-InceptionV3 ensemble with attention-driven feature fusion for 7-class classification on HAM10000
- Implement federated learning environment with 3 non-IID clients using Dirichlet sampling (α=0.5)
- Mitigate class imbalance using RandomOverSampler (34,573 oversampled training instances)
- Benchmark against strong baselines: VGG-16/19, ResNet-50, DenseNet-121, EfficientNet-B3, Inception-V3, ViT-B/16, GCN, GAT, Hybrid GCN-GAT
- Achieve privacy preservation with modest communication overhead and small accuracy gap vs centralized training
- Enable explainability through Grad-CAM and saliency maps for clinical deployment

Background Study

- **Melanoma:** Most dangerous skin cancer; early detection critical for survival
- **Dermoscopic Imaging:** Non-invasive imaging technique revealing subsurface patterns
- **HAM10000 Dataset:** 10,015 dermoscopic images across 7 diagnostic categories
- **Deep Learning Success:** CNNs match or exceed dermatologist performance in benchmark tests
- **Transfer Learning:** VGG, ResNet, DenseNet, EfficientNet show strong performance on dermoscopic tasks
- **Challenge:** Severe class imbalance and privacy constraints in real clinical settings
- **Federated learning:** Distributed machine learning where data stays on-premise; only model parameters shared
- **FedAvg Algorithm:** Server aggregates client gradients; privacy-preserving aggregation
- **Non-IID Data:** Real-world hospital data exhibits heterogeneous class distributions

Background Study

References	Objectives	Methods	Results	Limitations
Yaqoob et al. [1]	Privacy-aware skin cancer detection using federated learning with asynchronous training	Asynchronous Federated Learning with **CNNs** on HAM10000; local model training with delayed aggregation to handle client availability	**Accuracy: 87.3**%; improved privacy preservation; reduced communication rounds by 40% compared to synchronous FL	Limited to binary classification (melanoma vs non-melanoma); simple CNN architecture; no attention mechanism for interpretability
Hendrix et al. [2]	Evidential federated learning with prompt tuning for uncertainty-aware skin lesion classification	FedEvPrompt integrating Vision Transformers with b-prompts and t-prompts; knowledge distillation on attention maps; evidential learning framework on ISIC2019	**Accuracy: 91.4%;** improved uncertainty quantification; enhanced privacy by sharing attention maps only (not parameters); handles data heterogeneity	Computational overhead of Vision Transformers; requires pretrained ViT models; limited evaluation on HAM10000 multi-class problem
Tian et al. [3]	Communication-efficient federated learning using dataset distillation for skin lesion classification	Generalizable Dataset Distillation (GDD) synthesizing small representative datasets; Gaussian distribution modeling; transmits synthetic images instead of parameters	50× reduction in communication cost; **accuracy: 88.6%;** faster convergence (5 rounds vs 20 for FedAvg)	Synthetic images may lose fine-grained dermatoscopic details; limited to 3-class problem; privacy concerns with sharing synthetic data
Shetty et al. [4]	Multi-class skin lesion classification on HAM10000 using CNN and machine learning	**Custom CNN** architectures with data augmentation; comparison with **SVM, Random Forest, KNN** on HAM10000 seven classes	**CNN accuracy: 89.3%; SVM: 72.4%;** Random Forest: 68.1%; effective feature extraction for pigmented lesions	Centralized training (no privacy preservation); no federated approach; limited generalization to external datasets; class imbalance not addressed
Deng et al. [5]	Federated active learning for efficient annotation in skin lesion classification	Active learning strategy selecting informative samples for annotation; federated setup with uncertainty sampling	Achieved state-of-the-art performance using only 50% of annotated samples; reduced annotation cost by 50%	Requires initial labeled data; uncertainty estimation overhead; limited evaluation on rare lesion classes

Gap Analysis

- Most existing studies focus on binary melanoma vs non-melanoma, while real clinics need reliable seven-class skin-lesion classification on HAM10000.
- Prior federated learning work in dermatology often ignores non-IID client distributions or evaluates only on small or simplified settings.
- Many CNN and ensemble methods are centralized, so they do not protect patient privacy or support multi-institution collaboration.
- Severe class imbalance is usually handled only with simple augmentation, not with explicit oversampling and detailed ablation analysis.
- Very few methods combine a hybrid dual-backbone with attention, federated training, and Grad-CAM–based explainability in a single deployable system.
- Most prior works stop at offline model evaluation and do not provide an end-to-end deployable system or user interface for real clinical use.

Methodology

➢**Data Acquisition:** The HAM10000 dataset provides 10,015 dermatoscopic RGB images of seven clinically diagnosed pigmented skin lesion categories, along with patient and acquisition metadata, which are resized to 300×300 pixels for this study

➢**Data Preprocessing and Augmentation:**

All dermoscopic images are resized to 300×300 pixels in RGB space and normalized using dataset-specific statistics to ensure consistent input for the hybrid EfficientNetB3–InceptionV3 model. During training, random augmentations such as flips, rotations, zoom, and color jitter are applied to increase data diversity and reduce overfitting, especially for minority lesion classes.

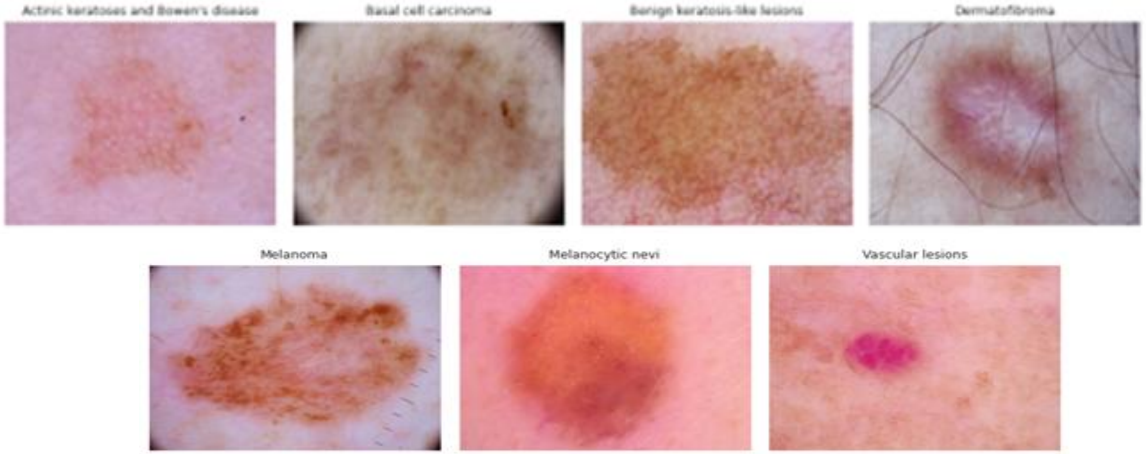

Fig. 1. Representative Dermoscopic Images of the Seven HAM10000 Skin Lesion Categories

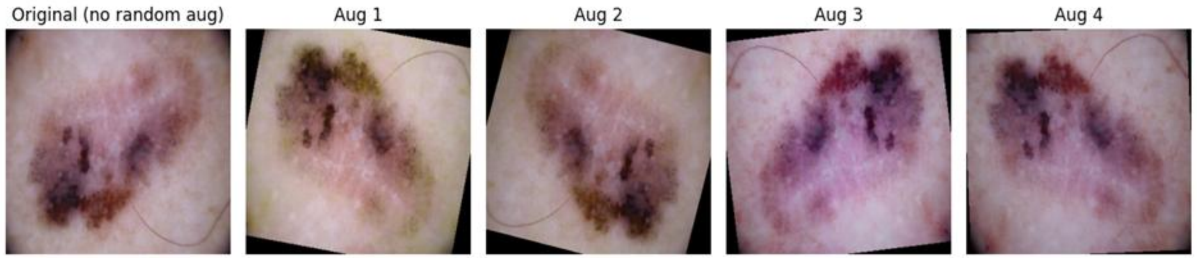

Fig 2: Examples of Random Data Augmentations Applied to a Dermoscopic Image

Methodology

➢**Class Imbalance Handling:**

RandomOverSampler is applied on the training split to expand minority lesion classes and transform the highly skewed HAM10000 distribution into a more balanced dataset of 34,573 images. This helps the federated hybrid ensemble avoid bias toward common benign lesions and improves sensitivity for rare but clinically critical classes such as melanoma and vascular lesions.

➢**Data Splitting:** Dataset divided into 80% training, 10% validation, and 10% testing sets.

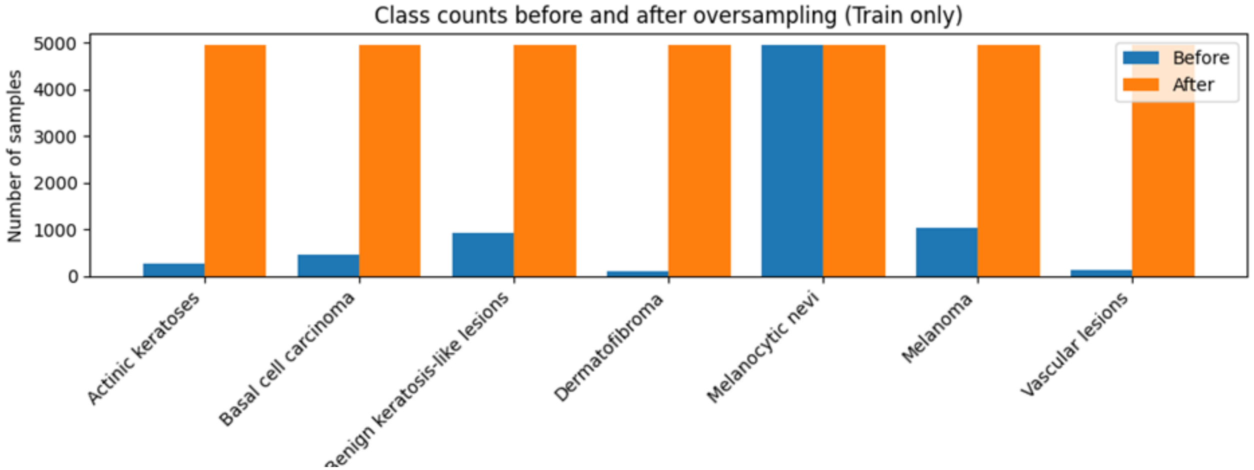

Fig. 3. Training Class Counts Before and After Oversampling

Table I: Training Set Class Counts Before and After Oversampling

Class	Count before oversampling	Count after oversampling
Melanocytic nevi	4,939	4,939
Melanoma	1,021	4,939
Benign keratosis-like lesions	923	4,939
Basal cell carcinoma	444	4,939
Actinic keratoses	267	4,939
Vascular lesions	116	4,939
Dermatofibroma	99	4,939

Proposed Model

- Build a federated hybrid model that combines EfficientNetB3 and InceptionV3 to classify seven types of skin lesions from the HAM10000 dataset.
- Train the model on three separate clients with non-IID data and aggregate them using FedAvg, so hospitals can collaborate without sharing raw images.
- Apply careful preprocessing, data augmentation, and RandomOverSampler to handle class imbalance and improve performance on rare lesions.
- Compare the proposed model with strong baseline networks such as VGG, ResNet, DenseNet, ViT, GCN, and GAT under the same evaluation setup.
- Generate Grad-CAM and saliency visualizations to highlight important lesion regions and support clinician-in-the-loop decision making.
- Package the best global model into an offline demo application for practical skin-lesion risk prediction in real clinical environments.

Proposed Model Architecture

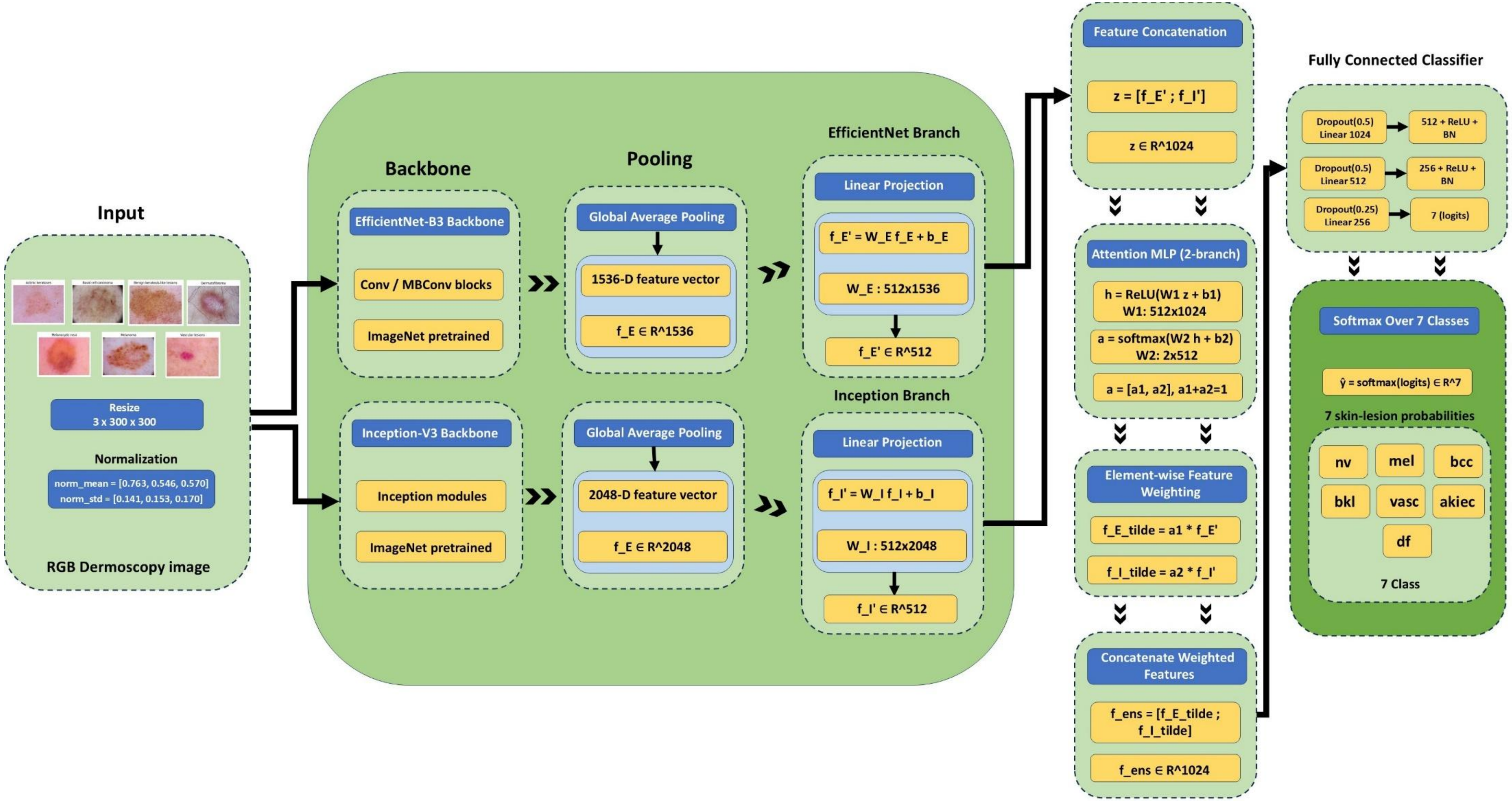

Fig 4: Architecture of Proposed Hybrid EfficientNetB3–InceptionV3 Ensemble

Result and Analysis

- The proposed hybrid EfficientNetB3–InceptionV3 model achieves the best test accuracy of 93.62%, higher than all individual CNN, ViT, and graph-based baselines on the seven-class HAM10000 dataset.
- Multi-class ROC curves show AUC between 0.97 and 1.00 for all lesion types, with perfect separation for Basal cell carcinoma and Dermatofibroma, confirming strong discriminative power.
- The normalized confusion matrix indicates that most samples are correctly classified, with remaining errors mainly between visually similar Melanoma and Melanocytic nevi.
- In the federated setting, the global model reaches 93.0% test accuracy, very close to centralized training, while keeping all patient images on their own clients.
- Ablation results show that oversampling and augmentation significantly boost performance, but the ensemble still outperforms baselines even when these components are removed.
- Grad-CAM, saliency, and SmoothGrad visualizations highlight key lesion regions, giving clinicians an interpretable view of why the model makes each prediction.

Result and Analysis

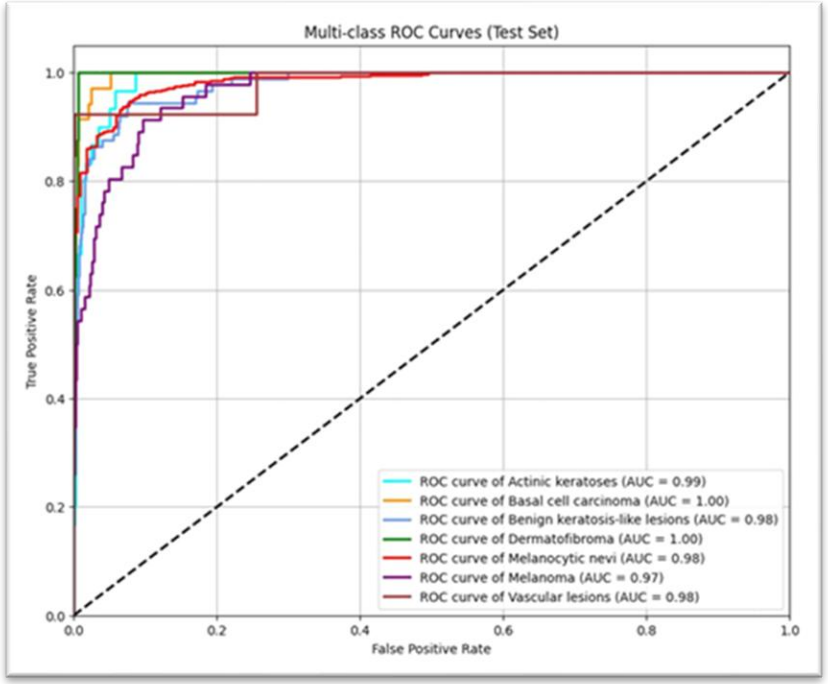

Fig. 5. Multi Class ROC-Curve of Proposed Model

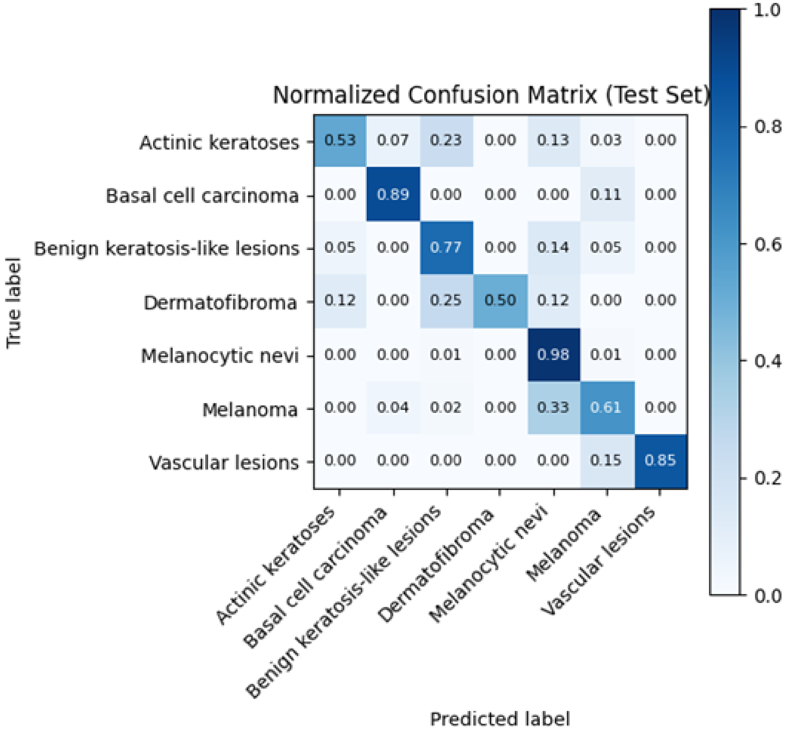

Fig. 6. Confusion Matrix of Proposed Model

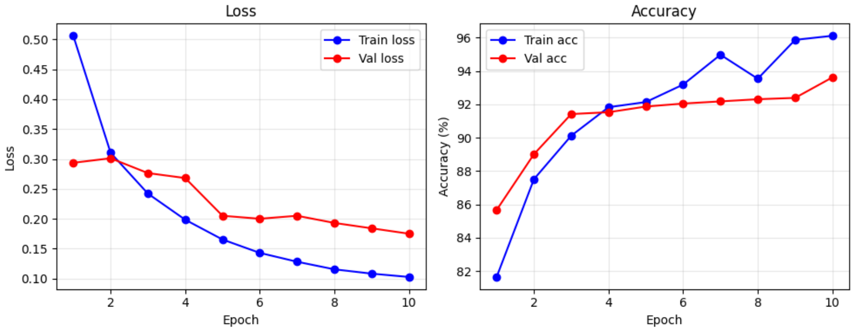

Figure 7. Training and validation loss and accuracy curves

Result and Analysis of Federate Clients

Table II: Proposed model and Baseline model comparison

Model	Accuracy
GCN	62.34%
GAT	61.23%
Hybrid GCN-GAT	54.76%
VGG-16	82.79%
VGG-19	86.23%
ResNet-50	89.67%
DenseNet-121	90%
EfficientNet-B3	91.02%
Inception-V3	89.35%
ViT-b-16	78.45%
Proposed model	**93.62%**

Table III: Federated Client-wise and Global Hybrid Model Accuracy

Component	Train Acc (%)	Test Acc (%)
Client 1	92.78	90.41
Client 2	96.18	93.50
Client 3	97.97	95.09
Global Model (Hybrid EfficientNetB3-InceptionV3 Ensemble Mode)	**94.97**	**93.00**

Fig. 8. Accuracy Comparison Bar Chart of Federate clients and Global Model

Analysis of Ablation Study

➢ This table compares different baseline models and shows that the proposed hybrid ensemble achieves the highest test accuracy (93.62%), and still performs best even when oversampling (SMOTE) or data augmentation are removed.

Table IV: Ablation study Of Proposed and baseline models

Model	Without SMOTE	Without Augmentation	Accuracy
GCN	52%	61%	62.34%
GAT	54%	62%	61.23%
Hybrid GCN-GAT	50%	54%	54.76%
VGG-16	68%	85%	82.79%
VGG-19	67%	84%	86.23%
ResNet-50	66%	88%	89.67%
DenseNet-121	72%	90%	90%
EfficientNet-B3	67%	90%	91.02%
Inception-V3	74%	89%	89.35%
ViT-b-16	59%	77%	78.45%
Proposed model	**80%**	**92.81%**	**93.62%**

Result and Analysis

- Grad-CAM heatmaps from EfficientNet-B3, InceptionV3, and their ensemble show that the hybrid model concentrates on the lesion itself, matching clinically relevant regions.

- Saliency and SmoothGrad highlight the pixels most responsible for the model's decision on a class-4 lesion, with SmoothGrad giving a cleaner focus on the lesion area.

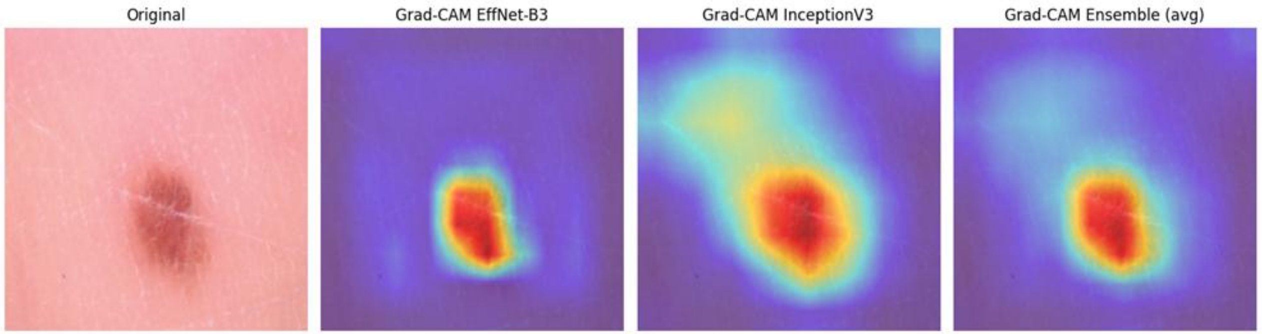

Fig. 9: Grad-CAM skin-lesion visualizations

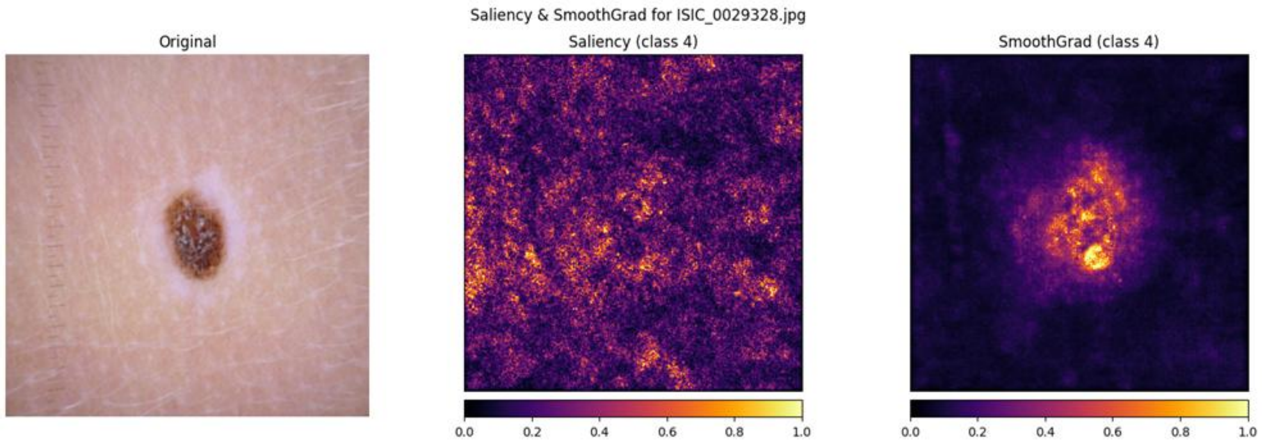

Fig. 10:Saliency and SmoothGrad maps for class-4 skin lesion

Result and Analysis

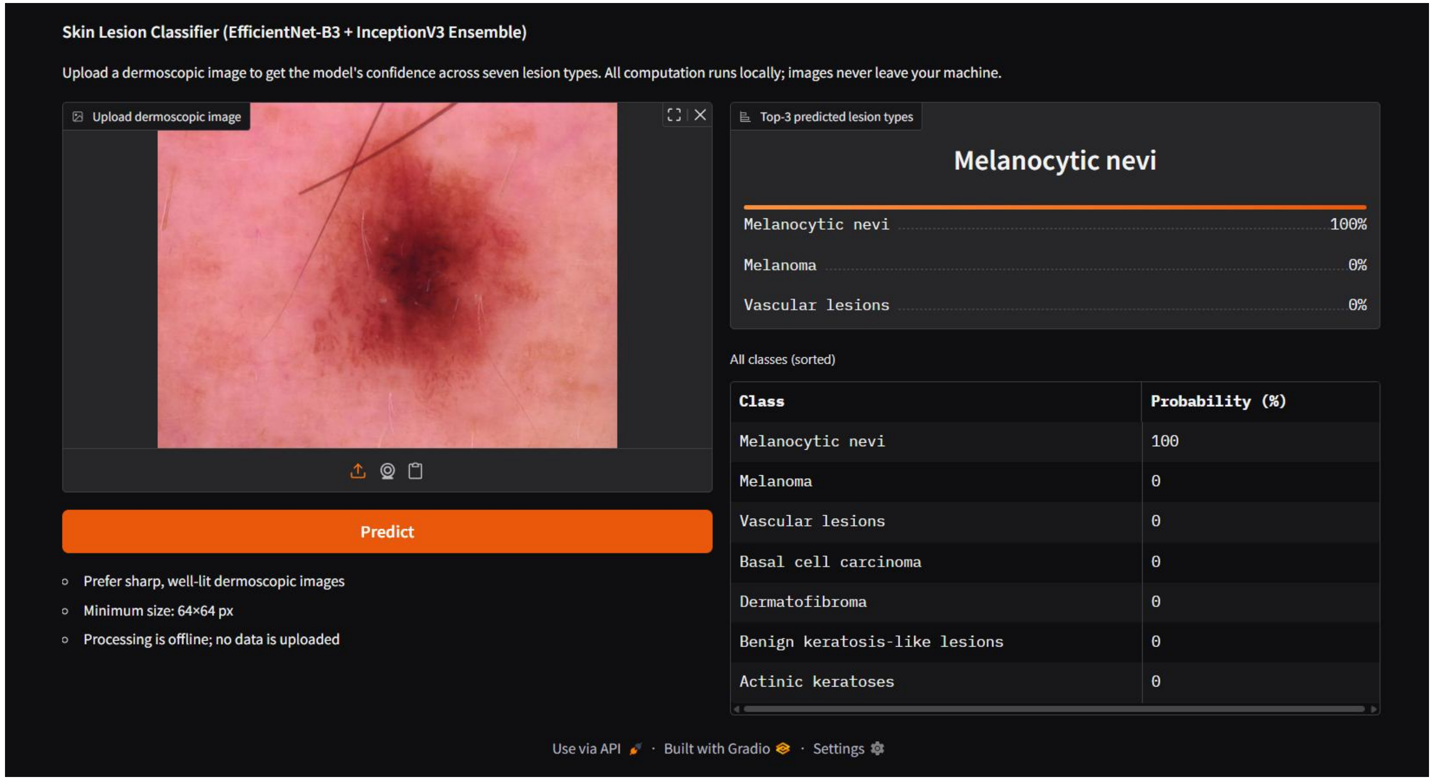

Fig. 11: User Dashboard of Skin Lesion Classifier

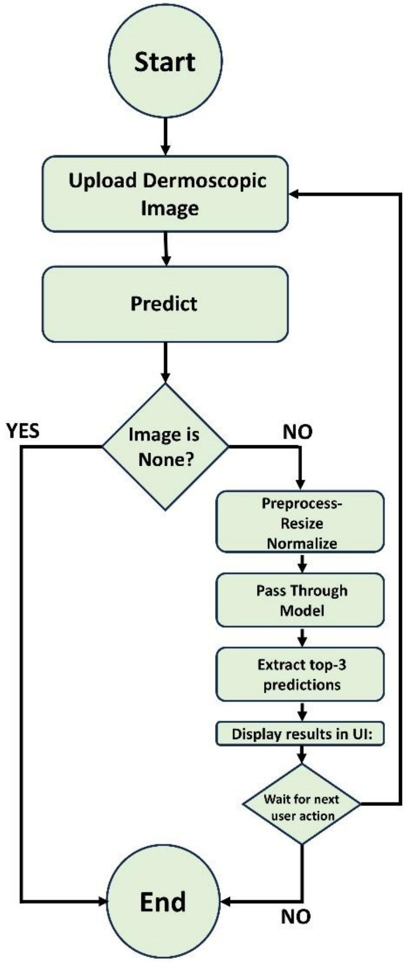

Fig. 12: Activity diagram of the skin-lesion classifier inference workflow

Novelty of the Work

- **Dual-Backbone Ensemble in Federated Setting:** First hybrid EfficientNetB3-InceptionV3 ensemble under federated learning for dermatology
- **Attention-Based Feature Fusion:** Sample-specific weighted fusion improves robustness vs. simple concatenation
- **Non-IID Class Imbalance Handling:** RandomOverSampler + Dirichlet partitioning + FedAvg addresses both challenges simultaneously
- **Comprehensive Multi-Class Benchmark:** 11 baseline comparisons (CNNs, Transformers, Graph-Based) on 7-class HAM10000
- **Explainability in Federated Dermatology:** Grad-CAM + Saliency maps for clinician-facing decision support
- **Practical Multi-Institutional Deployment:** Demonstrates privacy-preserving collaboration with minimal accuracy gap (0.62%)

Conclusion

Our federated hybrid EfficientNetB3–InceptionV3 model achieves over 93% accuracy on seven skin-lesion classes while keeping all images at their own hospitals, showing that privacy-preserving collaboration is both effective and practical. By handling class imbalance and providing Grad-CAM and saliency visualizations, the system offers reliable predictions and clear visual explanations that can support dermatologists in real clinical use and can be extended in future work to larger, multi-center deployments.

References

- [1] Muhammad Mateen Yaqoob, Musleh Alsulami, Muhammad Amir Khan, Deafallah Alsadie, A. Khader, and M. Alkhathami, “Federated Machine Learning for Skin Lesion Diagnosis: An Asynchronous and Weighted Approach,” *Diagnostics*, vol. 13, no. 11, pp. 1964–1964, Jun. 2023, doi: https://doi.org/10.3390/diagnostics13111964.
- [2] R. Hendrix, S. F. Proietto, C. Spampinato, S. Palazzo, and U. Bagci, “Evidential Federated Learning for Skin Lesion Image Classification,” *arXiv (Cornell University)*, Nov. 2024, doi: https://doi.org/10.48550/arxiv.2411.10071.
- [3] Y. Tian, J. Wang, Y. Jin, and L. Wang, “Communication-Efficient Federated Skin Lesion Classification with Generalizable Dataset Distillation,” *Lecture notes in computer science*, pp. 14–24, Jan. 2023, doi: https://doi.org/10.1007/978-3-031-47401-9_2.
- [4] B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, “Skin lesion classification of dermoscopic images using machine learning and convolutional neural network,” *Scientific Reports*, vol. 12, no. 1, p. 18134, Oct. 2022, doi: https://doi.org/10.1038/s41598-022-22644-9.
- [5] Z. Deng, Y. Yang, and K. Suzuki, “Federated Active Learning Framework for Efficient Annotation Strategy in Skin-Lesion Classification,” *Journal of Investigative Dermatology*, Jun. 2024, doi: https://doi.org/10.1016/j.jid.2024.05.023.

Any Questions?

THANK YOU