Title:
Federated Skin Cancer Detection Using CNNs and FedOPT Optimization: A Privacy-Preserving Dermatology AI Framework with Client-Level Metric Tracking and 3D Learning Dynamics

---

Executive Summary

This paper presents a federated learning (FL) framework for skin cancer classification using the HAM10000 dermatology dataset. The system employs a lightweight convolutional neural network (CNN), integrates Flower's FedOPT optimization strategy, and incorporates client-level upsampling to mitigate data imbalance. The study also introduces comprehensive metric tracking and 3D visualization techniques to analyze per-client and per-round learning behavior. Results demonstrate stable convergence, improved classification accuracy under non-IID data conditions, and meaningful insights into federated optimization dynamics.

---

1. Introduction

Skin cancer is a major global health concern, and early detection significantly improves patient outcomes. Machine learning systems trained on dermatoscopic images have shown promise in assisting clinical decision-making. However, the sensitive nature of medical data and regulatory constraints make centralized data aggregation impractical.

Federated Learning (FL) offers a decentralized alternative, allowing institutions to collaboratively train models without sharing raw patient data. This work develops an end-to-end FL pipeline for dermatology image classification, combining CNN modeling, adaptive federated optimization via FedOPT, client-level data balancing, and advanced evaluation through ROC curves and 3D surface visualizations.

---

2. Dataset: HAM10000

The HAM10000 dataset contains **10,015 dermatoscopic images** across seven diagnostic categories:

- mel (melanoma)
- nv (melanocytic nevi)
- bkl (benign keratosis)
- df (dermatofibroma)
- akiec (actinic keratoses/Bowen's disease)
- bcc (basal cell carcinoma)
- vasc (vascular lesions)

Challenges

- Severe class imbalance
- High intra-class variation
- Variability in image resolution and illumination

Preprocessing

- Images resized to 128×128
- RGB normalization
- Label encoding
- Division into training and test sets
- Subsequent partitioning into five federated clients

Representative sample images illustrate the diverse visual characteristics across lesion categories.

3. Federated System Architecture

The federated learning system consists of five simulated clients and a central server orchestrated using the Flower framework.

Key Components

1. **Client Nodes**
   - Receive unique, non-IID partitions of the training dataset
   - Maintain separate training and validation splits
   - Apply oversampling to balance class distribution
2. **Federated Server**
   - Performs FedOPT aggregation
   - Tracks per-round metrics
   - Updates and broadcasts global model parameters
3. **Training Protocol**
   - Five communication rounds
   - Full participation from all clients each round
   - Three local epochs per client

This architecture reflects real-world FL settings involving multiple healthcare institutions.

4. Model Architecture

A lightweight convolutional neural network (CNN) is employed to ensure computational efficiency in federated environments.

Conv2d(3, 32) → ReLU → MaxPool
Conv2d(32, 64) → ReLU → MaxPool
Conv2d(64, 128) → ReLU → MaxPool
Flatten → Linear(32768, 128) → ReLU
Linear(128, 7)

CrossEntropyLoss is used for multi-class classification, and the Adam optimizer is applied during local training.

---

5. Federated Optimization: FedOPT

FedOPT generalizes adaptive optimization techniques such as Adam to the federated setting.

Parameters Used

- Server learning rate $\eta = 0.01$
- Local learning rate $\eta\_l = 0.001$
- $\beta_1 = 0.90$
- $\beta_2 = 0.999$
- $\tau = 1e{-9}$

A customized FedOPT implementation is used to collect:

- Per-round training loss (per client)
- Per-round validation loss (per client)
- Per-round validation accuracy (per client)

These metrics enable fine-grained analysis of convergence behavior across heterogeneous clients.

---

6. Handling Non-IID Data Through Upsampling

The HAM10000 dataset exhibits substantial class imbalance. To reduce client drift and stabilize convergence, each client applies an oversampling procedure that:

- Identifies minority class samples
- Replicates them until all classes reach the majority-class frequency
- Randomizes dataset order after resampling

This strategy ensures more uniform class distribution across clients, improving both learning stability and evaluation consistency.

---

7. Training and Evaluation Methodology

Local Training

Each client conducts:

- 3 training epochs
- Batch size of 32
- CNN forward and backward propagation
- Loss computation and optimization

Local Validation

Each client computes:

- Validation accuracy
- Validation loss

Global Evaluation

After federated training concludes, the final global model is evaluated on a held-out test set using:

- Confusion matrix
- Class-wise ROC curves
- AUC scores

---

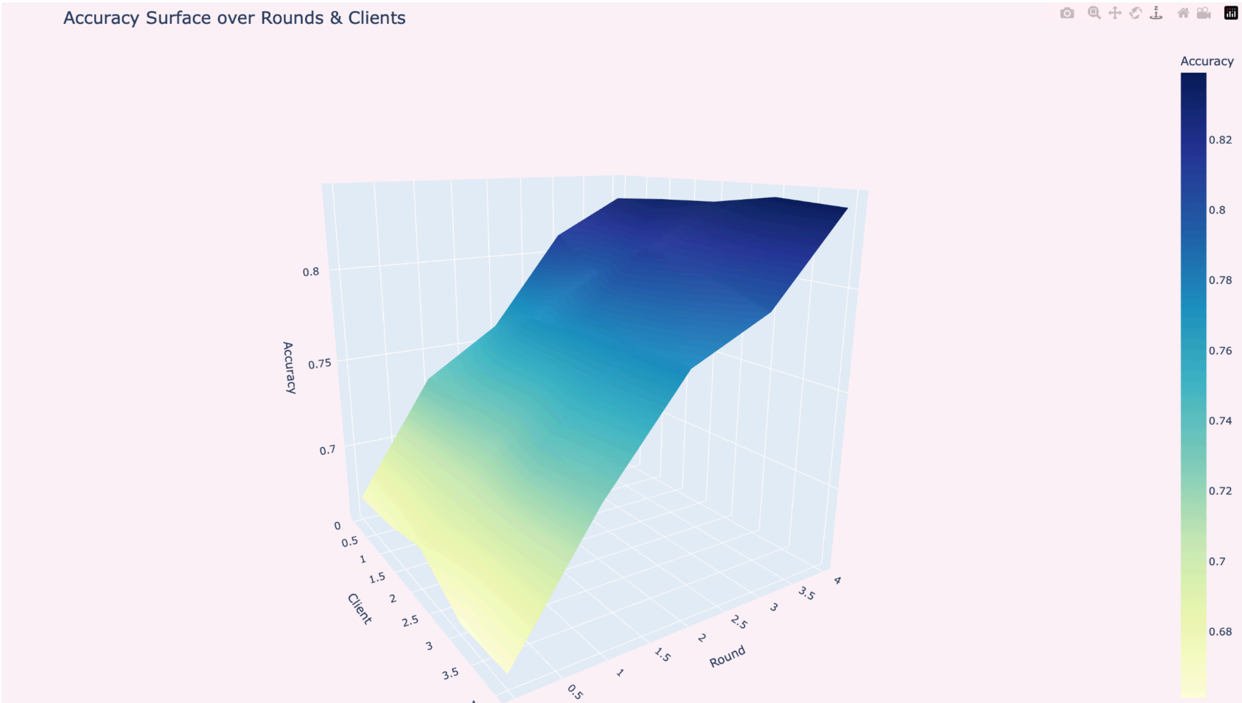
8. Results

8.1 Accuracy Surface

A 3D surface visualization of validation accuracy across rounds and clients shows:

- Consistent improvement over all five rounds
- Final-round accuracies ranging between ~0.80 and ~0.84
- Reduced variability among clients in later rounds

Interpretation:
FedOPT demonstrates stable convergence even with heterogeneous client datasets.

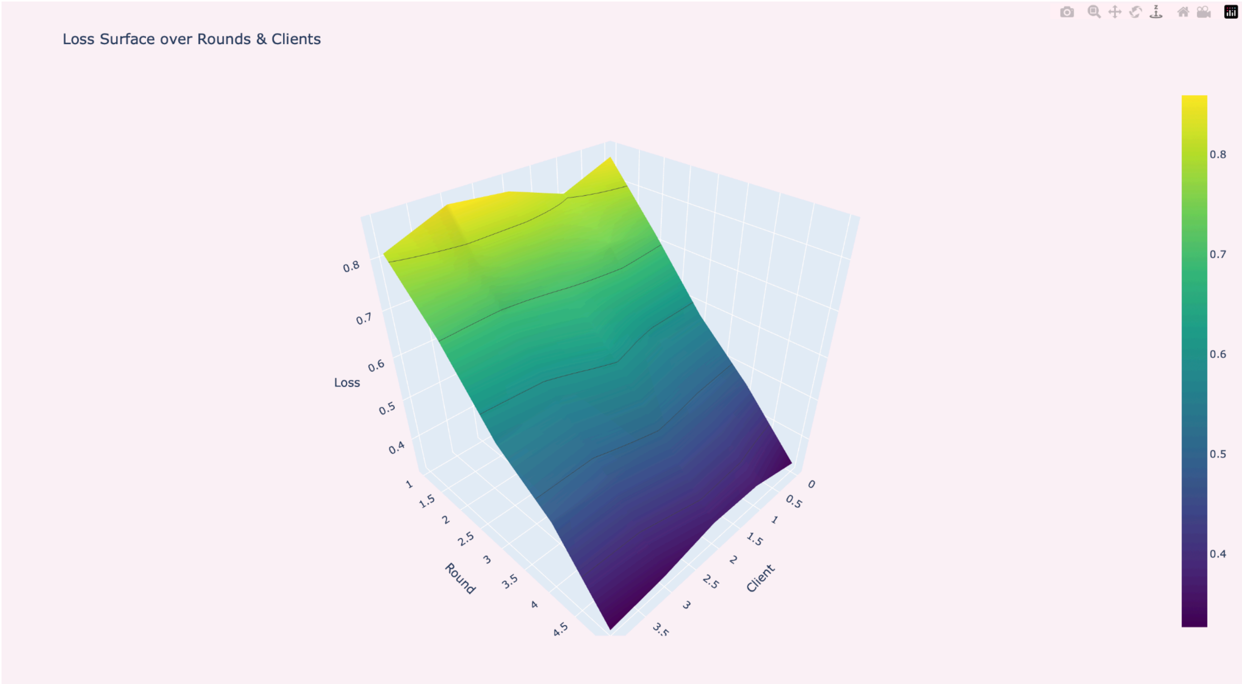


(Figure: Accuracy Surface over Rounds & Clients)

8.2 Loss Surface

Training and validation losses decrease steadily across rounds, with final validation losses in the 0.35–0.45 range.

Interpretation:
The downward trend indicates effective learning dynamics with no signs of divergence or overfitting.



(Figure: Loss Surface over Rounds & Clients)

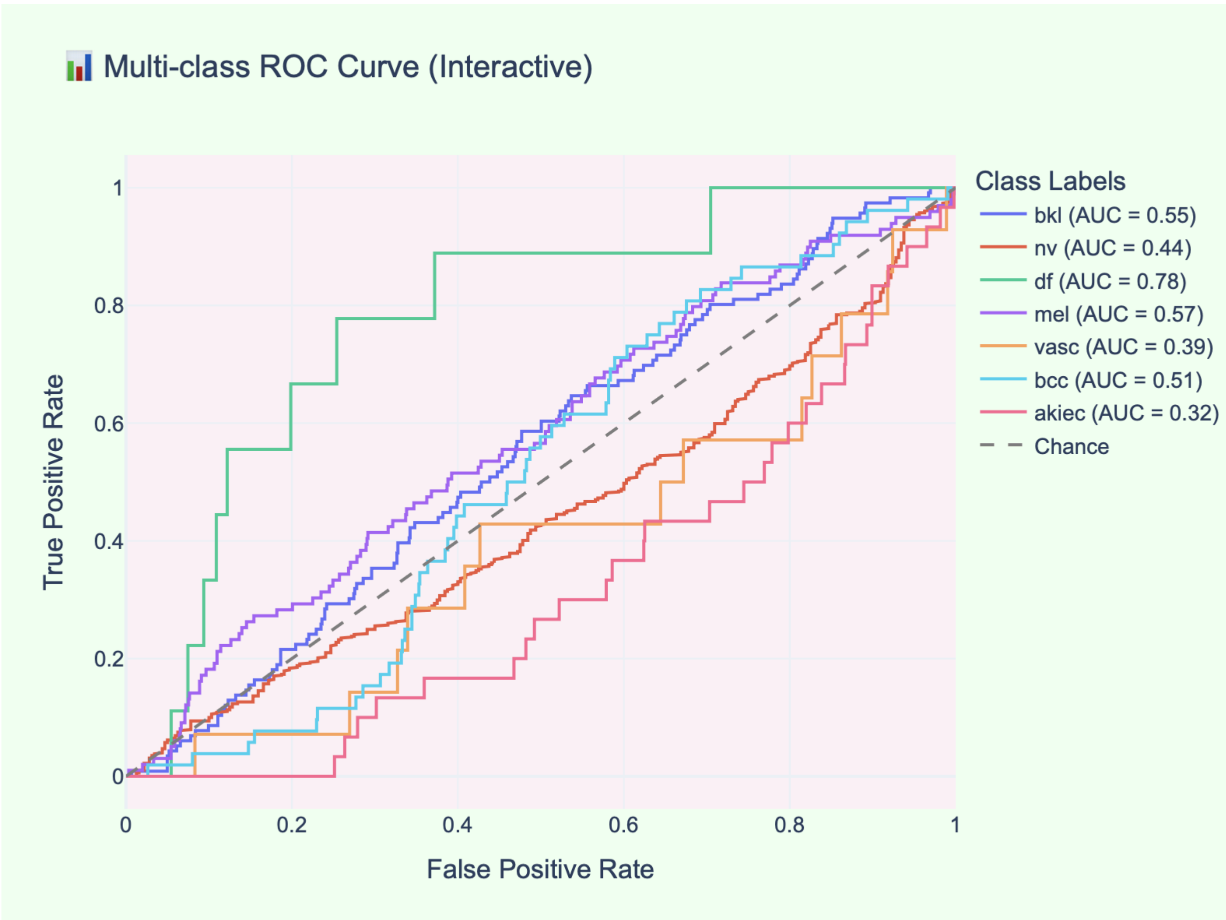
8.3 Multi-Class ROC and AUC

One-vs-rest ROC analysis yields the following AUC values:

Class	AUC
df	**0.78**
mel	0.57
bkl	0.55
bcc	0.51
nv	0.44
vasc	0.39
akiec	0.32

Interpretation:

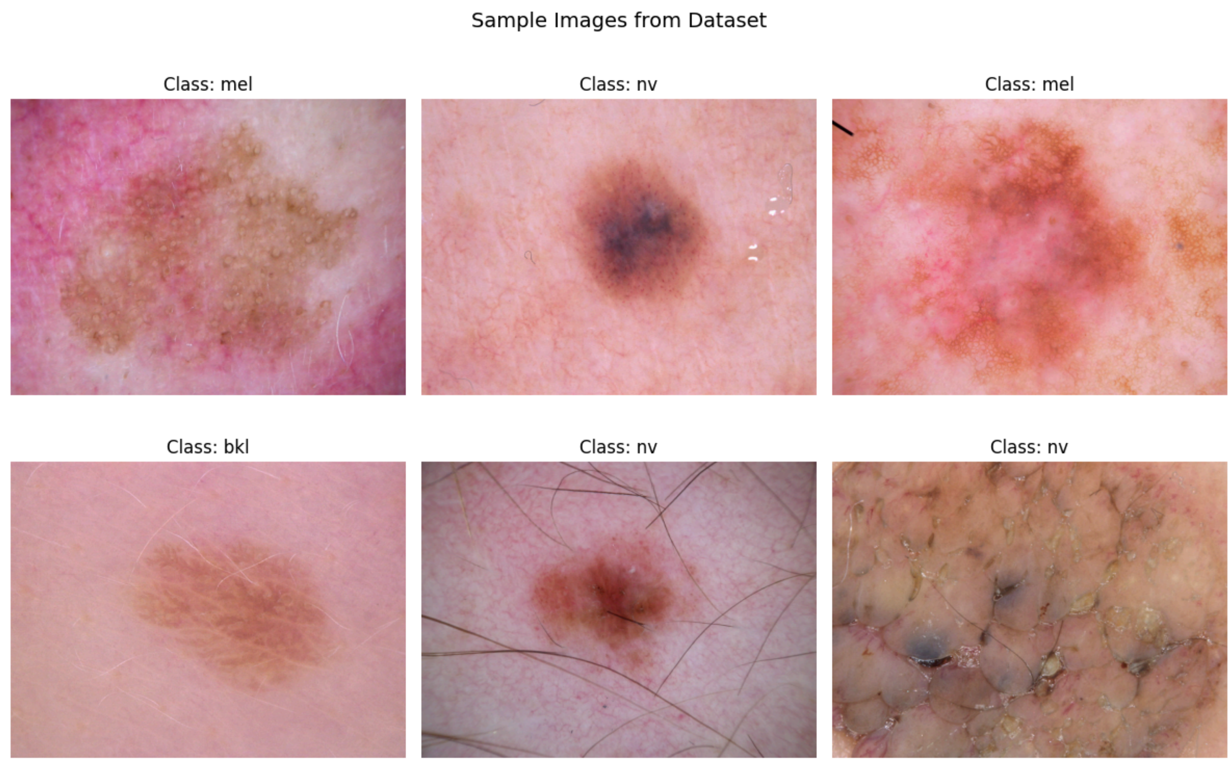
- Strong separability for *df*
- Moderate performance for *mel* and *bkl*
- Lower performance for *akiec* and *nv* due to high intra-class variability and limited samples



(Figure: Multi-class ROC Curve)

8.4 Sample Images

A grid of representative images demonstrates the inherent complexity of the classification task across lesion types.



(Figure: Sample Images from Dataset)

---

9. Discussion

The results demonstrate that:

- The FedOPT-based FL system converges reliably on non-IID medical image data
- Upsampling is effective in reducing performance disparities among clients
- 3D visualization is highly useful for diagnosing cross-client learning dynamics
- Despite architectural simplicity, the CNN achieves competitive performance

Limitations

- Minority classes remain challenging due to limited sample sizes
- A simple CNN architecture constrains maximum achievable accuracy
- Further improvements require more sophisticated model architectures or augmentation techniques

---

10. Future Work

Potential extensions include:

- Differential privacy integration for regulatory compliance
- Secure aggregation for enhanced confidentiality
- Personalized federated methods (e.g., FedPer, FedRep)
- Vision Transformers adapted for federated settings
- Automated hyperparameter tuning
- Deployment across real multi-institution networks

---

11. Conclusion

This study presents a practical and interpretable federated learning pipeline for skin cancer detection using non-IID dermatology image data.
The combination of:

- Lightweight CNN modeling
- FedOPT optimization

- Client-level upsampling
- Per-client metric tracking
- 3D convergence visualizations

collectively demonstrates the feasibility and value of applying FL to medical imaging tasks.

The framework serves as a foundation for further research and development in privacy-preserving clinical AI systems.

---

12. References

- Flower: A Friendly Federated Learning Framework
- Tschandl et al. HAM10000 Dataset
- Reddi et al., Adaptive Federated Optimization Algorithms (FedOPT)
- PyTorch Documentation
- scikit-learn Metrics Documentation
