Blockchain-Assisted Federated Hybrid Deep Learning for Secure and Privacy-Preserving Skin Cancer Diagnosis

Naresh Alapati[1,2], BVVS Prasad[3]

[1]Lincoln University College, Petaling Jaya, Selangor, Malaysia.

[2]Department of Computer Science and Engineering, Vignan's Nirula Institute of Technology and Science for Women, Peda Palakaluru, Guntur, India.

[3]School of Engineering (CSE), Anurag University, Venaktapur, Ghatkesar, Hyderabad, Telangana, India

E-mail:[1]pdf.Naresh@lincoln.edu.my

Abstract

Skin cancer is one of the most prevalent and fastest-growing forms of cancer globally, and late diagnosis is still a significant issue. Despite the success of deep learning for dermoscopic image analysis, centralized training is plagued by the limitations of data-privacy regulations (HIPAA, GDPR), non-uniform data distributions between rare lesion types, and problems of adversarial and poisoning attacks in common computing environments. The complementary solutions to these challenges are federated learning (FL), hybrid deep learning (HDL) and blockchain, respectively, as they preserve privacy, enhance the robustness of features by architectural fusion and offer tamper-resistant, audit-able aggregation. This paper summarizes the current status of the development of each paradigm, shows that, on their own, there are still significant knowledge gaps, and outlines an integrated approach and a pre-registered experimental protocol for a secure, scalable, multi-institutional skin cancer diagnosis system that brings all three paradigms together.

Keywords—Skin Cancer, Federated Learning, Hybrid Deep Learning, Blockchain, Medical Imaging, Privacy Preservation, Transfer Learning.

I. Introduction

Early detection increases the 5-year chances of survival for skin cancer by more than 50 percent. expertise-dependent, and expensive.However, while the survival rates range from approximately 20% to more than 95%, traditional dermoscopy and biopsy are expertise-dependent, slow, and costly. And highly variable between observers. CNN architectures like EfficientNet, DenseNet and ResNet In controlled environments, they can now match the accuracy of dermatologists, however, there are four challenges to real world use. Centrilized training data is constrained by data-privacy regulation and class imbalance degrades the quality. Difficulty with rare clinically important lesions; difficulty with domain shift between imaging devices and imaging protocols; hurts Centralized repositories can suffer from data leakage and attack; and generalization. Federated None of these barriers are solved by all three approaches, that is, learning, hybrid deep learning or blockchain. The three were combined into one pipeline. This paper provides a review of the advances made on each of these areas, (1) compares the existing work on these areas, and (2) suggests future directions. It identifies key evaluation dimensions, (3) establishes formalized knowledge gaps and (4) outlines an integrated approach. A closing, and evaluation protocol that closes them.

II. Literature Survey

Traditional ML methods (ANN, SVM, KNN) [1] are noise-sensitive and generalize poorly under class imbalance. Single- and multi-architecture CNN ensembles [2]-[4] reach state-of-the-art accuracy on ISIC benchmarks but rely on centralized training, incur high computational cost, or remain data-and-compute intensive, limiting deployment in resource-constrained clinics. Federated approaches [5],[9],[10] demonstrate privacy-preserving collaborative training, and blockchain-enabled FL [10] adds auditability, but at the cost of communication overhead and limited transaction throughput. Metadata-fusion [6], hierarchical taxonomy [7], and early CNN-vs-dermatologist studies [8] improve interpretability or establish accuracy benchmarks, but none combine privacy-preserving federation, hybrid feature fusion, and blockchain-secured aggregation in

one pipeline while also handling rare-class imbalance. Table 1 compares these works across the four evaluation dimensions used throughout this paper.

Table 1. Comparison of related works and this work across key parameters

Reference	Privacy-Preserving FL	Hybrid DL Architecture	Blockchain Integration	Rare Class Handling
[1] Yang et al., 2025	No	No	No	No
[2] Aboulmira et al., 2024	No	Yes	No	No
[3] Farea et al., 2024	No	Yes	No	Partial
[4] Balaha et al., 2023	No	Yes	No	No
[5] Chen et al., 2022	Yes	No	No	No
[6] Pacheco et al., 2020	No	Yes	No	No
[7] Barata et al., 2019	No	No	No	No
[8] Haenssle et al., 2018	No	No	No	No
[9] Nguyen et al., 2022	Yes	No	No	No
[10] Kumar et al., 2024	Yes	No	Yes	No
This Work	**Yes**	**Yes**	**Yes**	**Yes**

III. Knowledge Gaps and Motivation

Six knowledge gaps (G1-G6) motivate the method and experiments that follow:

- G1 – No integration: FL, hybrid DL and blockchain are used and studied individually, with none bringing all three together to one skin-cancer pipeline.
- G2 - No formal privacy guarantee: Federated approaches often do not have a formal differential-privacy bound, but they are competitive in most of the accuracy measures.
- G3: Poor rare-class handling: Current architectures and training strategies are not sufficient for class imbalance in the datasets of rare classes (lesions).
- G4 - High computational/communication cost: hybrid models and FL aggregation are often impractical for bandwidth- and GPU-limited clinics.
- G5 - Weak poisoning/Byzantine defenses: blockchain-based aggregation is a principled defense but is rarely used systematically for skin-cancer FL.
- G6 - Limited real-world validation: most systems are validated only under idealized, single-source conditions.

Table 2 maps each gap to the method component (Section IV) that closes it and the metric (Section V) used to verify closure.

Gap	Method component (Sec. IV)	Verification metric (Sec. V)
G1	Joint HDLM + FL + blockchain (IV.B-E)	Ablation study (V.C)

G2	DP-calibrated gradient sharing (IV.D)	Empirical privacy audit (V.D)
G3	Focal loss + CGAN oversampling (IV.F)	Per-class recall/F1 (V.D)
G4	Client-side training/aggregation design (IV.D)	Communication cost, convergence (V.D)
G5	Smart-contract validation (IV.E)	Poisoning-detection precision/recall (V.D)
G6	Multi-dataset, non-IID protocol (V.A-B)	Cross-institution validation (V.E)

IV. Definition of Methods

A. Problem Formulation

Let D = {D1,...,DN} be the private local datasets of N institutions, Dk = {(xi_k, yi_k)}, with yi_k one of C = 9 ISIC lesion categories. The goal is a global model f(theta) minimizing the sample-weighted aggregate risk across clients, such that (i) no raw image/label leaves its institution, (ii) every transmitted update satisfies (epsilon, delta)-differential privacy, and (iii) every update and aggregation step is logged as an auditable ledger transaction.

B. Hybrid Deep Learning Module (HDLM)

Two parallel encoders, EfficientNet-B4 and DenseNet-121, produce feature maps FE and FD from an input dermoscopic image, each projected to a common channel dimension via 1x1 convolution. EfficientNet-B4 captures fine-grained spatial detail through compound scaling; DenseNet-121 promotes feature reuse and gradient flow through dense connectivity.

C. Multi-Scale Attention Fusion

A channel branch computes ac = sigmoid(W2*ReLU(W1*GAP(F))) to emphasize informative feature channels; a spatial branch computes as = sigmoid(Conv7x7([AvgPool(F);MaxPool(F)])) to emphasize diagnostically salient regions (vascular patterns, asymmetric pigmentation, irregular borders). The fused representation F_fused = Concat(FE,FD) x ac x as passes through batch-normalized then dropout-regularized (p=0.4) fully connected layers to yield class probabilities.

D. Federated Learning Protocol with Differential Privacy

At round t, client k runs T local SGD epochs to produce update gk, clipped to L2 bound C and perturbed as gk_tilde = clip(gk,C) + N(0, sigma^2*C^2), with sigma set via the moments accountant for a target (epsilon, delta) budget over all rounds. The server aggregates via sample-weighted FedAvg, theta = sum_k (nk/n) * theta_k, so institutional size determines contribution weight without requiring data centralization.

E. Blockchain-Assisted Secure Aggregation

Each signed, DP-noised update is submitted as a transaction to a permissioned Hyperledger Fabric network. A smart contract (1) verifies signatures, (2) flags statistical outliers consistent with poisoning, (3) enforces a minimum-quorum rule before aggregating, and (4) commits the aggregated model hash and contributing-client record as an immutable, auditable block.

F. Class-Imbalance Handling

Local training minimizes a focal loss, L = -sum_c alpha_c*(1-p_c)^gamma*log(p_c), with alpha_c inversely proportional to class frequency, and is preceded by per-class CGAN oversampling of rare lesion categories, with the synthetic-to-real ratio capped per client and synthetic samples excluded from validation, testing, and aggregation-quality checks.

V. Experiments: Design and Evaluation Protocol

This section specifies the pre-registered protocol for evaluating the method in Section IV; as this is currently a conceptual contribution, no results are reported, and every design choice is traceable to the gap it targets (Table 2).

A. Datasets and Partitioning

ISIC 2019/2020, HAM10000 [11], and PAD-UFES-20 [6] are pooled and harmonized to nine ISIC categories, then split across N simulated clients via non-IID Dirichlet allocation over labels to emulate realistic inter-hospital skew. A stratified global test set is held out centrally and excluded from all training/aggregation decisions.

B. Preprocessing and Federated Simulation Setup

Images undergo color-constancy normalization, artifact removal, and standard resizing/normalization; CGAN oversampling is applied only to local training folds. The FL protocol is simulated (e.g., Flower) with fixed local epochs, communication rounds, and client participation fraction, sweeping epsilon to trace the privacy-utility trade-off; the blockchain layer is simulated as a configurable Hyperledger Fabric network to measure consensus latency and throughput.

C. Baselines and Ablation Study

Variant	**Isolates**
Centralized HDLM (pooled data)	Accuracy upper bound, no privacy/federation cost
FedAvg, single-backbone (EffNet-B4 or DenseNet-121 only)	Contribution of hybrid fusion
FedAvg + HDLM, no differential privacy	Accuracy cost of DP
FedAvg + HDLM, trusted-server aggregation	Overhead/benefit of blockchain layer
FedAvg + HDLM, cross-entropy + no CGAN	Contribution of rare-class handling
Reproduced prior methods [2]-[5],[10]	External comparison

D. Evaluation Metrics

Category	**Metrics (gap addressed)**
Overall	Accuracy, macro-F1, macro-AUROC
Rare-class	Per-class recall/F1, balanced accuracy (G3)
Privacy	Empirical epsilon, membership-inference success rate (G2)
Security	Poisoning-detection precision/recall (G5)
System cost	Communication volume, convergence rounds, consensus latency/throughput (G4)

E. Statistical Validation and Reproducibility

Institution-level k-fold cross-validation and paired significance testing (Wilcoxon signed-rank) against each baseline/ablation, reported as confidence intervals rather than point estimates. Seeds, hyperparameters, and the partitioning script are to be fixed and released for independent reproducibility.

VI. Conclusion and Future Work

Federated learning, hybrid deep learning, and blockchain each address a distinct barrier to real-world skin-cancer diagnosis, but no prior work unifies all three. This paper defines an integrated method, HDLM with attention fusion, DP-calibrated federated training, blockchain-secured aggregation, and focal-loss/CGAN rare-class handling, together with a pre-registered evaluation protocol that traces every metric back to a specific knowledge gap. Future work includes running this protocol on multi-institutional dermoscopic data, optimizing communication efficiency via gradient compression and asynchronous aggregation, exploring

lighter-weight consensus mechanisms, extending to multimodal clinical inputs, and developing post-hoc explainability for clinically interpretable diagnosis rationales.

References

[1] G. Yang, S. Li, Q. Zhao, and W. Zhang, "A comprehensive review of ML and DL techniques for skin disease classification," IEEE Access, vol. 13, pp. 11234–11258, 2025.

[2] A. Aboulmira, R. Hachimi, and H. Lakhbab, "SkinHealthMate: Web and mobile AI for skin disease using EfficientNet ensemble," Expert Syst. Appl., vol. 241, p. 122683, 2024.

[3] E. Farea, A. Al-Qahtani, and S. Hameed, "Hybrid DL model with ABC optimization for skin disease classification," Biomed. Signal Process. Control, vol. 88, p. 105634, 2024.

[4] H. M. Balaha and A. E. Hassan, "Skin disease classification using transfer learning and spatial self-attention mechanisms," IEEE Access, 2023.

[5] Z. Chen, W. Liang, and K. Li, "Personalized FL with progressive Fourier aggregation and conjoint prototype alignment for skin lesion classification," IEEE Trans. Med. Imag., vol. 41, no. 12, pp. 3597–3610, 2022.

[6] A. G. C. Pacheco et al., "PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images from smartphones," Data Brief, vol. 32, p. 106221, 2020.

[7] C. Barata, M. E. Celebi, and J. S. Marques, "Improving dermoscopy image classification using color constancy," IEEE J. Biomed. Health Inform., vol. 19, no. 3, pp. 1146–1152, 2015.

[8] H. A. Haenssle et al., "Man against machine: Deep learning CNN for dermoscopic melanoma recognition vs. 58 dermatologists," Ann. Oncol., vol. 29, no. 8, pp. 1836–1842, 2018.

[9] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for smart healthcare: A survey," ACM Comput. Surv., vol. 55, no. 3, pp. 1–37, 2022.

[10] S. Kumar, R. Patel, M. Singh, and A. Sharma, "Blockchain-enabled FL framework for secure healthcare applications," IEEE Access, vol. 12, pp. 45000–45015, 2024.

[11] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset of multi-source dermatoscopic images," Sci. Data, vol. 5, p. 180161, 2018.

[12] N. Codella et al., "Skin lesion analysis toward melanoma detection: ISIC 2018 challenge," arXiv:1902.03368, 2019.

[13] T. J. Brinker et al., "Skin cancer classification using CNNs: Systematic review," J. Med. Internet Res., vol. 20, no. 10, p. e11936, 2018.

[14] A. Mahbod et al., "Fusing fine-tuned deep features for skin lesion classification," Comput. Med. Imaging Graph., vol. 71, pp. 19–29, 2019.

[15] R. Sharma, S. Gupta, and A. Verma, "Hybrid DL with multi-scale feature extraction for skin disease," Expert Syst. Appl., vol. 215, p. 119433, 2023.