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

Presented by

Name: Amit Kumar Ghosh
ID : 221-15-4650
Department : CSE
Daffodil International University

Outline

Introduction

- Rising Global skin Cancer Rates
- Subjective Manual Diagnosis
- Automated AI Classification
- Federated learning privacy
- Hybrid Ensemble Architecture
- Clinical Impact and Performance

Problem Statement

- **High Mortality Risks:** Late detection kills
- **Subjective Human Error:** Doctors disagree often
- **Expensive Invasive Biopsies:** Painful and costly
- **Severe class imbalance:** Real world data imbalance
- **Data Privacy Concerns:** Patient data risks
- **Clinical deployment needed:** Health Aid System

Objectives

- Preserve patient privacy via federated learning while maintaining strong predictive performance.
- Deliver a deployable clinical decision-support health aid system.
- Design a dual-backbone ensemble classifier (EfficientNet-B3 + InceptionV3).

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

Feature	Centralized CNN on HAM10000 papers[1][4]	FL melanoma detection papers [2]	Ensemble papers HAM10000 (centralized)	Proposed federated hybrid
7-class HAM10000	Yes	No (binary)	Yes	Yes
Federated learning	No	Yes	No	Yes
Non-IID client distribution modeled	No	Partial	No	Yes
Explicit imbalance handling (oversamp)	Limited	No	Limited	Yes (RandomOverSampler)
Hybrid dual-backbone with attention	No	No	No	Yes
Explainability (Grad-CAM etc.)	Partial	No	Partial	Yes
Systematic ablation studies reported	Limited	No	Limited	Yes (imbalance, aug.)
Deployment and system implementation	Prototype / research code	Prototype only	Prototype only	Fully functional offline desktop system (no internet required)

Methodology

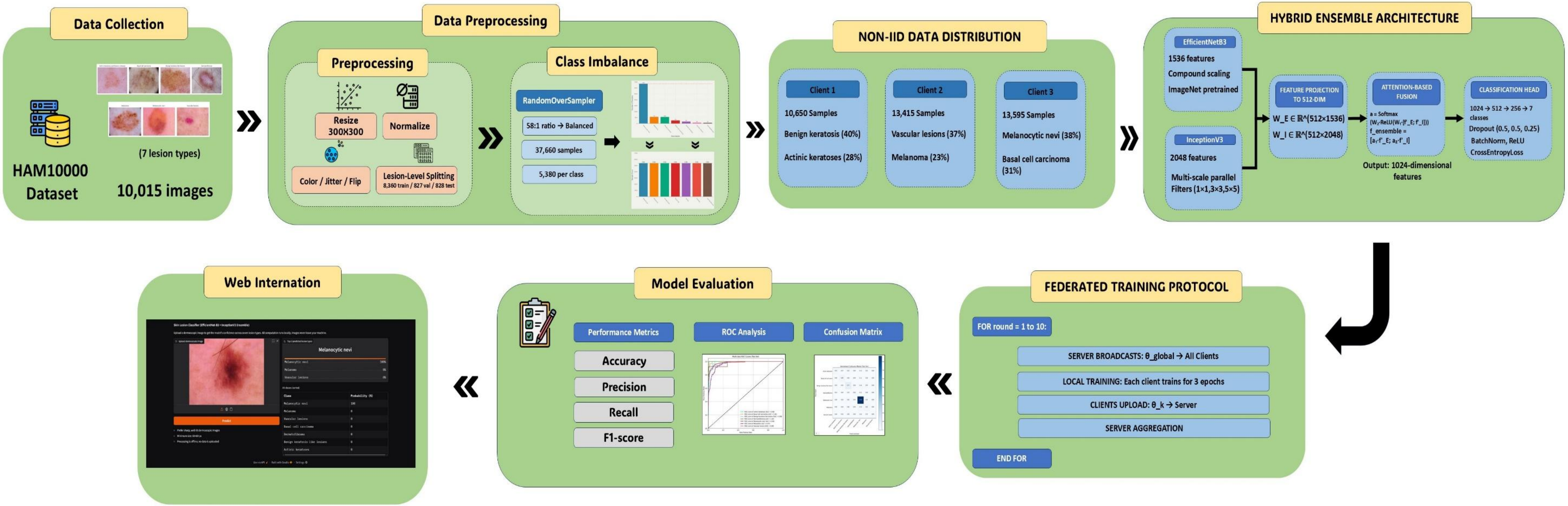


Fig 1: Proposed Federated Learning Methodology Workflow

Methodology

➢**Data Acquisition:** The HAM10000 dataset provides 10,015 dermatoscopic RGB images [6].

➢**Data Preprocessing and Augmentation:**

All dermoscopic images are resized to 300×300 pixels in RGB space and normalized using dataset-specific statistics. During training, random augmentations such as flips, rotations, zoom, and color jitter are applied.

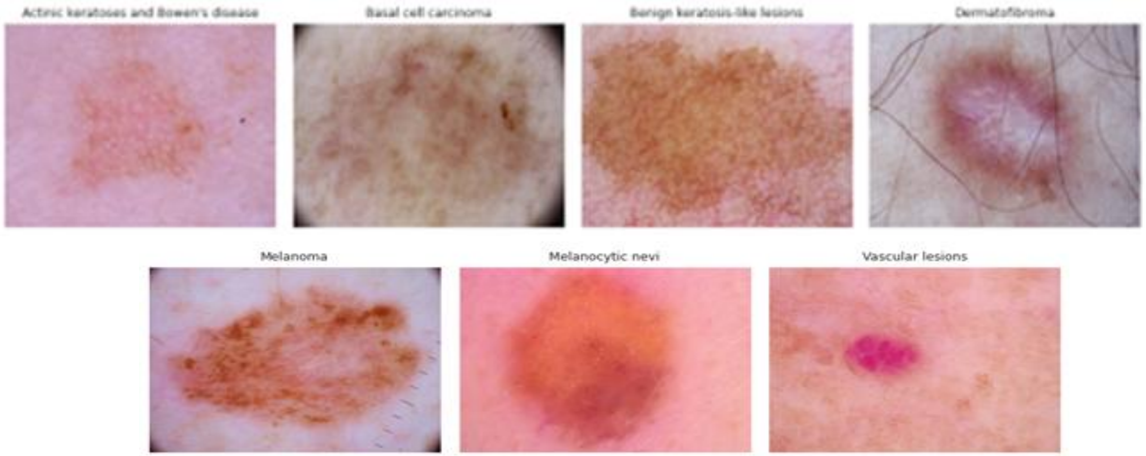


Fig. 2. Representative Dermoscopic Images of the Seven HAM10000 Skin Lesion Categories

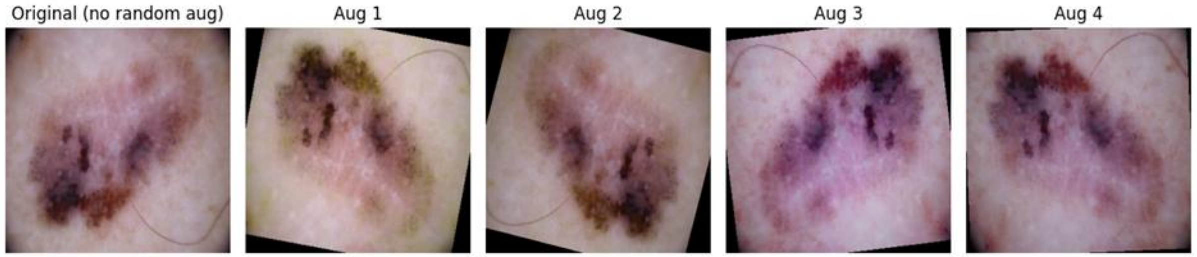


Fig 3: Examples of Random Data Augmentations Applied to a Dermoscopic Image

Methodology

➢**Class Imbalance Handling:** The HAM:10000 dataset is highly Skewed, if not addressed the model could get biased towards heavy clases.

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

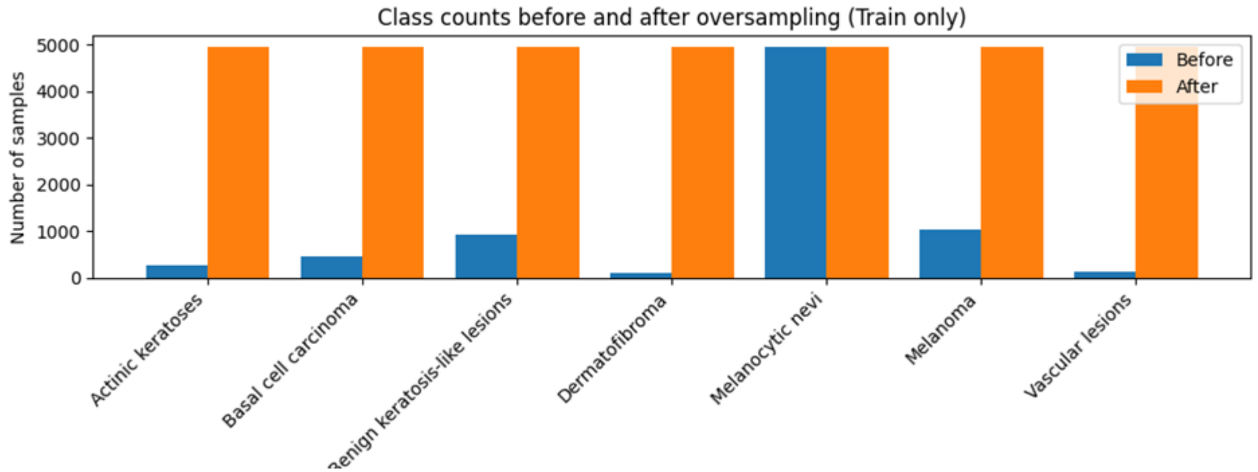


Fig. 4. Training Class Counts Before and After Oversampling

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

Methodology

➢**3 Federated Client:** The Train set was distributed into 3 client.

➢**Developing Proposed Model:** A Hybrid combination of EfficientNetB3 and InceptionV3 frameworks with attention-driven feature fusion

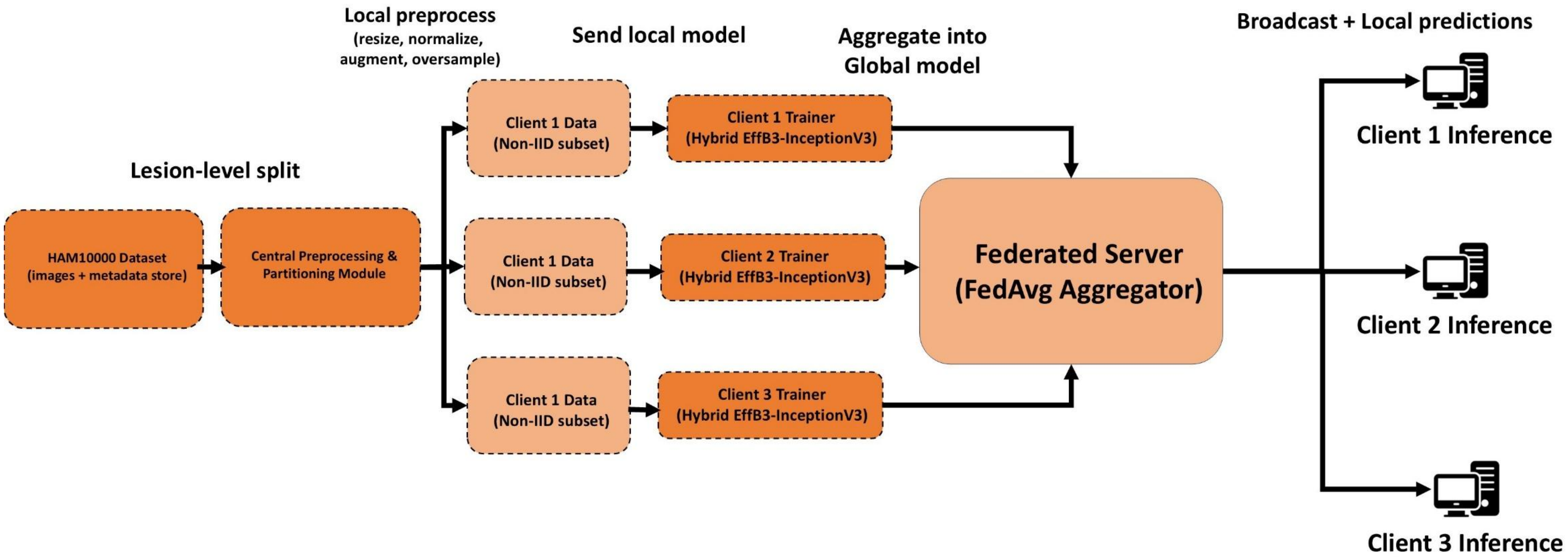


Fig 5: Data Flow of the Federated Clients

Proposed Model Architecture

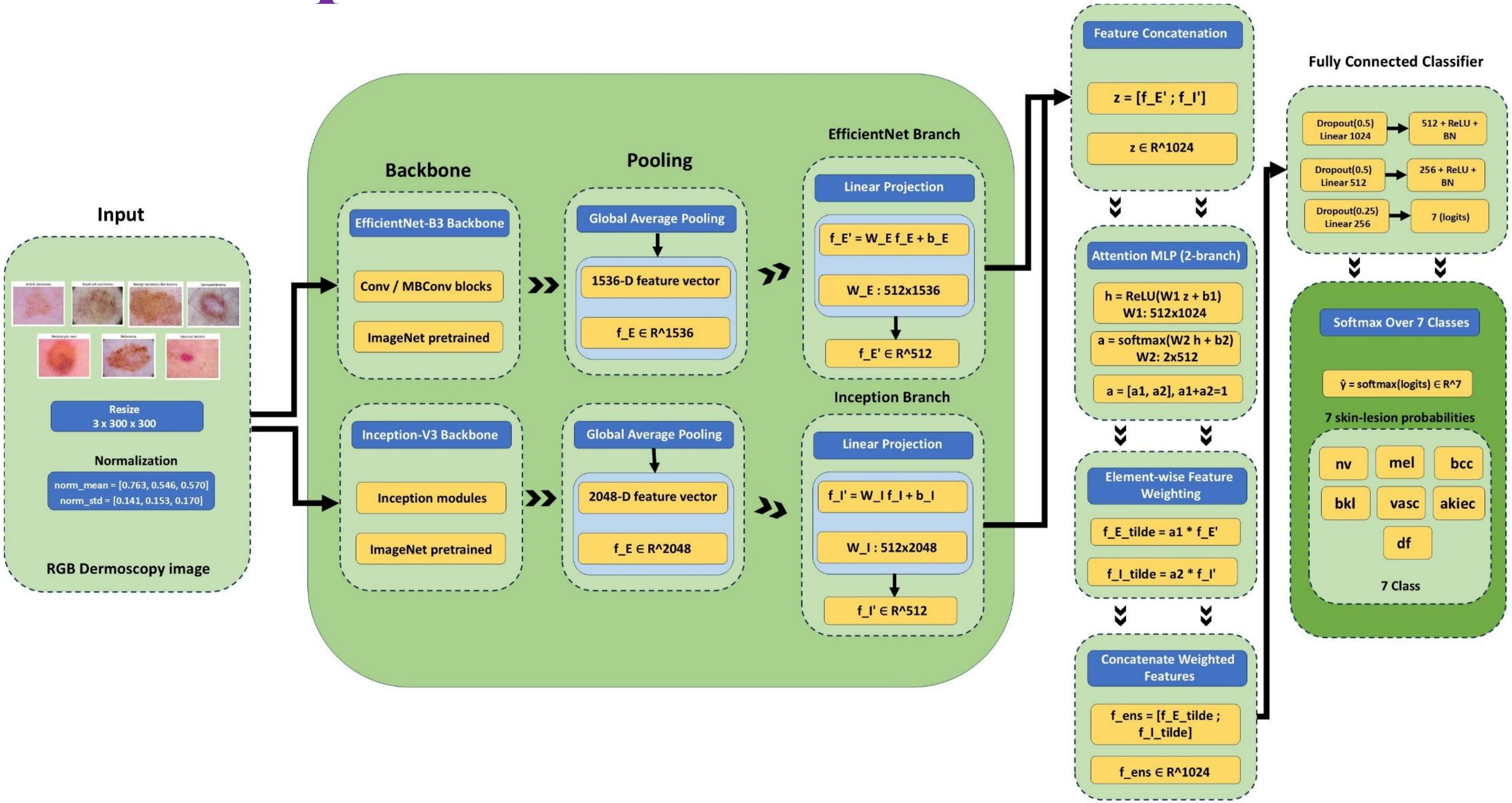


Fig 6: Architecture of Proposed Hybrid EfficientNetB3–InceptionV3 Ensemble

Result and Analysis

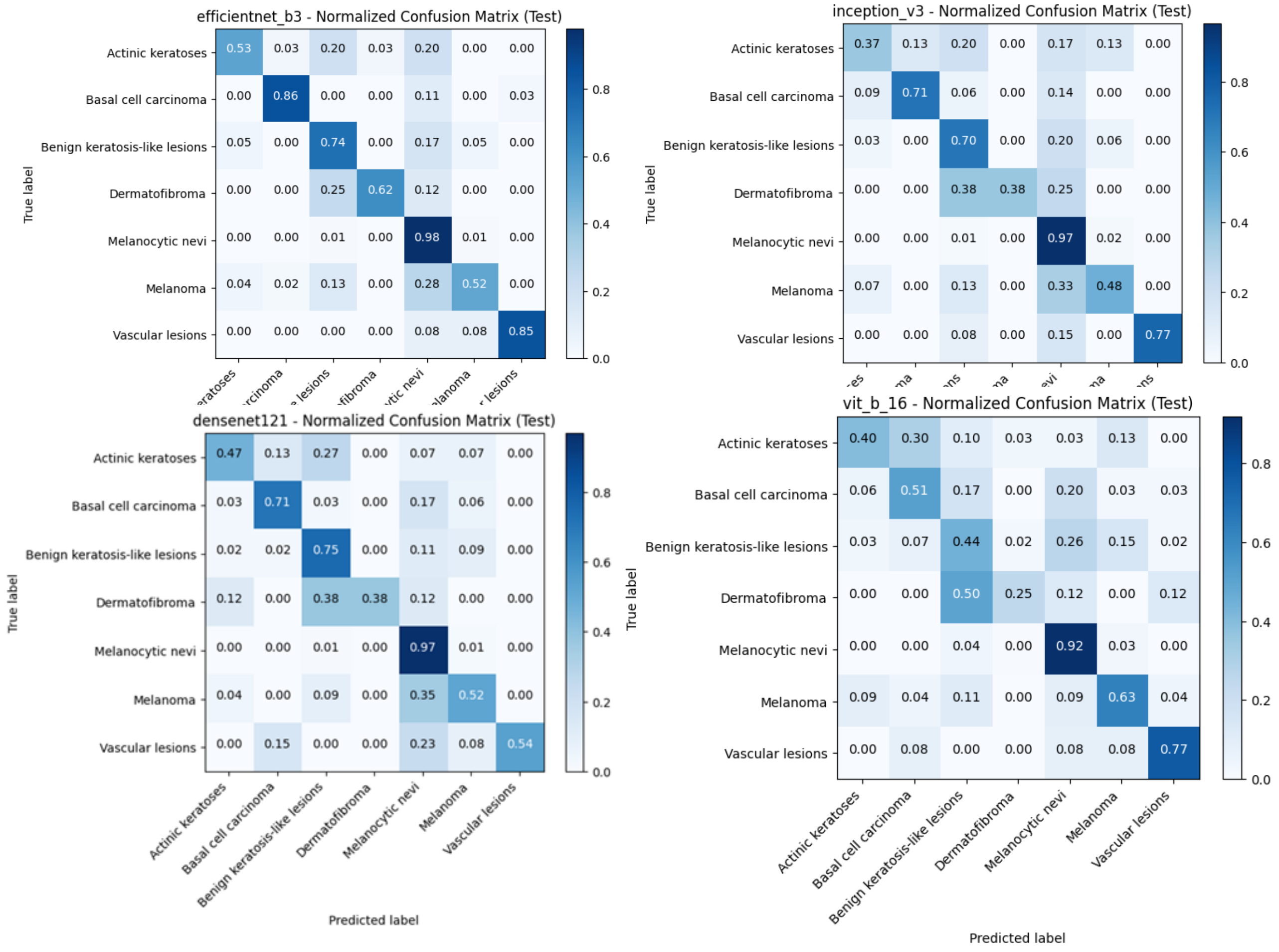


Fig. 7. Confusion Matrix of EfficientNetB3, InceptionV3, Densenet121, Vit_b_16

Result and Analysis

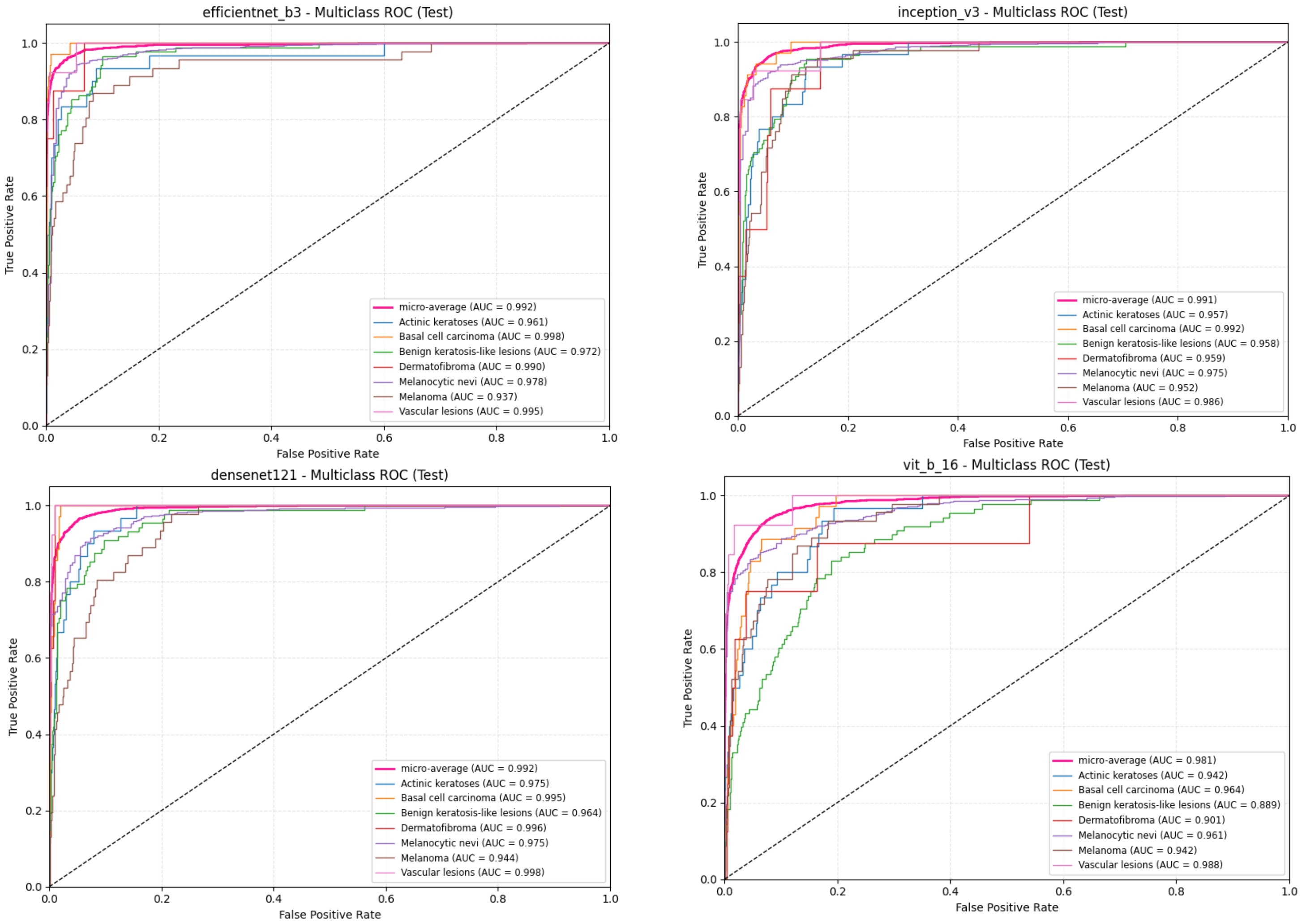


Fig. 8. Multi Class ROC-Curve of EfficientNetB3, InceptionV3, Densenet121, Vit_b_16

Result and Analysis

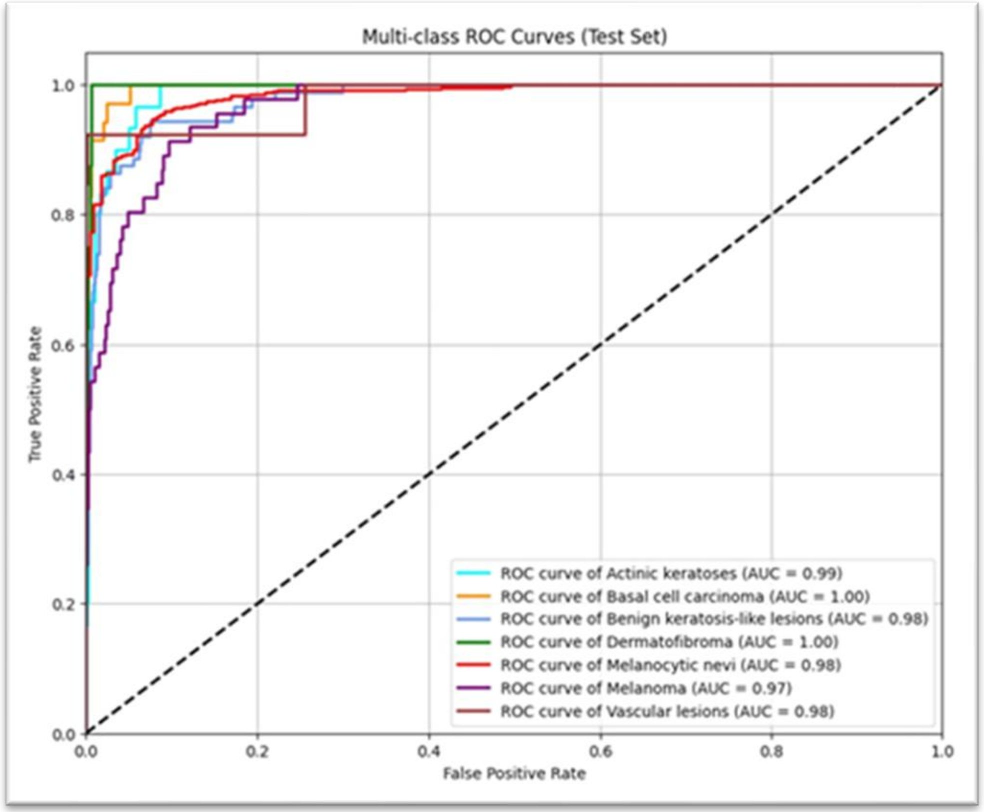


Fig. 9. Multi Class ROC-Curve of Proposed Model

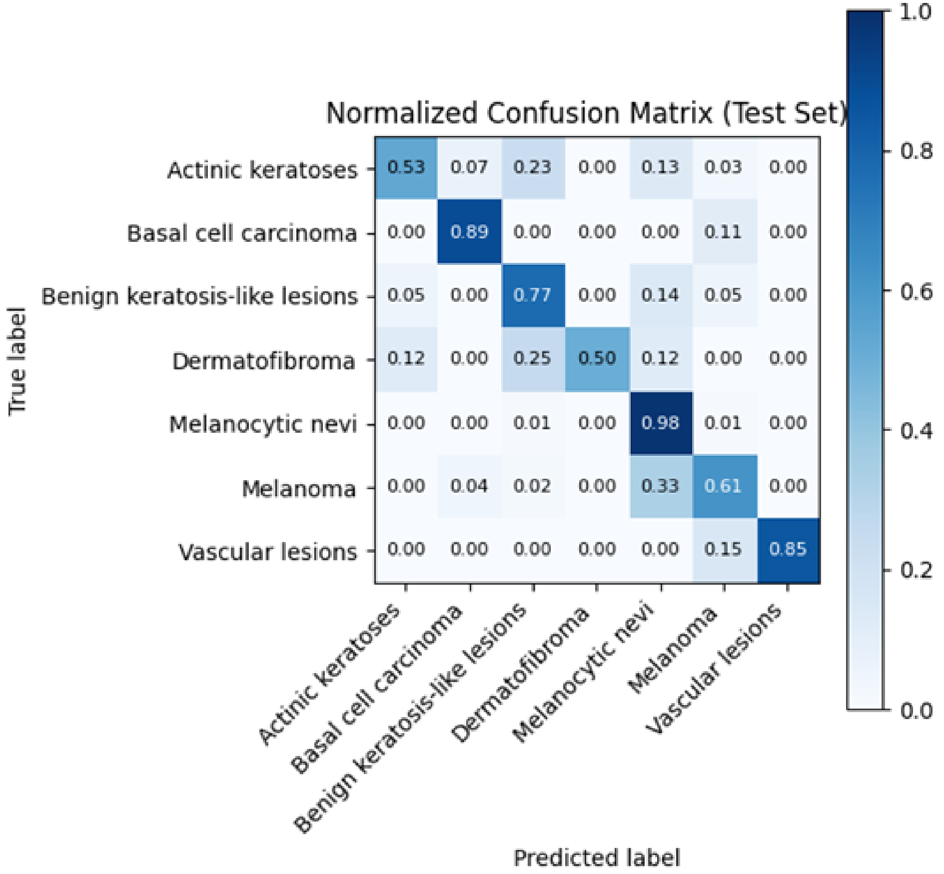


Fig. 10. Confusion Matrix of Proposed Model

Result and Analysis

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: Classification Report of the Proposed Model

Class	Precision	Recall	F1-score	Support
Actinic keratoses	0.72	0.60	0.65	30
Basal cell carcinoma	0.86	0.89	0.87	35
Benign keratosis-like lesions	0.83	0.77	0.80	88
Dermatofibroma	1.00	0.62	0.77	8
Melanocytic nevi	0.96	0.98	0.97	883
Melanoma	0.65	0.61	0.63	46
Vascular lesions	1.00	0.85	0.92	13
Accuracy			**0.9362**	1103
Macro avg	0.86	0.76	0.80	1103
Weighted avg	0.93	0.93	0.93	1103

Result and Analysis

Table IV: Training parameter for the proposed model

Hyperparameter	Value
Input Shape	300 x 300
Batch Size	32 (Effective)
Epochs	10
Optimizer	Adam
Learning Rate	1e-4 (0.0001)
Loss Function	Cross-Entropy

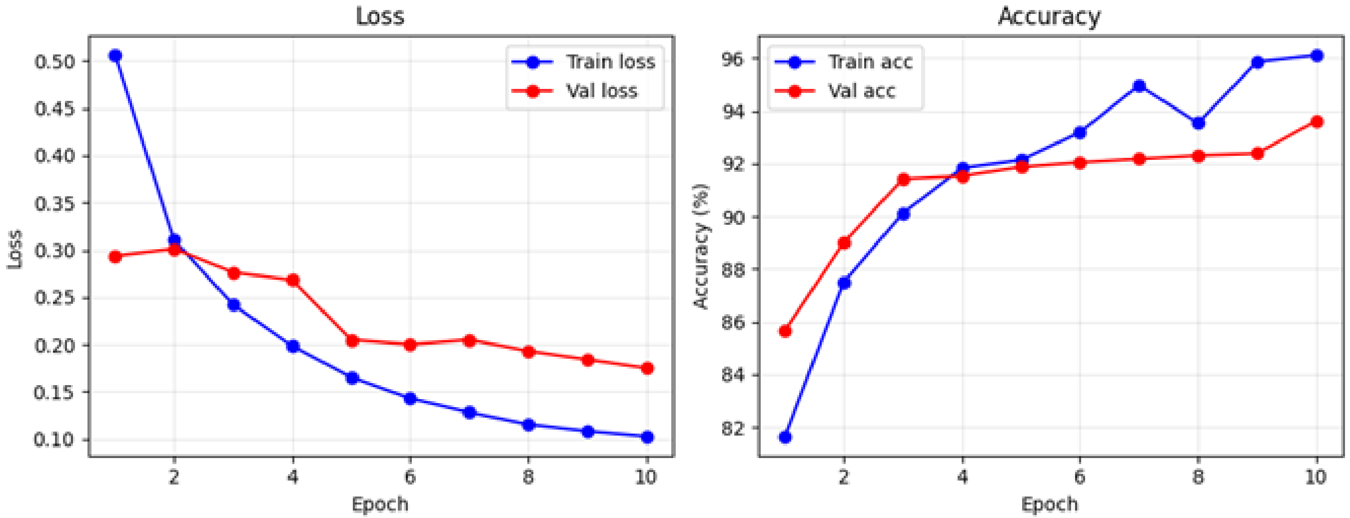


Fig 11. Training and validation loss and accuracy curves

Analysis of Federate Clients

Table V: 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**

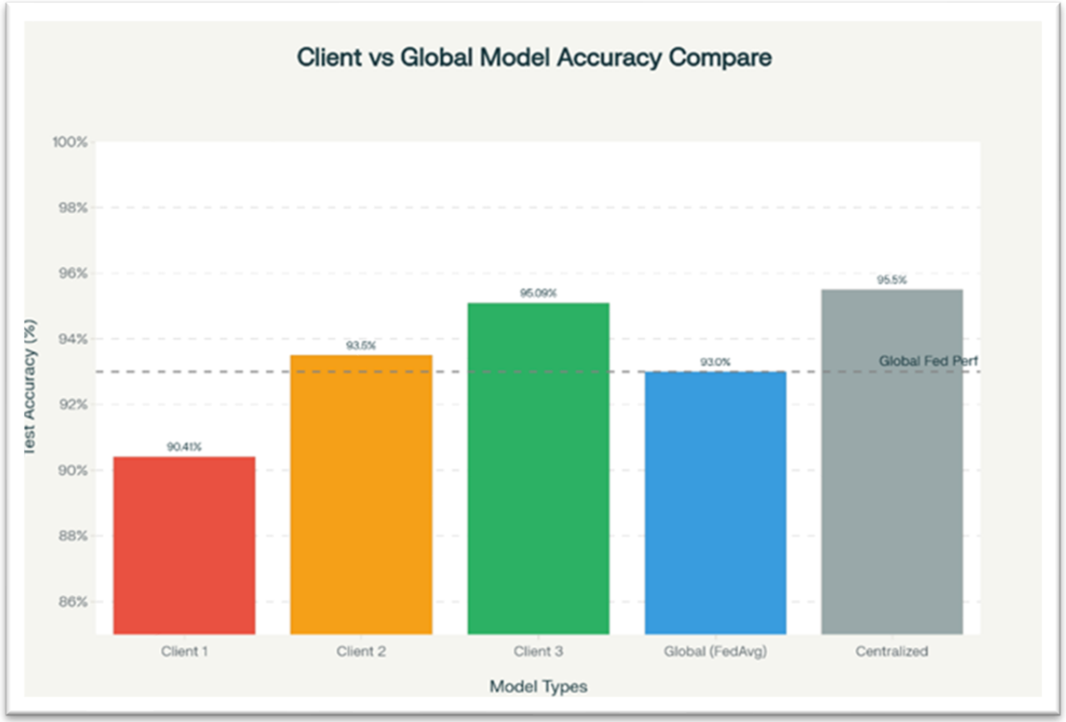


Fig. 12. Accuracy Comparison Bar Chart of Federate clients and Global Model

Analysis of Ablation Study

Table VI: Ablation study Of Proposed and baseline models

- 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.

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.

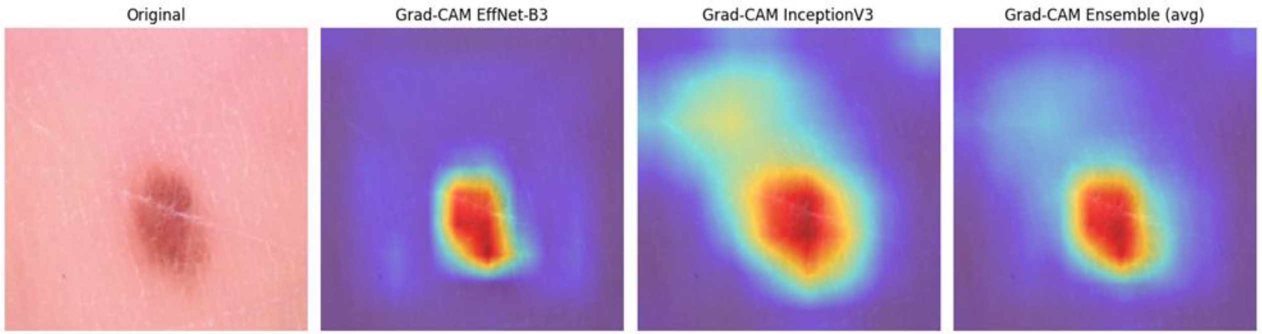


Fig. 13: Grad-CAM skin-lesion visualizations

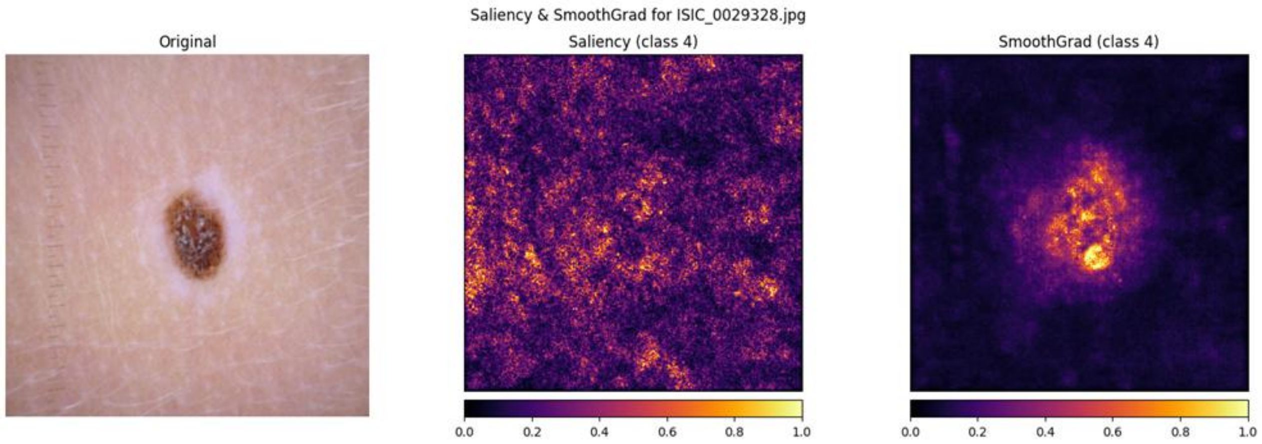
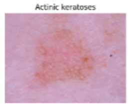
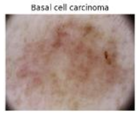
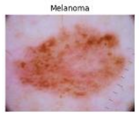


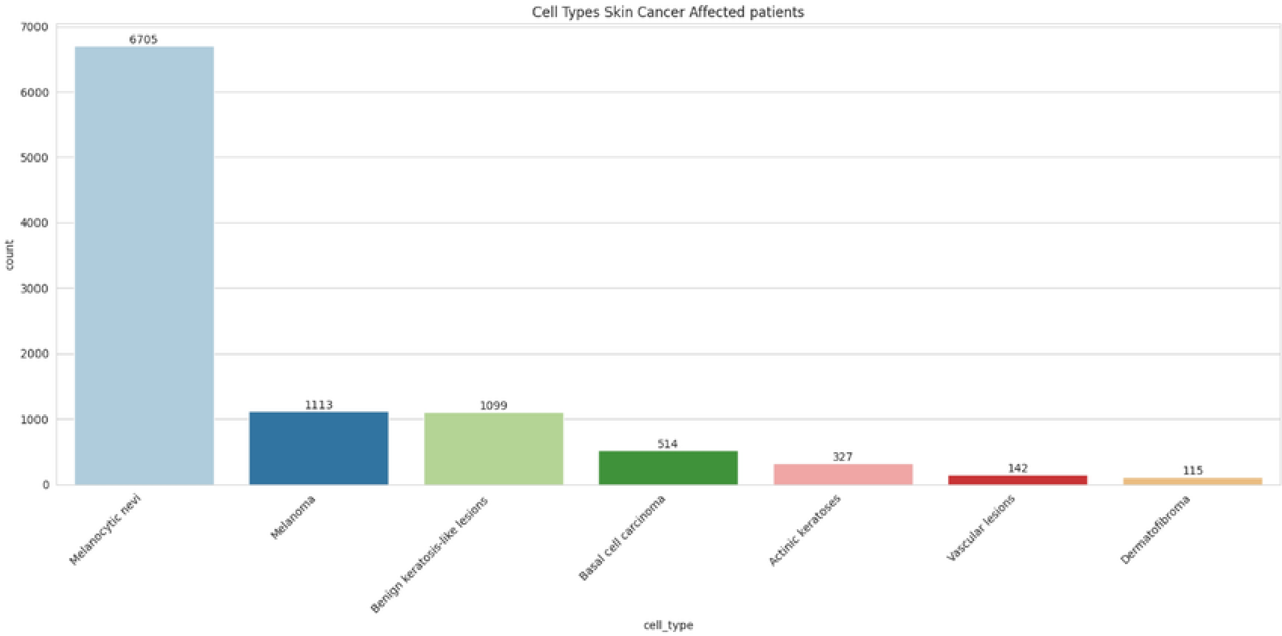
Fig. 14:Saliency and SmoothGrad maps for class-4 skin lesion

Novelty of the Work

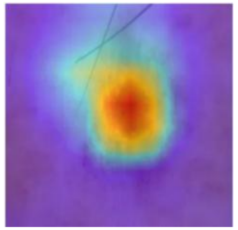
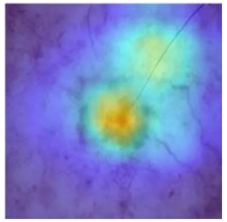
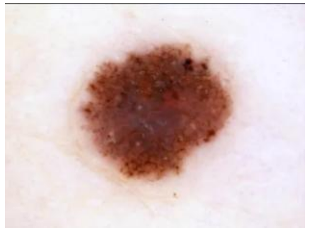
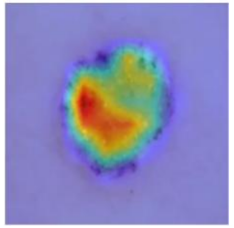
- **Dual-Backbone Ensemble in Federated Setting:** First hybrid EfficientNetB3-InceptionV3 ensemble under federated learning for dermatology
- **Attention-Based Feature Fusion:** Sample-specific weighted fusion improves robustness vs. simple concatenation
- **Non-IID Class Imbalance Handling:** RandomOverSampler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

Sample dataset

Class	Image
Actinic keratoses	
Basal cell carcinoma	
Benign keratosis-like lesions	
Dermatofibroma	
Melanocytic nevi	
Melanoma	
Vascular lesions	



Sample dataset and Expected output

True Class	Ground Truth Image	Model Prediction	Confidence	Grad-CAM
Melanocytic nevi		Melanocytic nevi	100%	
Melanoma		Melanoma (98%) Nevi (2%)	98% confident	
Melanoma		Keratosis *Misclassified*	65%	

Web Interface of Image Enhancement

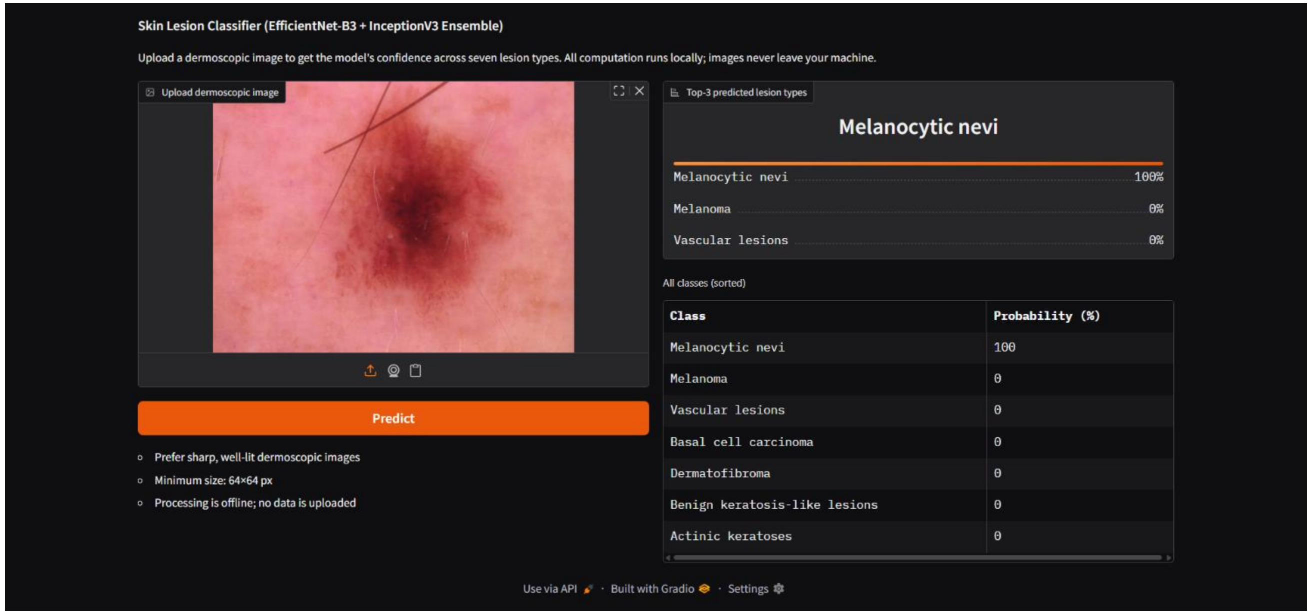


Fig. 14: User Dashboard of Skin Lesion Classifier

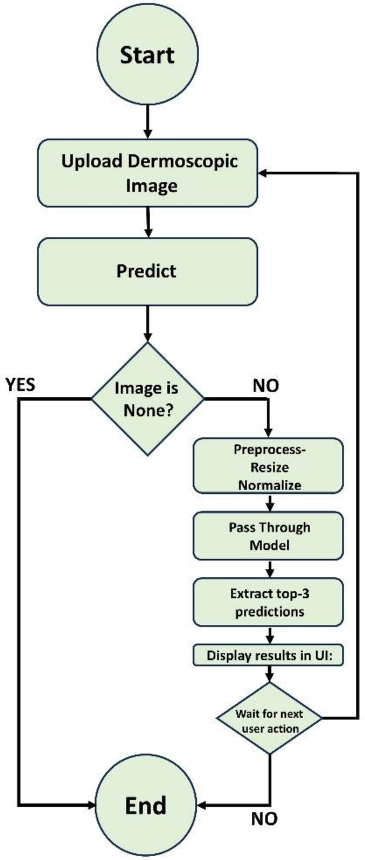


Fig. 16: Activity diagram of the skin-lesion classifier inference workflow

Comparative Analysis

Reference / Author	Methodology	Privacy Preserving (FL)	Multi-Class (7 Classes)	Addressed Class Imbalance?	Explainability (Grad-CAM)	Key Limitation
Yaqoob et al. [1]	CNN with Asynchronous Federated Learning	✅ Yes	❌ No (Binary)	❌ No	❌ No	Limited to only 2 classes (Benign vs. Malignant); no imbalance handling.
Hendrix et al. [2]	Evidential Deep Learning (EDL)	✅ Yes	✅ Yes	❌ No	❌ No	Focuses on uncertainty estimation but ignores the severe class imbalance in HAM10000.
Tian et al. [3]	Dataset Distillation with FL	✅ Yes	✅ Yes	❌ No	❌ No	Computationally expensive; struggles with "Non-IID" data (diverse hospital data).
Shetty et al. [4]	CNN + Transfer Learning	❌ No	✅ Yes	✅ Yes	❌ No	**No Privacy.** Requires centralized data sharing, violating medical regulations (HIPAA).
Proposed Model	**Hybrid Ensemble (EffNetB3 + InceptionV3)**	✅ Yes	✅ Yes	✅ Yes	✅ Yes	**None.** Combines privacy, high accuracy (93.62%), imbalance handling, and doctor-friendly explainability.

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 as a Health Aid tool and can be extended in future work to larger 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.
- [6] "Skin Cancer MNIST: HAM10000," *www.kaggle.com*. https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

THANK YOU