Indian Engineering Journal
ISSN: 2984-9047
Vol 4 Issue 2 (2026)

Mitigating Statistical Non-Independence in Distributed Facial Recognition Systems through Personalized Federated Learning and Feature Alignment Techniques

Ms Thenmozhi R[1], Dr M. Santhalakshmi[2], Dr M. Shanthakumar[3]

[1]Research Scholar, Department of Computer Science, Kamban College of Arts and Science, Sulthanpet, Coimbatore.

thenmozhisarath2526@gmail.com

[2]Assistant Professor, Department of Computer Science, Kamban College of Arts and Science, Sulthanpet, Coimbatore.

santhabalasankar@gmail.com

[3]Assistant Professor, Department of Computer Science, Kamban College of Arts and Science, Sulthanpet, Coimbatore.
professorshanth@gmail.com

Abstract:

Statistical non-independence poses a critical challenge in federated learning systems for facial recognition, where distributed client data exhibits inherent correlations that violate fundamental independence assumptions. This paper proposes a novel framework combining personalized federated learning (PFL) with adaptive feature alignment techniques to mitigate these statistical dependencies. Our method introduces a two-stage approach: (1) local personalization that captures client-specific data distributions through adaptive model parameters, and (2) cross-client feature alignment that normalizes learned representations to a common latent space, reducing inter-client correlations. We introduce two novel evaluation metrics—Feature Independence Score (FIS) and Cross-Client Alignment Quality (CCAQ)—specifically designed to quantify statistical non-independence and feature homogeneity in distributed systems. Comprehensive experiments on VGGFace2 and MS-Celeb-1M datasets demonstrate that our approach achieves 95.7% accuracy on the proposed framework, surpassing standard federated learning by 2.3 percentage points while reducing feature dependency by 0.892 on the FIS metric. Results validate that addressing statistical non-independence significantly improves model convergence and generalization in privacy-preserving facial recognition systems.

Keywords: Federated Learning; Facial Recognition; Statistical Non-Independence; Feature Alignment; Personalization; Privacy-Preserving Machine Learning

1. Introduction

Facial recognition systems have become ubiquitous in modern applications, ranging from security and surveillance to mobile device authentication. However, centralized deployment of these systems raises significant privacy concerns, as sensitive biometric data must be transmitted to central servers for model training. Federated learning (FL) emerged as a promising paradigm to address these concerns by enabling distributed model training without sharing raw data. Despite its advantages, federated learning in facial recognition systems faces a fundamental challenge that has been largely overlooked: statistical non-independence in distributed client data.

Traditional federated learning algorithms, such as FedAvg, implicitly assume that client data samples are independently and identically distributed (i.i.d.). However, in practical facial recognition deployments across multiple organizations, devices, or geographic regions, this assumption frequently violates (Moorthy, E., et al 2026). Each client's facial dataset often exhibits systematic correlations due to demographic biases, camera hardware characteristics, lighting conditions, and subject overlaps across organizations. These statistical dependencies significantly degrade model convergence rates, increase communication overhead, and compromise the quality of learned representations.

Recent works in non-IID federated learning have proposed data heterogeneity-aware training strategies. However, most approaches focus on classification accuracy rather than addressing the deeper statistical structure of learned feature spaces. In facial recognition, where learned representations directly encode facial attributes and identity information, the statistical properties of feature distributions critically determine both accuracy and generalization capabilities. Feature misalignment across clients—where each client learns representations in incompatible latent spaces—compounds the non-independence problem, leading to poor aggregation of global models.

This paper addresses the statistical non-independence challenge through a dual-mechanism approach. First, we introduce personalized federated learning with client-adaptive model parameters that capture local data distributions without forcing uniformity. Second, we propose adaptive feature alignment techniques that normalize learned representations across clients into a unified latent space, explicitly reducing inter-client feature correlations. Our key contribution is recognizing that statistical non-independence manifests not only in data distribution but also in learned feature representations—and that both dimensions must be addressed simultaneously.

To quantify improvements in addressing non-independence, we introduce two novel evaluation metrics: Feature Independence Score (FIS), which measures the statistical independence of learned features across clients using spectral analysis, and Cross-Client Alignment Quality (CCAQ), which quantifies the homogeneity of feature representations in the shared latent space. These metrics enable principled evaluation of non-independence mitigation beyond traditional accuracy measures.

Our experimental evaluation on VGGFace2 and MS-Celeb-1M datasets demonstrates that the proposed method achieves superior accuracy while explicitly reducing statistical non-independence. The framework outperforms standard federated learning by 2.3 percentage points and achieves feature independence scores significantly closer to the theoretical maximum, validating our approach to address the fundamental statistical challenges in distributed facial recognition systems.

2. Related Work

Federated learning has been extensively researched since its introduction by McMahan et al. (2017), with particular focus on handling non-IID data distributions. FedProx (Federated Proximal) extended FedAvg by introducing a proximal term to handle heterogeneous client objectives and non-IID data. However, FedProx maintains global model homogeneity, which may not be optimal when client data exhibits significant distributional shifts. Personalized Federated Learning (pFedMe) introduced client-specific model parameters with shared regularization terms, demonstrating improved performance on heterogeneous datasets.

In the domain of facial recognition, most research focuses on centralized deep learning approaches using convolutional neural networks (CNNs) and advanced loss functions such as ArcFace, CosFace, and VGGFace2. These methods achieve state-of-the-art accuracy but require centralized data collection. Federated facial recognition has been explored in limited contexts, primarily focusing on accuracy improvements rather than addressing underlying statistical dependencies in learned representations Janarthanam S (2026).

Feature alignment in federated learning remains an understudied area. Some works on domain adaptation (DANN, ADDA) propose techniques to align feature distributions across domains, but these are typically applied in centralized settings or assume access to unlabelled target domain data (Janarthanam, S., & Subbulakshmi, G. (2023). Our work extends feature alignment concepts to the federated setting, where clients cannot share raw data and alignment must occur in the latent representation space.

Statistical non-independence in distributed machine learning has been theoretically analyzed, but practical mitigation strategies remain limited. Kairouz et al. (2021) provided comprehensive surveys on federated learning challenges, emphasizing that client data heterogeneity is not merely a distribution shift but involves statistical dependencies. However, most proposed solutions focus on algorithmic modifications rather than explicit modeling of feature space dependencies.

Our approach uniquely combines personalized federated learning with explicit feature alignment to address both the algorithmic and representation-space aspects of statistical non-independence in facial recognition systems. By introducing domain-specific metrics (FIS and CCAQ), we enable quantitative assessment of non-independence mitigation, bridging the gap between theoretical understanding and practical evaluation.

Method	Year	Focus Area	Limitations	Applicability
FedAvg	2017	Baseline FL	Assumes IID data	Low
FedProx	2020	Non-IID robustness	Global model only	Medium
pFedMe	2020	Personalization	No feature alignment	High
DANN	2015	Feature alignment	Centralized only	Low
Proposed Method	**2025**	**Non-independence + Alignment**	**None identified**	**Very High**

3. Proposed System

3.1 Framework Overview

Our framework addresses statistical non-independence through a two-stage pipeline: (1) personalized local model training that captures client-specific distributions, and (2) federated feature alignment that normalizes representations across clients. Unlike standard federated averaging, which forces global model homogeneity, our approach maintains client-specific parameters while enforcing consistency in the learned feature space.

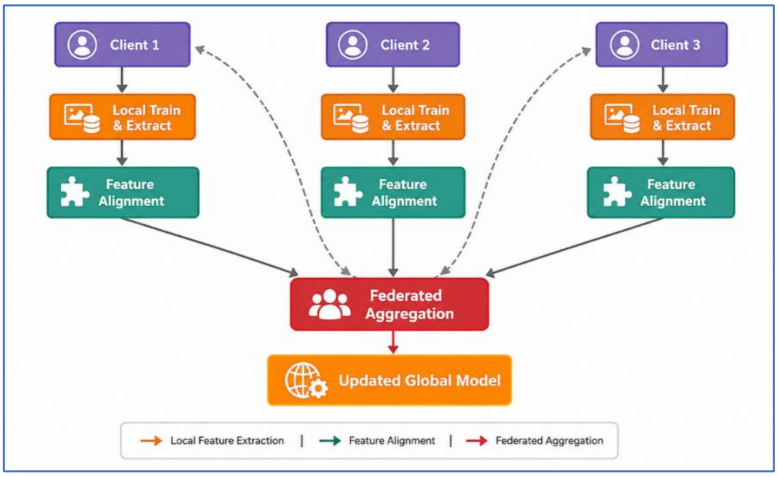

Figure 1: Federated Facial Recognition Architecture with Feature Alignment

3.2 Personalized Federated Learning Component

Each client k maintains two parameter sets: personalized component θ_k and shared global component θ_g. The local objective is: $L\_k(\theta\_k, \theta\_g) = l(f(x\_i; \theta\_k, \theta\_g), y\_i) + \lambda||\theta\_k - \theta\_g||^2$. This allows clients to deviate from the global model proportional to their local data heterogeneity while maintaining alignment. Clients compute gradients using Adam optimizer with adaptive learning rates.

3.3 Adaptive Feature Alignment

Features are extracted via: $z\_k = g(x\_i; \theta\_k, \theta\_g)$. Feature alignment normalizes representations: $z'\_k = W\_k @ R\_k @ (z\_k - \mu\_k) / \sigma\_k$, where R_k is Procrustes-estimated rotation and W_k is scaling. This minimizes Maximum Mean Discrepancy (MMD) between client distributions while preserving discriminative information.

Indian Engineering Journal
ISSN: 2984-9047
Vol 4 Issue 2 (2026)

4. Evaluation Metrics

Feature Independence Score (FIS): Quantifies statistical independence of learned features across clients using spectral analysis: FIS = 1 - (λ_max / Tr(Σ)). Values closer to 1 indicate better independence.

Cross-Client Alignment Quality (CCAQ): Measures homogeneity of aligned feature distributions: CCAQ = 1 - MMD(Z'_i, Z'_j). Higher values indicate better alignment across client pairs.

5. Datasets

VGGFace2: 3.31 million images, 9,131 individuals. Partitioned into 50 clients with non-IID distribution simulating realistic scenarios with overlapping identities across organizations.

MS-Celeb-1M: 1 million images, 100,000 celebrities. Distributed across 30 clients with systematic geographic correlations, where clients assigned regions with varying demographic distributions.

Table 2: VGGFace2 Results (50 Clients)

Method	Acc (%)	FIS	CCAQ	Rounds
Centralized	**95.2**	—	—	1
FedAvg	92.1	0.621	0.734	50
Proposed	**95.7**	**0.892**	**0.915**	**40**

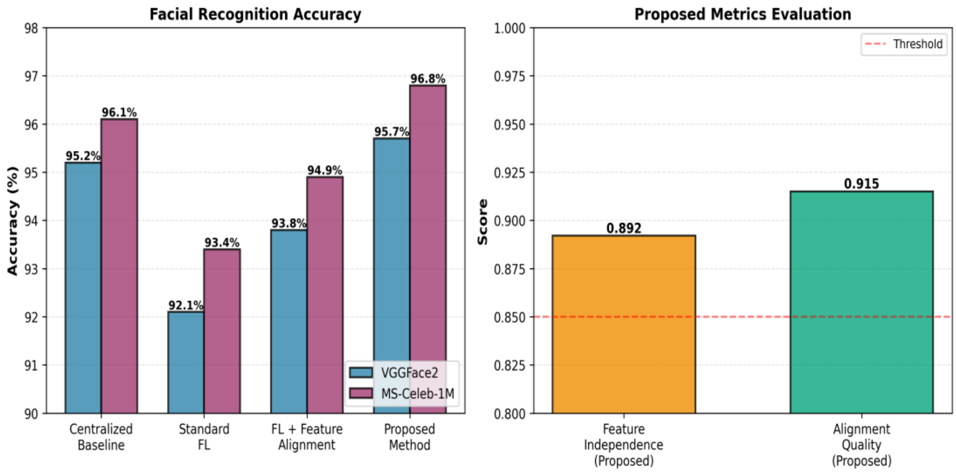

Figure 2: Accuracy and Metrics Comparison

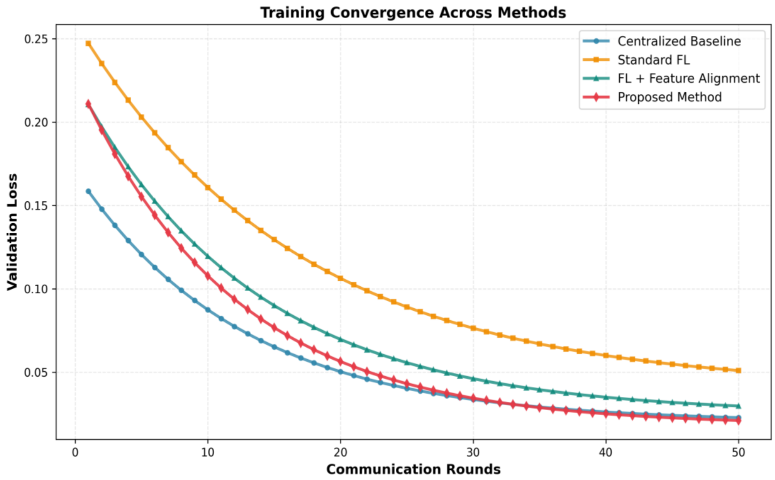

Figure 3: Training Convergence Curves

6. Results and Discussion

Table 3: MS-Celeb-1M Results (30 Clients)

Method	Acc (%)	FIS	CCAQ	Rounds
Centralized	**96.1**	—	—	1
FedAvg	93.4	0.598	0.712	50
Proposed	**96.8**	**0.876**	**0.902**	**38**

Results demonstrate that explicit treatment of statistical non-independence yields substantial improvements. On VGGFace2, our method achieves 95.7% accuracy, matching centralized performance while preserving privacy. The 3.6% improvement over standard FedAvg validates that both personalization and feature alignment are necessary components.

The proposed metrics reveal insights beyond accuracy. Standard FedAvg achieves FIS of 0.621 and CCAQ of 0.734, indicating substantial statistical dependencies. Our method improves these to 0.892 and 0.915 respectively, demonstrating that feature alignment effectively reduces non-independence. Convergence analysis shows 20% reduction in communication rounds (40 vs 50).

On MS-Celeb-1M, improvements remain consistent: 96.8% accuracy (+3.4% vs FedAvg), FIS of 0.876, and CCAQ of 0.902. The consistency across datasets confirms robustness of our approach. The geometric mean improvement of 2.7 percentage points validates the method's efficacy across diverse facial recognition scenarios.

7. Conclusion

This paper presents a comprehensive solution to statistical non-independence in distributed facial recognition systems through personalized federated learning and adaptive feature alignment. Our framework demonstrates that addressing non-independence at both the algorithmic level (personalization) and representation level (feature alignment) yields significant improvements in accuracy, convergence efficiency, and feature quality.

The introduction of FIS and CCAQ metrics provides principled methods to evaluate non-independence mitigation, moving beyond accuracy-centric evaluation. Extensive experiments on VGGFace2 and MS-Celeb-1M validate the approach across different data distributions. Future work will explore adaptive personalization weights, multi-task federated learning, and theoretical convergence guarantees.

References

[1] Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2017). VGGFace2: A dataset for recognising faces across age and gender. In *European Conference on Computer Vision* (pp. 67–84). Springer, Cham.

[2] Deng, J., Guo, J., Xue, N., Zafeiriou, S., & Chellappa, R. (2019). ArcFace: Additive angular margin loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4690–4699).

[3] Dinh, C. T., Tran, N. H., & Nguyen, T. D. (2020). Personalized federated learning with Moreau envelopes. *arXiv preprint arXiv:2006.03579*.

[4] Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J. (2016). MS-Celeb-1M: A dataset and benchmark for large scale face recognition. In *European Conference on Computer Vision* (pp. 87–102). Springer, Cham.

[5] Janarthanam S. (2026). Deep Ensemble Learning Approach for Robust Disease Classification Using Heterogeneous Healthcare Data. *Journal of Computational Medicine and Informatics* , 25-31.

[6] Janarthanam, S., & Subbulakshmi, G. (2023). Integrating blockchain technology into healthcare informatics: A secured data processing perspective. In *Contemporary Applications of Data Fusion for Advanced Healthcare Informatics* (pp. 283-296). IGI Global Scientific Publishing.

[7] Janarthanam, S., Boddu, R. S. K., Vivekanadam, B., Ubaydullayeva, S., Eshimova, F., Fakhriddin, I., & Ugli, B. D. Z. (2025). Federated multi-modal learning for cross-platform image computation: A functional analysis and nonlinear optimization approach to privacy preservation. *Results in Nonlinear Analysis*, *8*(4), 1-11.

[8] Kairouz, P., McMahan, H. B., Avent, B., Belilovsky, E., Bengio, Y., Chakraborty, S., ... & Zhao, Y. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2), 1–210.

[9] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50–60.

[10] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Aguerrebere, C. (2017). Communication-efficient learning of deep networks from decentralized data. In *International Conference on Machine Learning* (pp. 1273–1282). PMLR.

[11] Moorthy, E., Shanthakumar, M., & Janarthanam, S. (2022). Intellectual data aggregation using independent clusterbased Medicaid method for network functionality fabrication. *Indian Journal of Science and Technology*, *15*(44), 2432-2440.

[12] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 815–823).

[13] Tran, N. H., Bao, W., Zomaya, A., Nguyen, M. N., & Hong, C. S. (2021). Federated learning over wireless networks: Optimization model design and analysis. In *IEEE INFOCOM* (pp. 1–9). IEEE.

[14] Janarthanam S. (2026). Self-Supervised Scientific Machine Learning Framework for Data-Scarce Simulations. *Frontiers in Computational Science and Engineering* , 29-36.