



Towards federated learning for face recognition: a comprehensive review of secure and privacy-preserving techniques and prospective research

Ajnas Muhammed¹ · João Marcos¹ · Nuno Gonçalves¹

Received: 30 November 2025 / Accepted: 18 August 2026
© The Author(s) 2026

Abstract

Federated learning is a popular decentralized machine learning paradigm that enables multiple devices or entities to collectively train a common model without sending data to a central server. The distributed data availability improves privacy and security for each participating entity by keeping their data localized. Face recognition, a rapidly evolving field, uses deep learning techniques to achieve remarkable advances, frequently outperforming human performance on specific public datasets. However, the sensitive nature of facial data, as faces belong to personally identifiable information, raises serious privacy and security concerns on training data sharing. Federated learning has emerged as a promising solution to the privacy and security issues confronting the face recognition community. This paper provides a systematic, PRISMA-guided review of 50 publications released between 2020 and 2026, identified through searches of IEEE Xplore, ACM Digital Library, SpringerLink, arXiv, Google Scholar, and DBLP, on face recognition frameworks that use federated learning. The reviewed techniques are systematically classified into six categories, with an emphasis on their contributions and relevance to the larger domain of federated learning based face recognition. This work aims to summarize and analyze various federated facial recognition methods, including their underlying techniques, results, and goals. This survey also compares against twelve existing face recognition surveys to clarify its distinct and unique contributions. Furthermore, it provides a high-level detailed overview of how various federated learning functionalities and design principles have been used in face recognition applications. This review identifies key challenges while highlighting promising research directions for future advancements in the field.

Keywords Federated learning · Face recognition · Security · Privacy

1 Introduction

Federated Learning (FL) is an innovative decentralized machine learning training framework, introduced by McMahan et al. [1], that allows multiple participants to train a single model on distributed local datasets. FL is especially useful in different fields where data privacy, security, and regulatory compliance are critical, such as healthcare, finance, biometrics, and many more. In FL, by storing data locally and sharing only model updates, the framework reduces the risk of data breaches, improves privacy and security, and ensures compliance with regulations such as the European Union's General Data Protection Regulation (EU-GDPR) [2], California Consumer Privacy Act (CCPA) and its extension California Privacy Rights Act (CPRA) [3], and others. Adherence to the principles supported by various

João Marcos and Nuno Gonçalves have contributed equally to this work.

✉ Ajnas Muhammed
ajnas.muhammed@isr.uc.pt

João Marcos
joao.marcos@isr.uc.pt

Nuno Gonçalves
nunogon@deec.uc.pt

¹ Institute of Systems and Robotics, University of Coimbra,
3030-290 Coimbra, Portugal

regulations, such as the EU-GDPR and CCPA, is critical for any technology that handles personal data, including FL systems. Several studies have investigated how FL can be designed to meet different regulatory requirements [4]. Furthermore, as FL keeps the data local, it reduces the need for conventional large-scale data transfers, which saves bandwidth and computing resources. Due to these numerous benefits, FL has become a popular topic for academic research and industry applications [5].

Figure 1 depicts the centralized Learning (CL) versus FL. In CL, the initial step is to collect the training data from all participating entities (or clients). The next step involves transfer of all the collected data to the centralized location for model training. Later, the trained model is passed to the clients based on the design framework and/or authentic request. Whereas in FL, the framework works differently. Here, an initial model created by the server is distributed to participating clients. Each client performs local training with the local data and only the model updates are passed to the central server. Upon getting the local model updates, server aggregates to build a global model, and this process continues until the model converges.

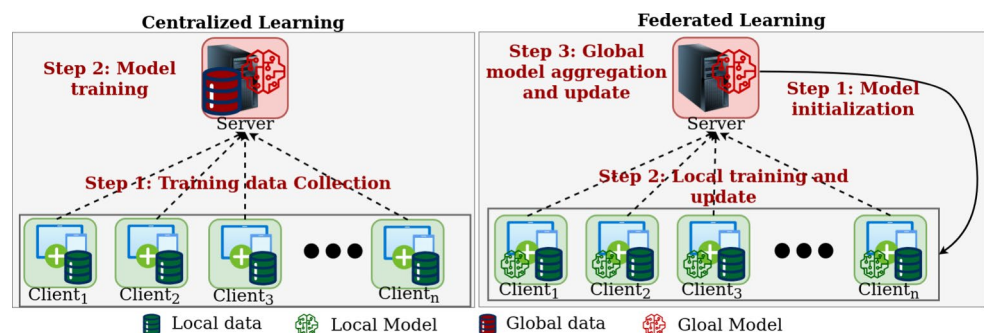
Face Recognition (FR) systems detect and verify individuals by examining their distinctive facial characteristics (or features). Over the years, FR has experienced remarkable enhancements in performance due to advances in machine learning algorithms, the accessibility of extensive datasets, and the augmented computational capabilities of contemporary hardware [6]. Recently, especially after COVID-19 pandemic, FR has emerged as a core element in modern security applications [7], social media platforms [8], border management, and numerous other crucial and critical applications [9]. Despite this popularity, FR systems pose substantial security and privacy concerns. Conventional FR methods rely on centralized repositories to gather and analyze sensitive facial data, increasing exposure to single-point attacks and data breaches [10]. As faces are part of human anatomy and face revocation is highly challenging and costly, privacy preservation in FR is of paramount importance. Unauthorized access or misuse may result in severe consequences such as surveillance overreach, identity theft, and abuses of individual freedom [11]. Consequently,

there is an increasing demand for methods capable of effectively safeguarding privacy while ensuring the accuracy and dependability of FR systems [10].

Although FR systems have attained notable accuracy and extensive applications, they encounter substantial challenges concerning data privacy, security, and regulatory adherence. Recently, the FR community has adopted FL as a promising approach to mitigate urgent privacy issues. FL enables decentralized data processing and model training, guaranteeing that sensitive facial data stays at local devices instead of being transmitted to a central server. This method not only preserve privacy but also enhance security by substantially decreasing the likelihood of data breaches. Furthermore, FL facilitates collaboration among diverse datasets, promoting the development of more robust and inclusive FR systems. By maintaining data localized and sharing solely model updates, FL adheres to rigorous privacy regulations, rendering it a feasible framework for developing ethical and compliant FR technologies. The rapid advancements in FR within FL have generated a pressing need for a comprehensive survey paper in this field.

Existing FR surveys consist of twelve related studies: [10–20, 22]. References [14] and [15] are surveys of deep learning-based FR, discussing architecture, loss functions and training pipelines. However, they remain centered on centralized training settings. Zeng et al. [12] discusses the technique focuses on privacy-preserving biometric identification and authentication, leveraging Homomorphic Encryption (HE) and Multi-Party Computation (MPC). Wang et al. [13] classifies privacy-preserving FR techniques into four paradigms: appearance-guided, identity-guided, reversible, and privacy-preserving FR systems. Surveys [10] and [11] discusses privacy leakage and presentation-level protection techniques on FR domain. Leyva et al. [16] and Roux et al. [17] discusses FR security through a broad attack-vector taxonomy and a focused review of backdoor attacks and defenses. Surveys [18] and [19] discusses democratic and racial fairness in FR. Hahn and Marcel [20] reviews biometric template protection techniques, and [22] conducts a systematic review of Convolutional Neural Network (CNN) architectures, databases, and performance metrics used in FR. Table 1 summarizes the scope, FL coverage, and key

Fig. 1 Block diagram representing CL vs. FL



gaps of these twelve surveys. Five of twelve surveys touch FL to some degree. References [10, 16], and [17] discuss FL only tangentially amid broader security concerns. Zeng et al. [12] and Wang et al. [13] mention it explicitly without focusing it as a main theme. The remaining seven do not discuss it at all. Hence, none of these surveys discussed FL as the primary axis. Therefore, this paper presents a comprehensive survey to systematically organize, and analyze the scope of FL on FR domain. With this review, our objective is to discuss the principal contributions, challenges and future works on this domain. This survey will function as a significant resource for researchers, practitioners, and policymakers engaged in the field of FL and FR. Table 2 presents the various papers examined in this survey, along with their publication years and primary contributions.

2 Review methodology

This paper follows a structured, PRISMA-style protocol with three search pattern: an initial database search (Round 1), a DBLP supplementary search (Round 2), and

an additional records identified outside the formal database queries (Round 3). Figure 2 summarizes the complete flow.

2.1 Databases searched

Round 1 searched IEEE Xplore, ACM Digital Library, SpringerLink, arXiv, and Google Scholar. Round 2 (y) queried DBLP (dblp.org) directly by screening results manually one at a time. Round 3 consists of a small number of records identified through other means (e.g., citation snowballing, author familiarity with the field).

2.2 Search strategy and keywords

Round 1 search strings combined terms from two concept groups, e.g.: (“federated learning” OR “FL”) AND (“face recognition” OR “facial recognition” OR “biometric recognition”), refined with secondary terms such as “privacy-preserving,” “secure aggregation,” “differential privacy,” “non-IID,” “fairness,” “domain adaptation,” and “edge/IoT deployment”, identifying 187 records. Round 2 ran two literal DBLP queries: (i) “*privacy face recognition*”

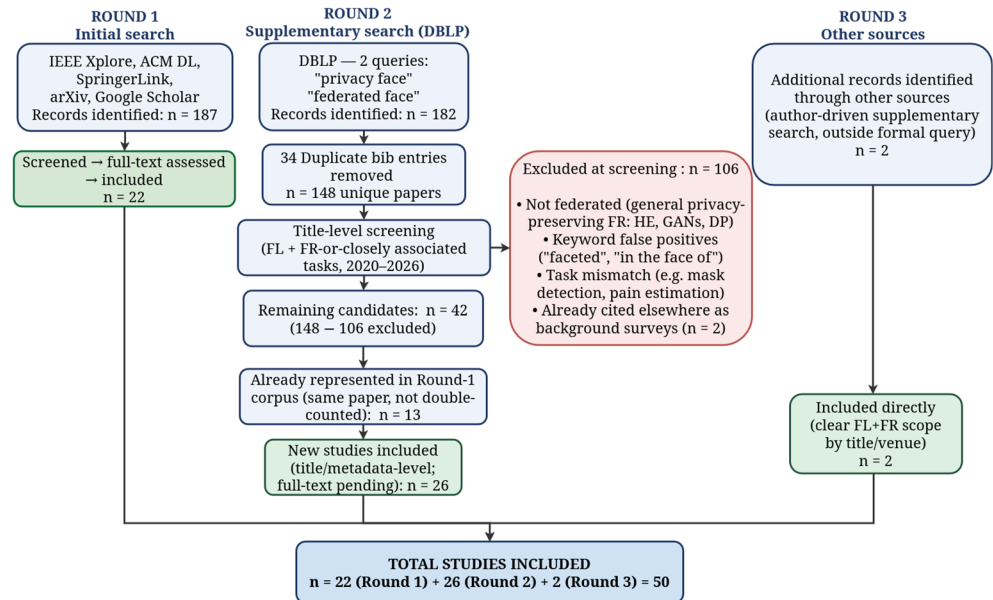
Table 1 The scope, FL coverage, key gaps of existing surveys on related field

Reference	Year	Scope	FL coverage	Key gaps
[10]	2025	Privacy leakages and solutions across the FR pipeline	Not centrally addressed	Organized by leakage/solution type, not by FL method; no FL-specific taxonomy or cross-method benchmark comparison
[11]	2023	Presentation-level privacy protection techniques	Not addressed	Scope limited to presentation-level (not training-level/federated) privacy protection
[12]	2025	Privacy-preserving biometric authentication protocols	Mentioned, not central	Broad biometrics scope, not FR-specific; not organized around FL
[13]	2025	Facial privacy via four paradigms	Mentioned, not central	Organized by privacy paradigm, not by FL method
[14]	2022	Broad, end-to-end deep FR (architectures, losses, pipelines)	Not addressed	Centralized-training-focused throughout; does not address distributed/federated training
[15]	2021	Broad deep FR survey (widely-cited reference)	Not addressed	Foundational but centralized; predates most FL-for-FR work
[16]	2026	Attack-vector taxonomy for FR systems	Not centrally addressed	Covers attack surface broadly; does not address FL-specific threats
[17]	2024	Backdoor attacks and defenses in FR	Not centrally addressed	Conceptually adjacent to our Sybil-attack discussion but not set in a federated/distributed training context
[18]	2026	Demographic fairness in FR	Not addressed	Fairness surveyed independent of training paradigm; does not connect fairness to FL-specific factors (client heterogeneity)
[19]	2025	Racial bias in FR	Not addressed	Fairness surveyed independent of training paradigm; does not connect fairness to FL-specific factors such as client heterogeneity
[20]	2023	Biometric template protection	Not addressed	Directly relevant to DeepShieldFed [21], but organized around template-protection technique generally, not federated training
[22]	2025	Systematic review of CNN architectures, databases, metrics for FR	Not addressed	Closest methodological peer (also a systematic review); organized around architecture/database taxonomy rather than training paradigm, and not FL-specific

Table 2 Details regarding the works reviewed in this survey

References	Year	Key contributions
[23]	2021	Early strategy paper proposing FL for FR training (labeling, data management, on-device evaluation, asynchronous update buffering); no experimental evaluation
[24]	2022	Proposed a gradient correction mechanism to address gradient divergence and improve FFR performance
[25]	2021	Integrated differential privacy into FL-based FR to ensure robustness and privacy
[26]	2021	Proposed a technique combining privacy, fairness, and explainability
[27, 28]	2021	Demonstrated FL for domain adaptation in FR while preserving data privacy
[29]	2021	Used ensemble models in FL for improved transferability and privacy
[30]	2022	Demonstrated feasibility of deploying FL-based FR in resource-constrained industrial settings
[31]	2022	Enhanced model performance and privacy by mitigating non-IID data effects. Also published at ICLR 2022
[32]	2024	Improved performance on imbalanced datasets via adaptive inter-class representation learning
[33]	2023	Demonstrated decentralized FL for FR, highlighting privacy and scalability advantages
[34]	2024	Introduced a self-supervised technique reducing reliance on labeled data. Also published at BMVC 2024
[35, 36]	2024	Proposed a privacy-preserving FL framework using differential privacy and secure aggregation
[37, 38]	2024	Introduced meta-learning to improve performance on imbalanced datasets and generalization
[39]	2023	Integrated FL-based FR with electrostatic discharge (ESD) testing
[40]	2024	Introduced a Sybil-aware FL framework (SaFL) that detects and mitigates malicious clients
[41]	2024	Introduced a supervised FL framework using secure aggregation
[42]	2024	Demonstrated real-world feasibility of differential privacy in FL-based FR
[43]	2025	Explored trade-offs between privacy, accuracy, and fairness using differential privacy and fairness-aware methods
[44]	2025	Proposed an adaptive FL framework using actor-critic client selection for privacy/fairness
[45]	2024	Compares multiple FL methods for FR specifically under non-IID data conditions
[46]	2024	Examines information-leakage risks in facial and object recognition operations conducted over FL
[47]	2026	Proposes a prototype-based framework for open-set heterogeneous FL based FR
[48]	2026	Proposes PriSecFedFR, combining FL with random projection for privacy-secure FR model training
[49]	2025	Proposes FFDL, a feature-fusion deep learning method utilizing FL for forged face detection
[50]	2025	Proposes Fed-StackFPAD, combining FL with stacking to address data heterogeneity in face presentation attack detection
[51]	2025	Proposes an adaptive FL approach for unsupervised multidomain FR using a bio-inspired optimization algorithm
[52]	2025	Combines self-supervised learning with FL for few-shot FR
[53]	2025	Proposes pFedFace, a personalized FL approach for FR
[21]	2025	Proposes DeepShieldFed, securing face templates via deep cancelable transforms combined with FL
[54]	2025	Explores unsupervised FL for FR in decentralized environments
[55]	2025	Proposes FORGET ME, a federated unlearning approach for face generation models
[56]	2024	Proposes a federated framework for generalized face presentation attack detection
[57]	2024	Proposes CL-FedFR, applying curriculum learning to FL based FR
[58]	2024	Proposes a federated face forgery detection approach using personalized representation learning
[59]	2023	Proposes FedForgery, a generalized face forgery detection method using residual FL
[60]	2023	Proposes FedSIS, combining federated split learning with intermediate representation sampling for face presentation attack detection
[61]	2023	Proposes F3, a fair and federated approach to face attribute classification under heterogeneous client data
[62]	2022	Proposes FedFV, performing federated face verification via equivalent class embeddings
[63]	2022	Proposes a dynamic-aware FL approach for face forgery video detection
[64]	2022	Proposes FedFR (Liu et al.), a joint-optimization federated framework for generic and personalized FR
[65]	2022	Examines robustness of smart-home federated FR under data-starved client conditions
[66]	2021	Proposes federated test-time adaptive face presentation attack detection with dual-phase privacy preservation
[67]	2021	Presents an early FL approach for FR (distinct from [23])
[68]	2021	Proposes a fair FL approach for heterogeneous face data (likely precursor to [61], overlapping authors)
[69]	2021	Proposes Federated Active Authentication (FAA); shares per-user feature statistics rather than raw features or full model updates to handle one-identity-per-device federated authentication
[70]	2020	Proposes FedGP, a federated GAN for privacy-preserving synthetic data release; evaluates resistance to model inversion via face detection/recognition rates on CelebA
[71]	2020	Presents an early FL approach for face anti-spoofing

Fig. 2 PRISMA-style flow diagram of the study selection process across all three identification streams. Total included: $n = 22$ (Round 1) + 26 (Round 2) + 2 (Round 3) = 50



and (ii) “federated ...face” (a DBLP proximity-style query matching “federated” and “face” co-occurring in a title), returning 182 raw records combined.

2.3 Inclusion and exclusion criteria

Inclusion criteria are as follows:

- Peer-reviewed articles, peer-reviewed conference papers, or arXiv preprints with technical contributions towards the intersection of FL and FR, or a closely-adjacent face-security task (presentation attack detection, anti-spoofing, forgery/deepfake detection).
- Published between January 2020 and 2026.
- Written in English.

Exclusion criteria are as follows:

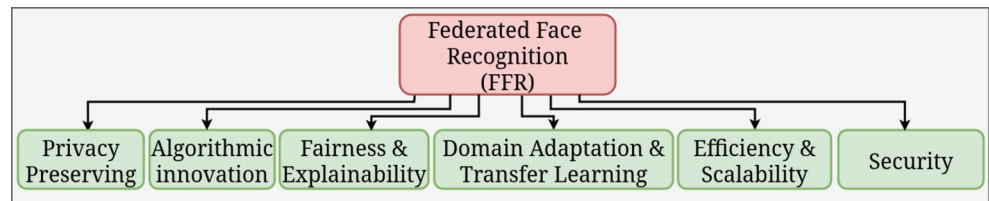
- Privacy-preserving or security-focused FR techniques without FL.
- Titles matching search keywords only incidentally. e.g., “multi-faceted” or “two faces of interference” matching on “face” as a substring or idiom rather than referring to facial recognition.
- FL papers using face imagery for a task other than identity recognition or a closely-adjacent security task (e.g., pain estimation, emotion recognition, mask-wearing detection).
- Workshop abstracts, posters, or papers without sufficient methodological detail to extract results.

Applying these criteria across all three rounds leaves the corpus summarized in Table 2: 50 studies, spanning 2020

to 2026. Two entries [23, 46] are conceptual or architectural contributions without empirical evaluation and are flagged as such where discussed. The next section groups these studies into six categories and discusses each in turn.

3 Federated face recognition

This section classifies 50 studies of Federated Face Recognition (FFR) methodologies into six categories: (1) privacy-preserving techniques (Sect. 3.1), (2) algorithmic innovations (Sect. 3.2), (3) fairness and explainability (Sect. 3.3), (4) domain adaptation and transfer learning (Sect. 3.4), (5) efficiency and scalability (Sect. 3.5), and (6) security—presentation attack detection and forgery detection (Sect. 3.6) as shown in Fig. 3. This classification is based on the primary area of focus, research direction and contribution. Privacy-preserving techniques include methods and frameworks exclusively aimed at improving data confidentiality within FFR systems. Algorithmic innovations emphasize innovative algorithms and optimization techniques designed to enhance the performance and robustness of FFR models. The fairness and explainability section examines methodologies that address ethical issues, bias reduction, and transparency within FFR systems. Domain adaptation and transfer learning emphasize methods that allow FR models to generalize efficiently across various domains and datasets. The efficiency and scalability section addresses methodologies for optimizing communication, minimizing latency, and improving computational performance in FFR systems. Finally, the security category examines methodologies for detecting presentation attacks and synthetic-media forgery within FFR systems, addressing

Fig. 3 FFR taxonomy diagram with all six categories

threats directed at the recognition target itself rather than at the training process.

Due to the fact that FFR models are trained locally on distributed devices and only model updates, rather than raw facial data, are shared, a direct comparison of the performance of two FFR techniques presents a challenge. The efficacy of FFR is contingent upon local data distribution, device characteristics, and the employed aggregation and communication protocols. The disparate operating environments, encompassing varying data privacy regulations, device specifications, and network parameters, inherent to each federated system render direct performance comparisons unfeasible. In a similar vein, the absence of a standardized benchmark or centralized dataset for evaluation complicates the comparison of different FFR systems. The FFR systems exhibit notable performance trade-offs relative to centralized (non-FFR) systems. Centralized, extensive datasets contribute to enhanced recognition performance. FFR facilitates facial data localization to enhance privacy and security; however, restricted facial data diversity and potential biases within local datasets may compromise performance. Given the inherent complexities and associated challenges, a fully controlled performance comparison between different FFR systems is not attempted in this study. However, for the subset of reviewed studies that report results on shared, widely-used benchmarks (LFW, IJB-B, IJB-C), Sect. 3.7 consolidates these already-published figures into indicative cross-method comparison tables, subject to the limitations regarding non-standardized training data, backbones, and client partitioning.

3.1 Privacy-preserving techniques

Privacy preservation methods in FFR focus on protecting sensitive facial data while maintaining optimal model performance. A method for preserving privacy of transferable facial images utilizing FL and ensemble models is described in [29]. This method establishes an FL model based on a distributed dataset. On the client side, a locally trained model based on the local data is provided as input to PcadvGAN [72]. On the server side, the discriminators of PcadvGAN and the FL model are combined through ensembling. Last, the global optimal aggregation parameter matrix of the GAN is computed. This method is capable of generating new, private facial images within a defined design framework. The

Table 3 FR accuracy and privacy-attack success rates (targeted/non-targeted) across five backbone architectures [29]

FR model	Accuracy (%)	Targeted attack	Non-targeted attack
VGG16	98.43	0.3901	0.2776
VGG19	98.92	0.3823	0.2765
Resnet18	97.91	0.3774	0.2855
Resnet50	99.01	0.3792	0.2799
Senet50	98.87	0.3845	0.2892

Table 4 Face detection and recognition rates (%) for images reconstructed via model inversion attack, CelebA [70]

Metric	Baseline (non-private)	FedGP
Detection	25.5	1.2
Recognition	2.8	0.1

experiments and results indicate that the generated facial images improve transferability and practicality. Table 3 displays the technique's experimental results, including the model, accuracy, and private image structural similarity (SSIM) for both targeted and non-targeted attack scenarios. They also contend that the implementation of Differential Privacy (DP) or HE could be employed to strengthen security and facilitate transferability.

FedGP [70] trains a federated GAN with FedAvg to synthesize artificial data that can replace clients' real data used for training. With this, no client ever shares their raw data that maps directly back to any individual. Beyond a differential-average-case-privacy analysis, the paper specifically stress-tests the approach against the model inversion attack on the CelebA [73] dataset. Reconstructed images are passed through OpenFace to measure how much identity information survives the attack. Table 4 shows that FedGP sharply reduces both detection and recognition success on reconstructed faces compared to a non-private baseline model trained directly on real images.

PrivacyFace, a framework developed to improve FFR by enabling the exchange of auxiliary and privacy-preserving information among clients, is presented in [31]. PrivacyFace incorporates an effective Differentially Private Local Clustering (DPLC) mechanism to derive sanitized clusters from local class centers, complemented by a consensus-aware recognition loss that improves the discriminative efficacy of the outcomes. PrivacyFace mathematically demonstrates that the technique introduces minimal overhead while delivering significant performance enhancements. Experimental results indicate that the technique incurs minimal overhead

while delivering enhanced performance. Table 5 presents the verification performance of various benchmarks employed in the experiment, reported as the True Acceptance Rate (TAR) at a fixed False Acceptance Rate (FAR) threshold. In Table 5, conventional FL (fine-tuned using the BUPT-Balancedface dataset [74]) is compared with PrivacyFace employing both CosFace and ArcFace loss functions. The experiments conducted on both loss functions demonstrate that PrivacyFace outperforms conventional FL across both loss functions.

An alternative FFR approach that integrates FL with both supervised and unsupervised FR classifiers, while preserving user privacy, is described in [35, 36]. In [35, 36], the authors introduce FL framework with and without aggregators, within the context of both supervised and unsupervised systems. Here, each local model is either transmitted to a secure aggregator (if available) or directly to the central server. A GAN model is utilized on edge devices to produce synthetic data. This process reduces the transmission load on the available bandwidth and reveals potential vulnerabilities through the disclosure of specific information regarding the local user. Subsequently, the secure aggregator or central server aggregates these individual models to form a comprehensive global model, which is then communicated back to the edge devices. The experimental findings on the CelebA [73] dataset demonstrate that FL offers substantial benefits in both supervised and unsupervised FR tasks. Experimental results demonstrate that implementing FFR system offers benefits such as improved privacy, since the original data remains on edge devices, and the aggregated model achieves performance nearly comparable to that of individual models, particularly when the FL model does not utilize a secure aggregator. Table 6 presents the resulting Equal Error Rate (EER) across supervised and unsupervised systems, with and without a secure aggregator and GAN-based synthetic data generation.

A privacy protection framework for FR, DPFedFace [42], is introduced for practical scenarios where each device exclusively holds the owner's facial data (one identity per client). The method protects user privacy by eliminating visually significant low-frequency components based on human and model perception. They also established a privacy cost allocation mechanism for learning that optimizes allocation strategies and introduces noise into frequency domain features, thereby safeguarding user privacy. Table 7 presents the experimental results, including accuracy between fixed and learnable privacy cost allocation methods, and PSNR values with and without low-frequency components. The experimental results indicate that the application of DP and noise addition enabled the method to achieve high recognition accuracy while ensuring robust privacy protection, thereby preventing an attacker from

Table 5 Verification accuracy (%) on RFW, IJB-B, and IJB-C, comparing conventional FL against PrivacyFace under CosFace and ArcFace losses [31]

Benchmarks	CosFace		ArcFace	
	Conventional FL	PrivacyFace	Conventional FL	PrivacyFace
RFW African	83.40	86.30	83.50	87.32
RFW Asian	83.38	84.58	83.08	84.55
RFW Caucasian	89.62	91.48	90.26	92.03
RFW Indian	87.23	88.92	87.32	88.90
IJB-B (TAR@FAR = 1e-4)	68.09	77.35	58.62	71.26
IJB-C (TAR@FAR = 1e-4)	70.16	83.20	60.98	79.74

Table 6 Equal error rate (EER, %) for supervised and unsupervised FFR, with and without secure aggregation and GAN-based synthetic data [35, 36]

System		With secure aggregator	Without secure aggregator
Supervised	Individual Model	2.11	
	FL model without GAN	2.61	1.55
	FL model with GAN	2.73	1.98
Unsupervised	Individual Model	2.57	
	FL Model	3.07	2.35

Table 7 Verification accuracy and privacy metrics (fixed vs. learnable privacy cost allocation, PSNR) across five benchmarks [42]

Benchmarks	Accuracy	Accuracy among different privacy cost allocation methods		PSNR	
		Fixed	Learnable	With low frequency	Without low frequency
LFW	98.48	75.96	98.52	51.529	20.512
CFP-FP	89.37	72.16	90.39	51.307	20.546
AgeDB-30	90.13	70.28	90.73	52.464	21.010
CALFW	90.95	—	—	—	—
CPLFW	98.54	—	—	—	—
IJB-B (TAR@FAR = 1e-3)	83.42	—	—	—	—
IJB-C (TAR@FAR = 1e-3)	87.76	—	—	—	—

reconstructing face images from the uploaded class embeddings, thereby improving the security and privacy.

In [44], the authors introduce a hierarchical FL framework designed to improve client selection and model aggregation within human-computer interaction systems, with a particular emphasis on FR. The proposed framework utilizes an Actor-Critic-based dynamic Client Selection mechanism. This approach leverages reinforcement learning to

adapt client participation in real-time throughout the training process, thus optimizing the equilibrium between exploration and exploitation. Consequently, model accuracy and training efficiency are enhanced. The multidimensional weighted aggregation methodology improves the resilience and efficacy of the global model by incorporating considerations such as data volume, loss values, and uncertainty within the framework of heterogeneous data. Table 8 presents the training accuracy achieved for both the CelebA [73] and WISDM [75] datasets under Random Selection (RS) and dynamic Actor-Critic client selection methodologies. Empirical results across diverse datasets demonstrate that the proposed methodology achieves accelerated convergence and enhanced accuracy compared to traditional RS approaches.

In [48], a privacy-secure FFR training method called PriSecFedFR is presented. The technique solves two independent problems: sharing non-IID data between clients and preventing biometric privacy leakage. To solve with the former problem, the server pre-trains a global model on CASIA-WebFace [76] dataset. For each round, it selects negative samples from the dataset for each client based on Euclidean distance in the feature embedding space. In this way, a client with only one local name can still figure out both positive and negative loss terms. The local model's drift related to the global model is limited even more by a contrastive regularization loss. In order to protect privacy, random projection (RP) is used to transform client feature embeddings before local training. RP is non-invertible and fulfills the properties of ideal biometric template protection (irreversibility, revocability, and unlinkability). RP also keeps pairwise distances almost similar. Table 9 reports the method's accuracy relative to a centralized upper bound and three federated baselines (FedFV, FedAwS, FedUV) across three benchmarks.

In [21], a federated cancelable-template protection framework for FR called DeepShieldFed is presented. A 512-D face embedding is extracted from each client using the InsightFace buffalo_1 model. Only the 96 most discriminative dimensions selected based on low intra-class variance and high inter-class separability are retained. These dimensions are then binarized using mean-thresholding, and the resulting bit-string is combined with a randomly generated 32-bit user-specific key through an XOR operation to generate a cancelable template. Local CNN classifiers are trained on these templates instead of raw embeddings. According to the authors, this is the first work to combine cancelable biometrics with FFR. It is tested against the ideal biometric template protection properties of revocability, unlinkability, security, and performance. Table 10 compares recognition accuracy, FAR, False Rejection Rate (FRR), and Genuine

Table 8 Training accuracy on CelebA and WISDM under Random Selection vs Actor-Critic client selection, cloud and edge settings [44]

Dataset	Client selection	Accuracy (%)
CelebA	Cloud RS	85.50
	Edge RS	81.70
	Cloud AC	84.34
	Edge AC	84.28
WISDM	Cloud RS	84.40
	Edge RS	79.14
	Cloud AC	88.26
	Edge AC	83.32

Table 9 Verification accuracy across LFW, AgeDB30, and CFP-FP, comparing PriSecFedFR against centralized upper bound and three federated baselines [48]

Method	LFW (%)	AgeDB30 (%)	CFP-FP (%)
Centralized (upper bound)	76.87	66.65	74.23
Federated, without RP	75.28	66.23	73.71
FedFV	73.55	55.32	58.56
FedAwS	72.87	57.53	69.56
FedUV	73.75	56.13	67.89
PriSecFedFR	74.23	64.54	72.31

Table 10 Comparison with state-of-the-art face template protection techniques, reporting accuracy, FAR, FRR and GAR [21]

Approach	Database	Accuracy (%)	FAR (%)	FRR (%)	GAR (%)
[77]	LFW	99.5	–	–	–
[78]	LFW	97.34	–	–	–
[79]	CASIAv5	94.5	–	–	–
[80]	CASIAv5	–	0.12	–	–
DeepShieldFed	CelebA	99.75	0.0002	0.0047	99.53
DeepShieldFed	LFW	99.65	0.0002	0.0057	99.4

Acceptance Rate (GAR) against prior centralized face template protection techniques.

The first federated unlearning framework for FR models, called FORGET ME (F^2GU), is presented in [55]. When a client asks for data removal under the "right to be forgotten," F^2GU modifies that client's generative trajectory (the path from an initial noise/latent vector to the final generated image) toward a target endpoint built using a Mirroring-guided Trajectory Optimization strategy, which avoids the quality loss seen with naive random-trajectory redirection. The goal of unlearning is driven by a trajectory-consistency loss, and the effect on the quality of generation for retained clients is limited by a multi-client regularization loss. One important difference between unlearning and retraining-based methods is that unlearning only needs one round of communication. There are three datasets that show that the method works on both GAN (EG3D/StyleGAN2) and Diffusion Model (Stable Diffusion v1.4) architectures. Table 11 reports representative results on VGGFace2 under the GAN architecture, contrasting identity-removal effectiveness

(ID, lower is better) against generation-quality preservation (FID, SSIM).

A summary of all privacy-preserving FFR techniques is presented in Table 12.

3.2 Algorithmic innovations

This section discusses various FFR techniques in relation to algorithmic innovations. The algorithmic innovation category is selected as FFR is a swiftly advancing domain, with numerous works presenting novel algorithms to enhance model performance, robustness, and adaptability. Algorithmic advancements in FFR seek to enhance performance and convergence.

Gradient correction is a vital methodology in FL designed to improve privacy preservation while sustaining model efficacy. One notable method is Federated Gradient Correction (FedGC) [24], which tackles the problem of gradient leakage. In this study, the authors investigate the concept of gradient correction within the context of backward propagation and introduce a softmax-based regularizer to adjust the gradients of class embeddings by accurately incorporating a cross-client gradient component. The method theoretically demonstrates that the FedGC functions as a loss function analogous to the conventional softmax. Table 13 presents the comprehensive experimental results demonstrating the superiority of FedGC, which can attain comparable performance to traditional centralized approaches.

Another technique, AdaFedFR [32] employs dynamically adaptive inter-class representations to address class imbalance. Authors assert that this technique improves the generalization of face models and increases the efficiency of federated training while maintaining strict privacy preservation. They employ feature representations of public identities as learnable negative knowledge to enhance the local objective within the feature space, thereby promoting the local model to acquire robust representations and optimize personalized models for clients. Experimental results across various datasets indicate that AdaFedFR surpasses the most popular techniques (at the time of their publication) on multiple prominent FR benchmarks within fewer than three communication rounds, demonstrating its communication

efficiency and high effectiveness. Table 14 presents the accuracy of AdaFedFR computed across various benchmarks. Extensive experiments across various public benchmarks demonstrate that the technique exhibits superior performance in terms of accuracy and efficiency.

In [34], the authors introduce a FFR framework utilizing intra-subject self-supervised learning (FedFS). FedFS is a novel FL architecture designed to develop personalized FR models without requiring subject-specific data. FedFS consists of two essential components: adaptive soft label construction and intra-subject self-supervised learning, which utilize aggregated features from both local and global models to collaborate with the representations of an off-the-shelf model. Furthermore, they incorporate a regularization loss to mitigate overfitting and maintain the stability of the optimized model. FedFS experiments are performed on various datasets, showcasing superior performance relative to prior methods. Table 15 presents the performance of the method across various participation rates on the DigiFace-1M [81] and VGGFace [82] datasets.

Liu et al. [64] is another basic joint-optimization framework that both [53] and [57] are built on. The technique works best when each client's local model starts from a backbone that has already been trained by the server instead of being trained from scratch. This also improves both a global representation that is generic and a representation that is unique to each client, so clients never have to send identity proxies, which is a privacy flaw [25]. A globally shared dataset D_g keeps local training from overfitting on the few local identities; a Hard Negative (HN) sampling strategy selects only the subset of D_g whose similarity to local data exceeds the CosFace margin ($t_{HN} = m = 0.4$), cutting the amount of global data used per round by about 10x for only a 0.2% generic-accuracy cost; and a contrastive regularization term keeps the local model close to the current global model while still letting it move from its own previous round. A Decoupled Feature Customization (DFC) module maps the generic feature into a client-specific space using a learned transformation and trains it with a margin-based binary cross-entropy over each local identity (one-vs-all), instead of a multi-class softmax. The ablation shows that this binary formulation outperforms a multi-class variant that is otherwise identical. Table 16 shows the main ablation from the paper (40 clients with 100 identities each; generic evaluation on IJB-C [83] and personalized evaluation on a held-out MS-Celeb-1M [84] procedure the authors introduce), showing how much each part contributes.

pFedFace [53] claims that FedFR's [64] fully-connected DFC transformation is like a "purifying" step that can get rid of just the domain-specific data that a personalized model needs while also increasing the cost of computing locally. pFedFace makes two copies of each batch-normalization

Table 11 Unlearning effectiveness vs. generation quality (VGGFace2, GAN architecture) [55]

Method	FID (↓)	ID (↓)	SSIM (↑)
FedAvg (no unlearning)	21.17	0.9014	0.8919
Fine-Tune	20.25	0.8901	0.8143
NegGrad	45.18	0.3789	0.1268
GUIDE (centralized, adapted)	16.53	0.0683	0.6297
Random redirection	55.82	0.0382	0.3268
F ² GU (proposed)	15.83	0.0443	0.6629
F ² GU + DP	17.47	0.0563	0.6046

Table 12 Summary of privacy-preserving FFR techniques

Paper	Dataset(s)	Metrics	FL framework	Key contribution
[29]	VGGFace, VGGFace2	Privacy, Accuracy	Ensemble FL	Ensemble models for privacy protection
[70]	MNIST, CelebA	Face detection/recognition rate, DAP privacy bound	FedAvg (federated GAN generator)	Federated GAN for synthetic-data release; reduces face detection/recognition success under model inversion attack from 25.5%/2.8% to 1.2%/0.1%
[31]	CASIA-WebFace, BUPT, IJB-B, IJB-C	Privacy, Accuracy	FedAvg + Clustering	Privacy-agnostic clustering
[35, 36]	CelebA	Privacy, EER	FedAvg	Privacy-preserving sup./unsup. FFR
[42]	CASIA-WebFace, LFW, IJB-B, IJB-C	Accuracy, DP metrics	DP-FedFace	Differential privacy in FFR
[44]	CIFAR-10, CelebA, WISDM	Accuracy, Privacy	Hierarchical FL	Actor-critic client selection
[48]	CASIA-WebFace (pretrain), VGG-Face (clients), LFW/CFP-FP/ AgeDB30 (test)	Accuracy (TAR@FAR = 1e-4)	FedAvg + Random Projection	Negative sample selection + random projection for Non-IID data and privacy
[21]	CelebA, LFW	Accuracy, FAR, FRR, GAR	FedAvg + Differential Privacy	Cancelable templates (quantize + XOR key) secured via FL; first FL+BTP integration
[55]	VGGFace2, CASIA-WebFace, CelebA-HQ	FID, ID, SSIM, IS	FedAvg + Trajectory Redirection	Federated unlearning for face generation via mirror-guided trajectory redirection

layer, one is shared and one is special to the client. The rest of the backbone architecture stays the same. The general branch ϕ_g, W_g is trained with the consensus-aware loss from PrivacyFace [31]. For privacy reasons, the classifier layers are kept locally and never transmitted. Then, a Personalized Discriminative Loss (PDLoss) examines at

Table 13 Verification accuracy (%) across nine FR benchmarks under Softmax, CosFace and ArcFace losses [24]

Benchmarks	Softmax	CosFace (m = 0.35)	ArcFace (m = 0.5)
LFW	98.40	98.83	98.65
CFP-FP	90.20	88.60	87.77
AgeDB-30	85.85	90.00	89.27
CALFW	81.47	87.82	86.47
CPLFW	71.88	76.72	75.17
SLLFW	90.38	94.02	93.58
FGG2-FP	87.64	85.74	84.80
MegaFace Ch.1 Verif	67.87	72.82	71.96
MegaFace Ch.1 Ident	61.26	70.96	68.75
IJB-B Verif	69.23	80.28	75.11
IJB-B Ident	78.52	82.10	78.33
IJB-C Verif	71.33	83.40	78.13
IJB-C Ident	79.52	83.44	79.28

Table 14 Verification accuracy (%) across LFW, CFP-FP, AgeDB-30, IJB-B, and IJB-C benchmarks [32]

Benchmarks	Accuracy
LFW	98.97
CFP-FP	83.34
AgeDB-30	90.56
IJB-B TAR@FAR = 1e-4	77.06
IJB-B TAR@FAR = 1e-3	86.12
IJB-C TAR@FAR = 1e-4	80.67
IJB-C TAR@FAR = 1e-3	88.78
IJB-C (1:N) TAR@FAR = 1e-2	63.30
IJB-C (1:N) TAR@FAR = 1e-1	74.23

the cosine-similarity structure generated by the local features versus that of the generic features for the same input batch. This pushes the local model toward bigger positive-pair and smaller negative-pair similarities than the generic model gets, within a margin τ (optimal $\tau \approx 0.2$). Gradients from this loss aren't transmitted back into the shared generic parameters on purpose so that cross-client consensus doesn't become unstable. In the experiments, BUPT-Balancedface [74] is split into four racial groups: African, Asian, Caucasian, and Indian. RFW [85] is used for personalized evaluation, and IJB-B [86] and IJB-C [83] are used for generic evaluation. The data is fed into a MagFace-trained ResNet18. The results for ArcFace loss are shown in Table 17. The paper also reports a closely matching trend under CosFace loss (personalized gains of 0.95–2.21 points over PrivacyFace across the four sub-datasets).

CL-FedFR [57] introduces Curriculum Learning (CuL) to FedFR [64] framework. CL-FedFR is the first study to look into CuL for FFR framework. A data-based curriculum sorts each client's training images from easy to hard by adding up the absolute head-pose angles. The authors test dual (P2), ternary (P3), quadrant (P4) splits, as well as a ternary

Table 15 True positive rate (TPR) at three FPR thresholds, across participation rates, on DigiFace-1M and VGGFace [34]

Rate	DigiFace-1 M			VGGFace		
	TPR@FPR	TPR@FPR	TPR@FPR	TPR@FPR	TPR@FPR	TPR@FPR
	0.1	0.01	0.001	0.1	0.01	0.001
0.01	66.23	31.42	13.97	96.80	90.92	82.90
0.1	83.83	60.57	39.66	96.88	91.49	83.91
0.3	86.03	61.64	41.13	97.65	92.50	85.41
0.5	87.52	65.67	45.86	97.66	92.66	85.86
0.7	89.05	69.27	50.72	97.86	93.37	87.21

Table 16 Ablation study on FedFR's components [64]

Setup	Generic 1:1 TAR@FAR = 1e-4	Generic 1:N TPIR@FPIR = 1e-1	Personal. 1:1 TAR@FAR = 1e-5	Personal. 1:N TPIR@ FPIR = 1e-4
Pre-trained (6k IDs)	84.58	80.30	72.50	82.33
FL, FedAvg only	83.71	78.53	85.70	92.27
+ Hard Negative sampling	84.64	80.76	91.91	96.09
+ Contrastive regularization	85.17	81.25	89.57	94.60
+ DFC (full FedFR)	85.21	81.27	95.46	97.94
Centrally trained (10k IDs, upper bound)	85.99	82.14	97.39	99.40

All results in %

Table 17 Verification accuracy (%) with ArcFace loss [53]

Method	RFW	RFW	RFW	RFW	IJB-B	IJB-C
	Cauc	Ind	Asian	Afr		
Pre-trained	90.82	88.10	83.22	82.92	4.63	6.93
FedAvg	92.15	90.00	86.48	85.57	74.64	78.28
PrivacyFace [31]	92.25	90.48	86.97	85.87	76.67	80.05
FedFR [64] (re-impl.)	92.68	91.37	88.10	86.45	76.11	79.30
pFedFace w/o PDLoss	93.03	91.60	89.27	88.03	76.63	80.02
pFedFace (full)	93.35	91.97	89.47	88.37	76.68	80.03
Global Training (upper bound)	93.05	91.77	88.53	89.08	80.24	83.48

Table 18 Personalized and generic FR results (%) [57]

Method	Personal. 1:1 TAR@FAR = 1e-5	Personal. 1:N TPIR@FPIR = 1e-4	Generic 1:1 TAR@FAR = 1e-4	Generic 1:N TPIR@ FPIR = 1e-1
FedFR [64] (baseline)	95.46	97.94	85.21	81.27
CL-FedFR-P2	96.19	98.34	85.57	82.12
CL-FedFR-P3	96.15	98.51	85.56	82.00
CL-FedFR-P4	96.12	83.32	85.61	82.08
CL-FedFR-P3-G	96.20	98.52	85.48	81.77
CL-FedFR-C (client-based)	96.31	98.52	85.16	81.66
CL-FedFR-P2-C	96.05	98.37	85.31	81.64
AntiCL-FedFR-P2 (control)	87.69	92.00	85.80	82.02

split that is applied to both local and global data (P3-G). A client-based curriculum instead ranks the 40 clients by their personalized performance after 5 initial communication rounds and trains first on the top-performing “easy” clients before folding in the bottom-performing “hard”

ones (CL-FedFR-C), with a combined client+data variant (P2-C). An anti-curriculum control is also used to find out if the way of the ordering is important. Using the same MS-Celeb-1 M/IJB-C setup as in [64], all forward-ordered CL variants make personalized recognition better than vanilla

Table 19 Centralized vs. federated accuracy (%), 5-way/5-shot, ImageNet-pretrained [52]

Backbone	Centralized	Federated	Drop (pts)
ResNet	99.64	84.88	14.76
DenseNet	99.60	86.02	13.58
MobileViT	99.92	91.16	8.76
ViT-Small	83.54	60.72	22.82
CvT	97.78	90.28	7.50
CoAtNet	99.36	81.96	17.40

FedFR, with smaller but still positive gains for generic recognition. However, Table 18 shows that no single variant wins on both fronts at the same time, and the anti-curriculum control has a clear cost for personalized accuracy even though it slightly helps generic accuracy. This shows that CuL's benefit in this setting is real but design-sensitive rather than automatic.

Federated Self-Supervised Few-Shot FR [52] is an intentionally placed empirical/benchmarking study. This combines three standard parts, SimCLR self-supervised pre-training, FedAvg, and prototypical-network few-shot learning. First, SimCLR contrastive learning is used on CASIA-WebFace [76] split Dirichlet-non-IID ($\alpha = 0.1$) across 20 clients. Next, federated few-shot learning with prototypical networks is used on LFW [87] across 20 IID clients for fine-tuning. This is tested on 3-way/5-way tasks with 3/5/10-shot support sets. They looked at six different backbone—ResNet, DenseNet, MobileViT, ViT-Small, CvT, and CoAtNet— and compare them with and without ImageNet pre-training. The experimental results show that from centralized to FL training, there is a consistent 12 to 30% drop in accuracy. The size of the drop depends a lot on the architecture: traditional CNNs (ResNet, DenseNet) are the most resistant to federated constraints (about 12 to 15 points), while transformer-based and hybrid architectures (ViT-Small, CvT in particular) can lose over 30 points without strong pre-training. This difference is shown clearly in Table 19 for the 5-way, 5-shot, ImageNet-trained setting across all six backbones. There are also diminishing returns after 5 shots (10 shots give less than 2% more accuracy for twice as much work per identity labeling), and domain-specific self-supervised pre-training on face data tends to help CNN backbones more than general ImageNet pre-training.

FedFV [62] addresses the extreme FL scenario: each client possesses images of just one identity, and class embeddings are not transferable among clients. This undermines traditional softmax-based feature representation losses, as the negative loss component requires embeddings from other classes. In the absence of additional classes, the softmax loss diminishes to zero, preventing a client from calculating the necessary term to maintain the distinctiveness of its embedding. Two previous strategies tackle this issue: FedAwS, which incorporates a server-side regularization

Table 20 Verification accuracy (%) vs. centralized training and prior federated methods [62]

Method	LFW	AgeDB-30	CFP-FP
Centralized (non-federated)	86.68	69.27	77.17
FedAwS	82.47	59.53	69.56
FedUV (code length 127)	78.83	56.35	68.03
FedUV (code length 255)	79.42	57.23	68.41
FedUV (code length 511)	81.60	58.93	68.63
FedFV-100 (Proposed)	82.63	66.68	74.10

term to disperse class embeddings, and FedUV, which allocates a codeword from an error-correcting code to each client, so ensuring embeddings are inherently separated; both approaches only partially mitigate the disparity. FedFV introduces a *equivalent class embedding*: the server computes the average and re-normalizes a random pair of non-participating clients' embeddings for that round, $\tilde{w} = (w_i + w_j) / \|w_i + w_j\|$, and transmits a batch of n such vectors to each participating client, along with the global model and the client's individual embedding. The client considers these as fixed negative-class anchors for the conventional softmax loss during local training, subsequently returning only its updated embedding and model weights, while the corresponding embeddings are never transmitted back. The paper demonstrates that a client cannot retrieve the original w_i, w_j from $\tilde{w}$, as an infinite number of embedding pairs on the pertinent semi-hypersphere could have generated it; an optional variant, FedFV(k), amalgamates $k > 2$ real embeddings per equivalent vector for enhanced security at a minor accuracy expense, with $k = 3$ suggested as the optimal balance of privacy, accuracy, and computational efficiency. Trained on CASIA-WebFace [76] (1,000 identities, one client each) and assessed on LFW [87], CFP-FP [88], and AgeDB-30 [89], FedFV-100 (100 equivalent embeddings per round) surpasses both FedAwS and FedUV across all three benchmarks, reducing the disparity with centralized training; the paper attributes this to FedFV's learned class-embedding similarity distribution being significantly closer to that of the centralized model compared to FedAwS or FedUV. Table 20 presents the primary comparison of the paper.

An overview of techniques for algorithmic innovation is presented in Table 21. The main challenges associated with these techniques encompass the complexities of handling non-independent and identically distributed (non-IID) data, improving model convergence in distributed environments, and addressing class imbalance within FR datasets.

3.3 Fairness and explainability

This section offers an overview of papers that address explainability, fairness considerations, and biases in FFR models. Fairness and explainability techniques ensure

Table 21 Summary of algorithmic innovation techniques

Paper	Dataset(s)	Metrics	FL framework	Key contribution
[24]	LFW, CFP-FP, CASIA-WebFace, AgeDB-30, SLLFW, CPLFW, CALFW, MegaFace, IJB-B	Accuracy, Robustness	FedAvg + GradCorrection	Gradient correction for non-IID robustness
[32]	MS-Celeb-1 M, IJB-B, IJB-C, LFW, CFP-FP, AgeDB	Accuracy, Inter-class rep	AdaFedFR	Adaptive inter-class representation learning
[34]	Digi-Face-1 M, VGGFace	Accuracy, Self-supervision	FedAvg + SelfSup	Self-supervised learning for FFR
[52]	CASIA-WebFace, LFW	Accuracy (N-way K-shot)	FedAvg + SimCLR + ProtoNet	SSL+few-shot benchmark across 6 backbones; 12–30% federated accuracy cost
[53]	BUPT-Balancedface, RFW, IJB-B, IJB-C	Verification accuracy	PrivacyFace + Adaptive (dual) BN	Client-specific BN layers + personalized discriminative loss (PDLoss)
[57]	MS-Celeb-1 M, IJB-C	TAR / TPIR	FedFR [64] + Curriculum Learning	Head-pose- and client-performance-based curricula atop FedFR
[62]	CASIA-WebFace, LFW, AgeDB-30, CFP-FP	Verification accuracy	FedAvg + Equivalent Class Embeddings	One-class-per-client setting; unshareable embeddings replaced by server-fused equivalents
[64]	MS-Celeb-1 M, IJB-C	TAR / TPIR	FedAvg + HN sampling + Contrastive + DFC	Joint generic/personalized FL; foundational framework for [53, 57]

that the FFR systems are equitable and transparent. To reduce bias in model predictions, approaches such as adaptive FL with actor-critic selection (Privacy Safeguarding for Human-Computer Interaction) are employed. Model

Table 22 Accuracy, fairness (DDP), and training time, with and without HE [26]

Privacy (HE)	Training samples	Accuracy	Fairness (DDP)	Time (sec/epoch)
Plain (No HE)	1000	~74–80%	~8–14%	< 1
Encrypted (HE)	1000	~70–82%	~6–13%	~3000

decisions are made interpretable through the incorporation of explainable AI (XAI) methodologies, which are essential for applications in healthcare and law enforcement. Nevertheless, the heterogeneity of data among clients and the challenges associated with auditing decentralized models have hindered the attainment of fairness in FFR. A selection of key techniques that elucidate fairness and explainability are as follows.

In [26], a technique is introduced to develop system capable of ensuring trustworthiness by providing privacy, fairness, and explainability through intentional design. The technique demonstrates the potential for concurrently acquiring data while maintaining privacy. To ensure the privacy of the individual, HE has been implemented. The technique attains fairness and interpretability without compromising the accuracy of the final models. The FairFace [90] dataset is utilized to substantiate the authors' claims. Table 22 shows the experimental results reported in [26], both with and without HE, along with fairness measured as the Difference of Demographic Parity (DDP).

A Hessian-Free Model-Agnostic Meta-Learning (HF-MAML) approach is proposed in an FFR through [37]. Compared to the existing FFR models, this HF-MAML attains superior scores in verification assessments. Specifically, the verification scores exhibit the most significant enhancement in heterogeneous data partitions. An embedding regularization term is incorporated into the loss function to maintain a balance between personalization and robustness of global model. It has been demonstrated that the accuracy of global model verification is improved when this term is integrated with HF-MAML. HF-MAML and its embedding regularization extension are additionally utilized to perform a fairness analysis, which reveals that the standard deviation of client evaluation scores can be decreased to improve fairness [38]. Table 23 presents the performance of the technique across the Equal Class Partition (ECP), Lognormal Class Partition (LCP), and Attribute-Based Partition (ABP) datasets at TAR@FAR = 0.1.

Another deep learning FFR model is introduced in [43] with both private and non-private models. Training is performed using BUPT-Balancedface dataset [74]. The private version of the model is adjusted to include DP measures to safeguard data confidentiality, while preserving the focus on fairness. The accuracy, fairness, and privacy of both private

Table 23 Verification performance (untuned vs. tuned, train vs. test) across three client partitioning schemes [38]

Partition	Split	Untuned train	Untuned test	Tuned train	Tuned test
ECP	Global	0.795 ± 0.007	0.705 ± 0.016	0.855 ± 0.005	0.762 ± 0.013
ECP	Local	0.804 ± 0.007	0.737 ± 0.014	0.852 ± 0.006	0.768 ± 0.013
LCP	Global	0.722 ± 0.012	0.695 ± 0.012	0.810 ± 0.014	0.737 ± 0.013
LCP	Local	0.725 ± 0.014	0.705 ± 0.016	0.802 ± 0.014	0.731 ± 0.012
LCP	Regularization	0.734 ± 0.011	0.726 ± 0.011	0.798 ± 0.011	0.735 ± 0.013
ABP	Global	0.784 ± 0.008	0.653 ± 0.015	0.838 ± 0.008	0.755 ± 0.016
ABP	Local	0.789 ± 0.009	0.679 ± 0.011	0.816 ± 0.006	0.759 ± 0.011
ABP	Regularization	0.789 ± 0.009	0.708 ± 0.014	0.812 ± 0.006	0.729 ± 0.011

Table 24 Verification accuracy across nine benchmarks and five DP privacy budgets (ϵ), with fairness SER and STD [43]

Benchmarks	$\epsilon = 1$	$\epsilon = 4$	$\epsilon = 8$	$\epsilon = 16$	$\epsilon = 32$
LFW	62.7	75.6	91.9	94.3	98.4
CALFW	49.3	65.1	83.8	87.6	92.7
CPLFW	37.4	51.8	78.8	82.4	86.3
AgeDB	54.4	63.9	86.2	90.7	94.9
CFP-FP	45.7	62.2	83.9	87.5	91.9
RFW African	32.6	48.1	64.3	69.7	79.5
RFW Asian	25.7	36.3	54.8	59.6	70.3
RFW European	33.8	48.6	65.2	69.9	82.3
RFW Indian	27.2	39.3	57.5	68.4	78.6
Fairness SER	1.32	1.34	1.19	1.17	1.17
Fairness STD	3.44	5.38	4.41	4.25	4.47

and non-private models are evaluated and compared. The experimental results indicate that differential privacy diminishes accuracy among various demographic groups in an inconsistent manner, and that modifying the privacy budget allows for a more suitable compromise between accuracy, fairness, and privacy. Furthermore, they performed experiments to investigate real-world bias by training the private model on the imbalanced CASIA-WebFace [76] dataset. This dataset accentuates the variability in accuracy and fairness disparities, illustrating the impact of dataset composition on the relationship among privacy, accuracy, and fairness. Table 24 displays the experimental results obtained through the technique.

F3 [61] addresses Face Attribute Classification (FAC, e.g., gender classification) under federated data heterogeneity. Here, each client's local data belongs to only a *single* demographic group for some sensitive attribute (e.g., age), rather than a representative mix of groups—for instance, a client might hold only “ ≤ 30 ” samples or only “ > 30 ” samples, never both. The paper proves, in its Proposition 1,

Table 25 Accuracy (%) and fairness-violation (Δ , lower is better) under FL with Data Heterogeneity [61]

Dataset	Method	Accuracy	Δ_{EOPP}	Δ_{EO}	Δ_{AP}
FairFace	FedAvg-DH (baseline)	79.7	1.0	16.3	17.3
	FairBest	72.0	0.6	8.2	8.8
	10-FairAvg	74.2	3.3	10.2	13.5
	10-FairAccAvg	73.3	0.8	9.6	10.4
FFHQ	FedAvg-DH (baseline)	90.8	2.4	10.5	13.0
	FairBest	74.7	1.8	1.9	3.2
	10-FairAvg	82.2	2.7	2.7	4.3
	10-FairAccAvg	82.7	1.8	2.0	3.9
UTK	FedAvg-DH (baseline)	94.1	1.5	11.9	13.5
	FairBest	83.9	1.9	6.0	10.0
	10-FairAvg	93.5	0.9	6.1	7.8
	10-FairAccAvg	93.7	3.1	5.3	6.8

that this setting breaks existing fair-FL methods such as the Lagrangian Multiplier Method. A client's local fairness-violation term requires samples from every demographic group to compute. Under this heterogeneity, that term diverges to infinity, so fairness cannot be enforced during local training at all. F3's response is to move fairness enforcement to the *aggregation* step instead. Clients train purely for accuracy. The server then evaluates each client's submitted model against a small held-out “aggregator set” containing samples from every demographic group, and aggregates using one of three heuristics: *FairBest* keeps only the single most-fair client model; α -*FairAvg* averages the top- $\alpha\%$ most-fair models; α -*FairAccAvg* averages the top- $\alpha\%$ models by accuracy-to-fairness-violation ratio. As α grows, both averaging heuristics converge toward plain FedAvg, giving a tunable accuracy/fairness dial. The authors evaluate F3 with age as the sensitive attribute and gender as the classification target, on FairFace, FFHQ, and UTK, using 50 clients and a ResNet-18 backbone. The baseline is FedAvg under the same heterogeneous split, referred to as “FedAvg-DH.” F3's heuristics improve fairness—measured as lower Δ_{EOPP} , Δ_{EO} , and Δ_{AP} violation—by 25–82% over FedAvg-DH. This comes at an accuracy cost of 0.4–17%, depending on dataset and heuristic (Table 25).

In [68], another FFR technique is introduced. The paper states that under single-category-per-client heterogeneity, the Lagrangian Multiplier Method's (LMM) fairness-violation term is uncomputable at the client level, since the false-positive/false-negative rates for categories absent from a client's data are undefined. The paper introduces four aggregation heuristics that transfer fairness enforcement to the server, assessed using a limited validation set retained solely by the aggregator: *FairBest* identifies the singular most-fair client model; *FairAvg* computes the average

of the top- κ most-fair models; FairAccRatio chooses the model with the optimal accuracy-to-fairness-violation ratio; and FairAccDiff selects the model that maximizes a weighted disparity between accuracy and fairness violation. The experiments encompass CelebA [73] (forecasting attractiveness and gender as the sensitive attribute), UTK, and FairFace [90] (both forecasting gender, with ethnicity/race as the sensitive attribute), involving 50 clients utilizing a ResNet-18 architecture. Table 26 presents the UTK findings, which are the most transparent of the three: FairAccRatio achieves an accuracy of 80.8%, just 4.5 points shy of FedAvg-DH (the simplistic accuracy-only heuristic under the same heterogeneous distribution), while reducing the Δ_{EOpp} fairness infringement by nearly 80% in comparison to that benchmark ($23.055 \rightarrow 4.828$) and it approaches FedAvg- λ 's accuracy while substantially exceeding it on fairness (4.828 vs. 17.453).

Table 27 presents a summary of the methodologies within the category of fairness and explainability. The primary challenges associated with this technique in this category encompass, but are not limited to, ensuring fairness among various demographic groups. Ensuring an appropriate balance between fairness and model performance, while also maintaining the transparency and interpretability of FFR systems.

3.4 Domain adaptation and transfer learning

FFR models are essential for generalizing across diverse domains, including different lighting environments and demographic groups. In this section, techniques focus on domain adaptation and transfer learning as strategies to improve model performance across diverse environments. Domain adaptation and transfer learning methodologies enhance the effectiveness of FFR models in novel environments.

An effective framework for executing FL on AIoT devices is proposed in [30]. The method is especially appropriate for AIoT devices owing to its significant decrease in both computational requirements and communication rounds. The authors demonstrate that the method experiences a negligible reduction in performance relative to its fully centralized counterpart, as evidenced by experiments on real-world datasets. Table 28 illustrates the verification rates across the pre-trained model (without fine-tuning), SOLO (fine-tuned with local data), and FL scenarios.

Another technique in [27] allows models to adapt to new domains without the need for labeled data. Ensemble models improve transferability through the integration of multiple models. In practical applications involving heterogeneous data distributions, these methods are essential for the implementation of FFR systems. A new federated

Table 26 Accuracy (%) and fairness violation on UTK under federated data heterogeneity [68]

Method	Accuracy	Δ_{EOpp}	Δ_{EO}	Δ_{AP}
NonFed (centralized, no FL)	87.9	18.645	31.910	13.265
FedAvg (balanced data, no fairness)	86.8	25.164	40.112	14.948
FedAvg- λ (balanced data, LMM fairness)	84.6	17.453	24.171	6.718
FedAvg-DH (heterogeneous data, no fairness)	85.3	23.055	32.906	9.851
FairBest	71.6	2.930	3.185	0.255
FairAvg ($\kappa = 3$)	80.6	4.976	9.847	4.976
FairAccRatio	80.8	4.828	5.192	0.364
FairAccDiff ($\gamma = 2$)	79.6	7.801	14.053	7.801

Table 27 Summary of techniques favoring fairness and explainability techniques

Paper	Dataset(s)	Metrics	FL framework	Key contribution
[26]	FairFace	Accuracy, Fairness, Privacy	Custom FL	Combines privacy, fairness, explainability
[37, 38]	CelebA	Accuracy, Imbalance	Meta-Learning FL	Meta-learning for imbalanced data
[43]	RFW, BUPT-Balancedface	Accuracy, Privacy, Fairness	Custom FL	Privacy/accuracy/fairness trade-offs
[61]	FairFace, FFHQ, UTK	Accuracy, Δ_{EOpp} , Δ_{EO} , Δ_{AP}	FedAvg + fairness-aware aggregation heuristics	Fair face attribute classification when clients are demographically homogeneous individually (FL with Data Heterogeneity)
[68]	CelebA, UTK, FairFace	Accuracy, Δ_{EOpp} , Δ_{EO} , Δ_{AP}	FedAvg + fairness-aware aggregation heuristics (FairBest/Avg/AccRatio/AccDiff)	Precursor to [61]; validation-set-driven heuristics for single-category-per-client heterogeneity

Table 28 TAR at two FAR thresholds, comparing pre-trained, locally fine-tuned (SOLO), and federated models [30]

Method	TAR@FAR = 1e-3	TAR@FAR = 1e-4
Pre-trained	64.32%	59.68%
SOLO (local fine-tune)	67.15% $\pm$ 0.19%	61.34% $\pm$ 0.11%
FL	72.83% $\pm$ 0.22%	63.51% $\pm$ 0.25%

unsupervised domain adaptation method for FR, is presented here. The technique initially enhances a hierarchical clustering algorithm to generate pseudo-labels for the target domain by leveraging the model pre-trained in the source

Table 29 Verification (TPR at three FPR thresholds) and rank-1 identification accuracy, by racial groups [27]

Metric	African	Caucasian
Verif. TPR@FPR = 0.1	91.57	96.00
Verif. TPR@FPR = 0.01	76.97	85.60
Verif. TPR@FPR = 0.001	49.40	70.93
Identification Rank-1	93.26	93.59

Table 30 Accuracy, F1, precision, recall, and AUC, comparing DP against FL [39]

Metric	DP	FL
Accuracy	84%	92%
F1 Score	85%	93%
Precision	87%	95%
Recall	83%	91%
ROC-AUC	90%	98%

Table 31 IJB-C 1:1 verification, TAR (%) at various FARs and AUC [47]

Method	TAR@1e-6	TAR@1e-5	TAR@1e-4	AUC
Pre-trained	21.80	36.52	55.90	98.417
Central (upper bound)	27.60	44.25	62.29	98.579
Local (no FL)	21.70	37.39	55.90	98.610
FD	13.20	23.25	38.25	97.573
LG-FedAvg	22.56	37.88	56.40	98.630
FML	14.27	24.48	39.78	96.508
FedProto	22.12	37.52	55.99	98.616
FedTGP	21.92	37.27	55.92	98.615
FedMRL	14.27	24.48	39.78	96.508
FedPFR (Proposed)	23.39	38.79	57.14	98.634

domain. Then, the framework conducts FL across domains by employing labeled data from the source domain and unlabeled data with pseudo labels from the target domain. The efficacy and importance are evaluated through comprehensive experiments conducted on various benchmarks. Table 29 presents the performance metrics for face verification across various false positive rates and for identification at rank 1.

In [39], a novel methodology for integrating FFR with Electrostatic Discharge (ESD) testers is presented. This integration utilizes the FL model with the preFedAvg algorithm, ensuring data privacy and security through a distributed learning approach that maintains the high accuracy and efficiency of the ESD testing systems. The performance of the proposed approach is evaluated using metrics such as accuracy, precision-recall, and F1 score and Area Under the ROC Curve (AUC). The results demonstrate that this approach surpasses the traditional method while concurrently protecting data against various vulnerabilities. The electronic manufacturing sectors, quality assurance divisions, and safety oversight units are integral components of

Table 32 Facial expression recognition comparison (%) [51]; not a face recognition result

Method	Accuracy	Precision	Recall	F1
PAFedFV	93.06	93.07	93.01	93.03
RUCR	92.75	92.72	92.76	92.71
GA-SVM	95.91	96.08	95.92	95.98
3D-CNN	96.17	96.23	96.19	96.20
DCAE-ViT-SDeceFL	95.42	95.69	95.42*	95.51
FedDAT	94.50	94.65	94.62	94.62
ResNet50	96.42	95.88	95.49	96.64
Inceptionv3	97.04	97.09	97.06	97.07
Proposed (GCRA)	98.32	98.33	98.40	98.37

this strategy. Table 30 presents the comparison of different performance metrics with Differential Privacy and FL.

FedPFR [47] addresses FFR when clients use different model architectures and hold entirely different sets of identities. Conventional FL and previous prototype-sharing techniques are inadequate in this scenario, as the aggregation of identity prototypes from separate clients undermines their significance. Rather, each client independently uploads its identity prototypes without averaging; the server merely aggregates them and redistributes the collection to all clients. Local training integrates a conventional classification loss with an innovative prototype-anchor contrastive loss that attracts embeddings towards their respective prototypes while repelling them from others. Across four distinct backbone architectures, FedPFR surpasses previous heterogeneous FL techniques on IJB-C verification at all thresholds. Table 31 illustrates the primary comparison; FedPFR surpasses FedAvg-style and earlier prototype-based (FedProto, FedTGP) heterogeneous-FL benchmarks at each FAR threshold, while also attaining the greatest AUC among all evaluated FL approaches.

Karra and Kakhandki [51] introduces an adaptive FL methodology that integrates optimization strategy, the Greater Cane Rat Algorithm (GCRA), for determining federated aggregation weights, subsequently employing a fused extreme learning machine classifier. Although the title references "face recognition," the paper's true focus, datasets, and assessment pertain solely to facial expression/emotion recognition (happy/sad/angry/surprised/neutral on SFEW, NVIE, ExpW, and EmotioNet), rather than face identity verification or identification, which is the focus of this survey; we include it for thoroughness with this caveat noted instead of presenting it as an identity-FFR outcome. The proposed technique surpasses previous methods in several expression-recognition evaluations. Table 32 shows accuracy, precision, recall, and F1 score for the proposed method against eight prior expression-recognition and FL techniques

Nusa et al. [45] compares FFR on a small non-IID LFW subset. The research assesses multiple pre-trained

architectures in conjunction with a bespoke CNN-Transformer hybrid (TFR), utilizing either FedAvg or FedProx aggregation, and also examines DP across varying noise intensities. TFR combined with FedProx consistently yields superior results in all scenarios, but a standard CNN lacking pretraining struggles to acquire knowledge. The authors further demonstrate that FL accommodates greater privacy noise when expanded to a larger number of clients with a reduced sample size per iteration. Table 33 reports accuracy and average running time per round for five model architectures under both FedAvg and FedProx aggregation.

Table 34 presents a summary of these techniques. The principal challenges associated with these techniques encompass adapting models to new domains while preserving privacy, ensuring robustness in cross-domain FR tasks, and effectively managing domain shifts within distributed environments.

3.5 Efficiency and scalability

The computational and communication challenges associated with FFR are explained in efficiency and scalability section. In environments with limited resources, such as edge devices and IoT systems, these techniques are crucial for the development of real-time FR applications.

Kim et al. [23] is a short position paper that proposes strategies for FFR model training, and claims to be the first publication to do so. It employs the ArcFace loss within MobileFaceNets and recognizes what it terms the “continuous” characteristic of FR: a singular identity produces novel facial data as time progresses, while new identities are introduced over time, resulting in an ever-increasing number of tasks. Four approaches tackle this issue. A labeling approach deduces positive labels en masse when the majority of frames within a window are positive. A data management approach mitigates catastrophic forgetting within mobile memory constraints by preserving solely the subset of embedding vectors that optimizes inter-vector distance, instead than maintaining all previous images. An assessment methodology divides devices into training and evaluation functions at a predetermined ratio (e.g., 80:20), enabling the validation of an updated global model on-device while maintaining the confidentiality of private data, with the option to revert to the most effective previous model if updates are rejected over consecutive iterations. An update buffer technique permits devices to train whenever they are accessible and stores their weight modifications, facilitating asynchronous involvement rather than server-directed device selection for each round. The document presents no experiments, datasets, or quantitative findings; it is provided as an initial conceptual contribution rather than as an empirical outcome.

Table 33 Accuracy and Average Running Time (ART, seconds/round) by architecture and aggregation method [45]; 5 clients, 5 communication rounds

Architecture	FedAvg Acc.	FedAvg ART	FedProx Acc.	Fed- Prox ART
CNN (no pretraining)	0.000	24	0.000	21
ResNet50	0.147	388	0.000	267
EfficientNet	0.216	301	0.311	310
MobileNetV2	0.143	256	0.137	221
TFR (Proposed)	0.392	24	0.437	31

Table 34 Summary of domain adaptation and transfer learning techniques

Paper	Dataset(s)	Metrics	FL framework	Key contribution
[30]	MS-Celeb-1 M, Private	Accuracy, Latency	Industrial FL	FL for AIoT/ industrial FR
[27, 28]	MS-Celeb-1 M (constructed benchmark), RFW	Verification, Rank-1 Id	FedAvg + clustering (C-FINCH) + DCL	Unsupervised domain adaptation via pseudo-labels + domain constraint loss
[39]	Custom	Accuracy, F1, Privacy	Custom FL	FL + ESD testing integration
[47]	MS-Celeb-1 M subset, IJB-C	TAR@ FAR, AUC	Prototype-based (no averaging) + PAC loss	Open-set + model-heterogeneous FFR via prototype pools
[51]	SFEW, NVIE, ExpW, EmotioNet	Accuracy, Precision, Recall, F1	GCRA-weighted FL + fused ELM	Metaheuristic-optimized FL aggregation; see scope caveat below
[45]	LFW (5 identities)	Accuracy, ART	FedAvg / FedProx	Architecture (incl. CNN+Transformer hybrid) and DP comparison under non-IID splits

Table 35 Verification accuracy on LFW, IJB-A, and IJB-C, comparing a centralized fine-tuned baseline, random initialization, and FedFace [25]

Method	Training data	LFW acc.	IJB-A@0.1	IJB-C@0.1
Fine-tuning baseline	Centralized	99.32%	84.18%	88.76%
Random Init	Distributed	94.61%	70.13%	69.30%
FedFace	Distributed	99.28%	83.79%	88.21%

Aggarwal et al. [25] introduces an FFR framework referred to as FedFace. FedFace utilizes face images collected from multiple clients to develop a robust and accurate FR model. The facial images stored on each client are not disclosed to other clients or the central server, and each client is a mobile device that contains the facial images of its

owner (one identity per client). The experiments substantiate the effectiveness of FedFace in enhancing the verification performance of pre-trained FR systems across a range of face verification benchmarks. Table 35 illustrates the FR accuracy on different FR benchmarks such as LFW [87], IJB-A [86] and IJB-C [83].

Federated Active Authentication (FAA) [69] tackles the identical one-identity-per-device implementation context as FedFace [25], although focuses on continuous or active authentication instead of singular verification, demonstrating that standard FedAvg is inadequately equipped for this. The technique initially measures this: inside the FAA framework, each device retains information from precisely one user (the paper's specific $q_{IID} = 0$, the most extreme non-IID scenario), and the detection accuracy of FedAvg deteriorates to random-guess levels when data deviates from IID. To tackle this issue, the suggested approach pre-trains an FR model using a public dataset, divides it into a feature extractor and a classifier, and transmits solely the static feature extractor to devices. Each device subsequently transmits only the average and covariance of its respective user's characteristics ("user impressions"), avoiding the transmission of raw images or complete model updates; the server generates synthetic features from these individual user Gaussians to train a multi-class classifier, which is then returned to the devices along with the feature extractor. This necessitates merely one instance of client-server interaction, in contrast to Split Learning, which demands several. Table 36 presents a comparative analysis of headlines across three active-authentication benchmarks (MOBIO, UMDAA-01, UMDAA-02): FedAvg exhibits the poorest performance by a significant margin (with an average detection accuracy as low as 51%, only slightly above random chance), Split Learning shows improvement yet remains inferior, while the suggested method surpasses both, along with nine previous one-class classification benchmarks from the active-authentication field.

In [33], a decentralized framework for the design and development of FFR solutions is presented. In this technique, the devices employed for image capture lack the capacity to train models capable of attaining high accuracy on extensive datasets. Consequently, a method is proposed that conducts the local detection of human faces and the extraction of features utilizing a pre-trained FaceNet model. These attributes are conveyed to a resilient processing unit, where a comprehensive model is constructed and subsequently relayed back to the clients for recognition. The experiment's findings demonstrate that the solution is both highly precise and efficient. Table 37 depicts the FR accuracy of LFW [87] dataset with different algorithmic setups.

An alternative approach advocates for the utilization of FL as a means to ensure the confidentiality of facial image

Table 36 Average detection accuracy (%) comparing FedAvg, Split Learning, and the FAA method [69]

Method	MOBIO	UMDAA-01	UMDAA-02
FedAvg	61.2	52.4	51.0
Split Learning (SLA)	92.0	89.0	62.0
FAA (Proposed)	99.8	95.4	81.2

Table 37 Recognition accuracy comparing a plain CNN against FaceNet+SVM, at two identity-pool sizes [33]

Algorithm	Accuracy (%)	No. Identities
CNN (500 epochs)	66	50
FaceNet+SVM	90	50
CNN (500 epochs)	75	30
FaceNet+SVM	96	30

Table 38 Equal error rate for individual and federated models, with and without a secure aggregation and GAN augmentation [41]

Models	With secure aggregator	Without secure aggregator
Individual Model	2.22	—
FL without GAN	2.72	1.66
FL with GAN	2.84	2.09

Table 39 Global (clean) and robust accuracy (%) at ABR = 50% under three adversarial attacks [65]

n_a	Global Acc.	FGSM	FFGSM	Square
1	85.7	54.1	54.2	60.5
2	83.1	66.7	67.9	71.0
3	73.9	71.9	72.3	72.5
4	30.7	74.1	75.0	66.6

data on edge devices, especially within recognition systems [41]. The approach emphasizes decentralized training significantly, thus removing the requirement to transmit raw image data to centralized servers. A secure aggregator aggregates the local models into a comprehensive federated model, which is then transmitted to individual devices through the primary server. The approach presented in this study provides two principal benefits. Initially, it ensures the confidentiality of facial images by securely storing the original data on individual devices. Secondly, empirical data demonstrates that the federated model exceeds the performance of individual models in terms of average Equal Error Rate (EER) when a secure aggregator is absent as shown in Table 38.

FLATS [65] discusses federated adversarial training for smart-home FR under a "starving" federated setting, where each client holds very minimal data. A randomly selected group of clients engages in adversarial training during each iteration, governed by two hyper-parameters: n_a (the number of clients undergoing adversarial training) and ABR (the proportion of their local batch that is adversarial). Benign federated training by itself renders the global model nearly

unprotected, achieving a robust accuracy of less than 7% against FGSM. Table 39 illustrates that augmenting n_a significantly enhances robustness, albeit at the expense of clean accuracy, with $n_a = 3$ achieving the optimal equilibrium (73.9% clean accuracy paired with 70–72% robust accuracy across three assaults); increasing n_a to 4 results in a drastic decline in clean accuracy to 30.7% despite attaining the highest robust scores. The analysis also validates that these resilience improvements last in non-IID scenarios (alterations in pixel and brightness, occlusion of the ocular region).

Bai et al.'s FedFace [67] is a separate publication from the similarly titled FedFace [25]. This introduces Partially Federated Momentum (PFM), which locally estimates a global momentum term to rectify client drift without incurring the expense of computing gradients twice per iteration, and Federated Validation (FV), which seeks improved aggregation weights by evaluating candidate models against private validation datasets maintained by distinct validator entities. In both 3-party and 12-party configurations on MS1M, Asian-DeepGlint, and VGGFace2, PFM consistently decreases training loss relative to FedAvg, often falling within the centralized benchmark. Table 40 reports final training loss for the IR-18/ArcFace setting; the paper's main verification benchmarks (LFW, IJB-B/C, MegaFace) are reported only as relative bar charts against a centralized baseline rather than as a table, so are not reproduced numerically here.

Vyas and Torrens [46] proposes an edge-based FL architecture for facial and object recognition in IoT video surveillance, aiming to avoid transmitting any raw images or video to a central server. The solution integrates differential privacy (noise incorporated into model updates) with homomorphic encryption, enabling the aggregator to process encrypted updates without decryption; only alerts and analytics, rather than unprocessed film, would be transmitted from the edge device. The document is clearly identified as a work in progress: it outlines the architecture and a test-bed strategy (NVIDIA Jetson) but does not provide a trained model, dataset, or accuracy metrics. This publication does not have a table, as it presents no empirical findings to disclose.

Solomon [54] discusses unsupervised FL for face verification on CelebA, using GAN-generated impostor images so edge devices need not hold genuine impostor samples, and evaluates FL with and without a secure aggregator via Equal EER. Table 41 shows FL without the aggregator improving on both individual per-device models (2.33% vs. 2.55% EER) and FL with the aggregator (3.05%), while a centralized baseline (all identities pooled) reaches 2.70%, illustrating a privacy/performance trade-off; both differences are reported as statistically significant ($p < 0.05$).

Table 40 Final training loss (IR-18 backbone, ArcFace loss, 3-party setting) [67]

Party	Solo	Centralized	FedAvg	PFM
MS1M	4.21	4.63	5.44	4.89
Asian-DeepGlint	5.78	4.61	5.17	3.88
VGGFace2	6.11	5.70	6.15	5.25

Table 41 Equal error rate (%) of individual, federated (with and without secure aggregator) and centralized settings [54]

Setting	EER (%)
Individual Model	2.55
Federated (with Secure Aggregator)	3.05
Federated (without Secure Aggregator)	2.33
Centralized (all data pooled, single model)	2.70

The summary of these techniques is provided in Table 42. The primary challenges associated with these techniques encompass minimizing communication overhead in FL, scaling FFR systems to accommodate extensive datasets, and ensuring real-time performance in industrial applications, among other issues.

3.6 Security: attacks on the training process and recognition target

FFR systems are particularly prone to two groups of attacks: the first group targets the training process: A malicious or Sybil client submit poisoned model updates to corrupt the global model. Whereas the second group targets the recognition target itself: presentation attacks, in which an attacker impersonates a live face using a printed image, recorded video, or a 3D/silicone mask, and synthetic-media forgery, where deepfake technology creates an artificial face that has never existed. This section explains several existing works in both this groups.

3.6.1 Process-level security: robustness against malicious clients

Although secure aggregation techniques such as HE and secure MPC enhance certain security and privacy metrics, they also introduce significant threats, including poisoning attacks, backdoor attacks, and free-riding attacks. In [40], the authors propose a novel defense strategy against poisoning attacks in FL, known as Sybil-aware Federated Learning (SaFL), which reduces the influence of Sybils through an innovative time-variant aggregation method. In the experiment, they evaluated the method in comparison to a baseline without protection and two additional commonly employed mitigation strategies (FoolsGold and Multi-Krum) across three cosine-distance thresholds ($\nu \in \{0.6, 0.8, \text{Decay}\}$). Table 43 summarizes the reported attack success rate and training-loss behavior across these methods and settings.

Table 42 Summary of techniques highlighting efficiency and scalability

Paper	Dataset(s)	Metrics	FL framework	Key contribution
[23]	None (no experiments)	N/A	ArcFace + buffered asynchronous aggregation	Early strategy paper for FL-based FR; labeling, memory-bounded data retention, on-device evaluation, and asynchronous update buffering
[25]	LFW, IJB-B, IJB-C	Accuracy, Scalability	FedFace	Collaborative FR, improved scalability
[69]	MOBIO, UMDAA-01, UMDAA-02	Average detection accuracy	Feature-statistic sharing (no FedAvg)	Federated Active Authentication; shares per-user feature mean/covariance instead of raw features or full model updates
[33]	LFW	Accuracy, Decentralization	Decentralized FL	Decentralized FL, reduced central reliance
[41]	LFW, IJB-C	Accuracy, Security	Supervised FL	Security/efficiency via supervised FL
[54]	CelebA (2,000 subjects)	EER	Unsupervised FL + secure aggregator (SMC)	GAN-generated impostors; privacy/performance trade-off of secure aggregation
[65]	Kaggle celebrity faces (105 classes)	Global/Robust accuracy	FedAvg + adversarial training	Federated adversarial training for smart-home robustness under starving data
[67]	MS1M, Asian-DeepGlint, VGGFace2	Training loss, relative TAR/accuracy	PFM + Federated Validation	Momentum-based drift correction + validation-searched aggregation weights
[46]	None (no experiments)	N/A	FedAvg/FedOpt (proposed)	Edge FL architecture proposal for leak-free facial/object recognition; work in progress

Table 43 Attack success rate and training-loss behavior presented under Sybil poisoning attack, across baseline, prior defenses, and SaFL [40]

Method	Setting	Attack success rate	Training loss behavior
Baseline	No Sybils	–	≥ 0.8 , fluctuates ~ 1.0 for ~ 200 iterations
	Single-target, 3 Sybils	~ 0.9	> 2 throughout, across all tested settings
Multi-Krum	Single-target, 4 Sybils	1.0	
	Multi-target, 3 Sybils	~ 0.6	
FoolsGold	Single-target, 3–4 Sybils	0.0	≈ 3.8 (worst case, 4 Sybils)
	Multi-target, 4 Sybils	–	≈ 6 (worst case, underfitting)
SaFL ($\nu = 0.6$)	Single-target, 3 Sybils	0.3 (worst case for SaFL)	–
SaFL ($\nu = 0.6$)	Multi-target, 3 Sybils	~ 0.3	–
SaFL ($\nu = 0.8$)	Single-target	–	Close to Baseline
SaFL ($\nu =$ Decay)	Multi-target	~ 0.0 (comparable to FoolsGold)	Best approximation of Baseline among all compared methods

The experimental results indicate that the method achieves a high level of protection with minimal impact on performance, thereby strengthening FL defenses against targeted poisoning attacks.

3.6.2 Target-level security: presentation attack detection and forgery detection

Fed-StackFPAD [50] operates with a self-supervised ViT-MAE encoder. Fed-StackFPAD is optimized by FedAvg utilizing four dataset clients. Each client also develops its own localized (non-federated) variant of the same encoder. A meta-classifier employing SVM then consolidates the probabilities from both the global and local models on a validation dataset located on the server side. This stacking method reduces the global model's tendency to overfit the primary client. Table 44 demonstrates that HTER diminishes by roughly 14 points in comparison to the most resilient prior FL baseline (FedGPAD) over all four leave-one-domain-out segments.

FedSIS [60] develops a hybrid Vision Transformer (ViT) that incorporates a CNN tokenizer with a transformer encoder. The tokenizer and classifier are tailored for each client and are aggregated by FedAvg, whereas the common encoder is trained using split learning. This approach synchronizes the decision boundary between genuine and malicious instances among clients. Additionally, a block-sampler randomly chooses one of the 12 intermediate ViT

Table 44 Cross-domain HTER (%) on the O/C/I/M leave-one-domain-out benchmark, vs. prior federated baselines [50]

Method	O&C&I→M	O&M&I→C	O&C&M→I	I&C&M→O
Fed-PAD (prior FL)	19.45	42.27	32.53	34.44
FedG-PAD (prior FL)	12.73	28.69	10.97	21.95
Fed-Stack-FPAD (proposed)	4.07	7.91	2.92	4.50

O = OULU-NPU, **C** = CASIA-FASD, **I** = Replay-Attack (Idiap), **M** = MSU-MFSD. Column headers denote leave-one-domain-out splits, e.g., O&C&I→M: train on O, C, I; test on unseen domain M

Table 45 Average cross-domain HTER (%) on Benchmark 1 (MCIO) vs. FedAvg, split-learning and centralized upper bound [60]

Method	Avg. HTER (%)
FedAvg (plain FL baseline)	10.15
FeSTA (split learning, no sampling)	10.64
Centralized ViT-H-IS	5.56
FedSIS (Proposed)	3.91

MCIO = MSU-MFSD, CASIA-MFSD, Idiap Replay-Attack, OULU-NPU (same four datasets as Table 44's O/C/I/M, averaged over all four leave-one-domain-out splits)

Table 46 Test accuracy, precision, and F1 on the merged FaceForensics++/Deeperforensics-1.0/WildDeepfake dataset, single-backbone vs dual-backbone fusion [49]

Method	Testing Acc.	Precision	F1
Xception (single backbone)	0.9125	0.9101	0.9101
MobileNetV2 (single backbone)	0.8856	0.8855	0.8855
FFDL (Proposed, fused)	0.98995	0.9863	0.98963

blocks in each iteration, mitigating non-IID data concerns and enhancing characteristics for new domains. Table 45 illustrates that FedSIS achieves the minimal average HTER, surpassing its centrally-trained equivalent.

FFDL [49] employs two pre-trained models for each client, Xception and Inception, to derive facial features. Xception incorporates depthwise separable convolutions to capture complex textural details. Inception use concurrent filters of differing sizes to identify patterns at several scales. These two feature sets are interconnected and unified by Canonical Correlation Analysis (CCA), which determines the most correlated components between them. The consolidated vector is then classified utilizing an optimized CNN head. Clients perform training within their respective local settings, but FedAvg integrates their models across several iterations. Table 46 compares the FFDL with single-backbone federated benchmarks in the integrated FaceForensics++/Deeperforensics-1.0/WildDeepfake dataset. The

Table 47 Hybrid-domain forgery detection: accuracy, and AUC (%) vs. the best non-federated baseline and a centralized ablation [59]

Method	Accuracy (%)	AUC (%)
DCL (best non-private baseline)	86.82	94.48
FedForgery (centralized ablation)	87.36	93.23
FedForgery (federated, proposed)	85.55	91.12

Table 48 AUC (%) on FaceForensics++ under varying client counts [63]

Setting	Deepfakes	Face2Face	FaceSwap	NeuralTextures
Centralized, c0	99.40	99.94	99.94	98.42
Federated, K = 2, c0	99.99	99.92	99.93	98.78
Federated, K = 5, c40	91.21	94.28	77.27	73.02

integrated two-backbone model attains enhanced testing accuracy and F1 score, exceeding the performance of any separately trained backbone under same circumstances.

FedForgery [59] trains a network comparable to a variational autoencoder (VAE) at each client. This network reconstructs the genuine facial image. The model then computes the residual: the difference between the input and its reconstruction. Classification is performed on this residual instead of the raw pixels. The concept suggests that leftover artifacts demonstrate enhanced generalization in hybrid or unfamiliar forging categories relative to unprocessed pixel patterns. Clients perform training within their respective local settings, whereas FedAvg aggregates their models from a collective of 10 clients. Table 47 displays a unique result. The federated approach falls short of the ideal non-private benchmark regarding AUC. It additionally adheres to its own centralized ablation. The authors distinctly express this inequality, highlighting the costs linked to preserving anonymity.

Dynamic-Aware FL [63] is claimed to be the initial FL framework specifically developed for identifying face forgeries in videos. It utilizes traditional FedAvg and main contribution is an Inconsistency-Capture Module (ICM). The ICM consists of two sectors: Global and Local. Both assess a target frame in comparison to a reference frame, sampled at 5/15 frames separated. The Global branch evaluates structural (dis)similarity throughout the complete face. The Local branch focuses solely on critical regions: the eyes, nose, and mouth, acquired by Region of Interest (RoI) Pooling. The fundamental idea is that altered videos display inconsistencies between frames, even when each separate frame is genuine. Table 48 illustrates outcomes similar to centralized training when the client count is restricted (K = 2). Nonetheless, performance declines markedly with an increase in the number of clients. This occurs as a result of

the reduction of per-client data with an increase in the client base.

FedPR [58] divides the forgery detector of every client into two distinct elements. First component functions as a shared feature extractor and is aggregated across all clients. The final component is a customized feature extractor and is kept confidential for every individual client. To force the customized branch to specialize, the method interpolates feature statistics (mean and variance) between genuine and synthetic samples, employing an AdaIN-style approach. An adversarial loss is also included that drives the collective sector toward a consistent representation that seems uniform across clients, regardless of their regional counterfeit discrepancies. The method is evaluated with a specially constructed hybrid dataset. It incorporates FF++, WildDeepfake, CelebDF-v2, and FMFCC-V cross its clients. Table 49 demonstrates that the federated model either equals or exceeds the performance of its centralized (non-FL) equivalent in three out of four sub-datasets.

FedPAD [71] describes the first publication in face presentation-attack detection (fPAD) as an FL challenge. This paper identifies two roles: data centers, which possess both authentic and counterfeit facial data and can train local fPAD models, and users, who solely have genuine images and lack the ability to train a model. Every participant locally trains a ResNet-18 fPAD classifier with cross-entropy loss. A server aggregates its parameters in a FedAvg over connection cycles, without ever accessing local image data. Users solely obtain the integrated global model during inference. The research encompasses seven publicly accessible datasets: Oulu-NPU, CASIA-MFSD, Idiap Replay-Attack, MSU-MFSD, SiW, 3DMAD, and HKBUMARsV2, featuring print, video-replay, and 3D-mask attacks. The approach utilized is leave-one-dataset-out: each dataset successively represents the unobserved user distribution. Table 50 outlines the principal comparison: Single (employing a singular data center), Fused (calculating averages from separate models), FedPAD, and All, which represents a non-private upper bound that directly consolidates the raw data from all data centers. FedPAD exceeds all privacy-preserving standards across the three criteria and nears the non-private maximum average AUC by 3.9 points. The research also indicates that a rise in data centers does not uniformly yield advantages: moving from 2 to 3 centers slightly lowers HTER before further expansions improve it again.

Federated Test-Time Adaptive fPAD [66] is an extension FedPAD [71]. It incorporates an extra phase for safeguarding privacy during testing. FedPAD independently maintains in demonstrating restricted generalization to attacks originating from domains that significantly differ from any training data center. This technique presents an extra level of testing: after acquiring the globally trained federated

Table 49 Per-subset accuracy (%) on the Forgery Source Hybrid Dataset, federated vs. centralized vs. best non-private baseline [58]

Method	CelebDF-v2	FF++ DF	FMFCC-V	Wild- Deepfake
RFM (best non-private baseline)	86.44	94.77	99.22	65.92
FedPR (w/o FL, centralized ablation)	84.29	94.12	99.36	70.05
FedPR (Proposed, federated)	89.90	94.81	99.60	70.81

Table 50 Comparison of models trained by data from a single data center vs. multiple, averaged across the four leave-one-out user datasets [71]

Method	Avg. HTER (%)	Avg. EER (%)	Avg. AUC (%)
Single	36.43	34.31	70.36
Fused	35.75	31.29	73.89
FedPAD (proposed)	32.17	28.84	76.51
All (upper bound, non-private)	27.26	25.09	80.42

Table 51 Effect of adding test-time adaptation on top of federated training, averaged (in %) across the four leave-one-out user datasets [66]

Method	Avg. HTER	Avg. EER	Avg. AUC
Single	41.61	36.66	67.07
Fused	35.75	31.29	73.89
FedPAD ([71], no test-time adaptation)	32.17	28.84	76.51
All (upper bound, non-private)	27.26	25.09	80.42
Proposed (FedPAD + test-time adaptation)	23.18	23.88	83.40

model, each user personalizes it locally. Adaptation operates by diminishing the entropy of the model's predictions on its own unannotated test dataset. No identifiers, gradients, or information are transmitted to the server, guaranteeing that this method is essentially privacy-protective. The research reproduces the exact experimental design of [71]: employing the same seven datasets, the identical ResNet-18 architecture, and the corresponding leave-one-dataset-out approach. Table 51 illustrates the key discovery: the integration of test-time adaptation with the federated model elevates the average AUC from 76.51% (FedPAD alone) to 83.40%, yielding a 6.89-point enhancement. It exceeds a benchmark that employs the same test-time adaptation on a single (non-federated) model (Avg. AUC 68.00%). Numerous additional results are displayed as incremental modifications of identical metrics instead of separate tables: the benefits increase in accordance with the number of data centers; it improves generalization to new 2D attack types, resulting in a 12.31 point enhancement in HTER and a 7.47

point increase in EER compared to FedPAD when each data center confronted only one 2D attack type; and it also enhances the generalization to previously unobserved 3D mask attacks, raising HTER by 17.73 points in comparison to FedPAD.

FedGPAD [56] uses FedPAD [71] as a baseline and presents an additional framework, FedGPAD, aimed at improving generalization in the face of unforeseen risks. The work used five solely 2D datasets (Oulu-NPU, CASIA-MFSD, Idiap Replay-Attack, MSU-MFSD, SiW), omitting the two 3D-mask datasets mentioned in [71]/ [66]. The fundamental principle of FedGPAD is the separation of federated domains. Each data center categorizes its architecture into a domain-invariant segment (feature extractor, classifier, depth estimator) and a domain-specific segment (feature extractor, decoder). Four local losses concurrently instruct both elements: classification, depth estimation, reconstruction, and an orthogonality term that preserves the distinction between the two components. The server receives solely the domain-invariant element, while the domain-specific element stays local. Table 52 displays the primary outcome. FedGPAD markedly exceeds FedPAD, achieving an average AUC of 89.25%, which is higher than the non-private standard. The upper threshold (80.42%) signifies that disentanglement surpasses the amalgamation of raw data.

Table 53 summarizes the security category, spanning attacks on training process, presentation-attack detection, and synthetic-media forgery detection. Unlike the other classifications, several of these techniques yield a federated result that aligns with or exceeds its own non-federated reference (FedGPAD, Fed-StackFPAD, FedSIS all demonstrate parity or superiority to either a centralized maximum or an earlier federated standard), while others explicitly indicate the reverse—a discernible privacy cost in relation to a centralized ablation (FedForgery). The foremost obstacles prevalent in this domain include the ability to generalize to attack or artifact types that are not included in the training data of any specific data center, alongside the challenge of balancing accuracy and privacy, which diverges across the two target level sub-clusters (presentation-attack compared to forgery detection).

Table 52 Comparison of models trained by data from a single data center vs. multiple, averaged across the four leave-one-out user datasets [56]

Method	Avg. HTER (%)	Avg. EER (%)	Avg. AUC (%)
Single	36.43	34.31	70.36
Fused	35.75	31.29	73.89
FedPAD	32.17	28.84	76.51
All (upper bound, non-private)	27.26	25.09	80.42
FedGPAD (Proposed)	18.59	17.48	89.25

3.7 Cross-method comparison: presentation-attack detection

Table 54 consolidates cross-domain HTER across the five presentation-attack detection papers that share the identical leave-one-domain-out protocol on Oulu-NPU (O), CASIA-MFSD (C), Idiap Replay-Attack (I), and MSU-MFSD (M). This represents an aggregation of previously published data rather than a systematic re-evaluation; however, the likelihood of incompatibility is diminished in this instance: [50] and [56] have already established benchmarks directly against [71] in their respective published tables, employing the same splits presented below. FedSIS [60] presents solely the overall average for all four divisions in its source table, lacking a detailed per-split analysis, so the per-split entries are denoted as "—"; the authors provide the average directly. The authors do not specify Fed-StackFPAD's [50] average, which is calculated here as the unweighted mean of its four disclosed splits. The table illustrates a distinct, consistent enhancement throughout the lineage: FedPAD (2020) → Federated Test-Time Adaptive fPAD (2021) → FedGPAD (2024) → Fed-StackFPAD / FedSIS (2023–2025), with the mean HTER decreasing from 32.17% to below 5% throughout five years of further research on the identical benchmark.

Three of these five papers—[66, 71], and [56]—also report per-split AUC and EER on this identical protocol, since all three build directly on the same original experimental setup. Fed-StackFPAD [50] and FedSIS [60] are not included in Tables 55–56, as neither reports these two metrics on a per-split basis in the source material. Table 55 shows that AUC improves in lockstep with the HTER gains above, and Table 56 shows the same for EER—both metrics improve most sharply on the O&C&M→I split, where FedGPAD's federated domain disentanglement (Sect. 3.6) yields its single largest gain over FedPAD (AUC 73.58%→95.34%; EER 26.54%→11.11%).

The four forgery-detection papers in this category [49, 58, 59, 63] are deliberately excluded from a combined cross-method table. Although their datasets overlap by name (FaceForensics++ and WildDeepfake recur across three of the four), each paper reports a different metric on a different constructed split: [49] reports Test Accuracy/Precision/F1 on a merged FF++/Deeperforensics-1.0/WildDeepfake set; [59] reports Accuracy/AUC on its own Generalized-domain held-out-artifact-type split; [63] reports per-manipulation-type AUC on FaceForensics++ at varying client counts; and [58] reports per-subset Accuracy on a four-dataset custom hybrid that includes CelebDF-v2 and FMFCC-V, absent from the other three.

Table 53 Summary of security techniques: attacks on the training process and recognition target

Paper	Dataset(s)	Metrics	FL framework	Key contribution
[40]	VGGFace2	Attack success rate, training loss	Time-variant aggregation (SaFL)	Process-level security: Sybil-aware aggregation defending against poisoning attacks on FL training
[71]	Oulu-NPU, CASIA-MFSD, Idiap Replay-Attack, MSU-MFSD, SiW, 3DMAD, HKBUMARsV2	HTER, EER, AUC	FedAvg (FedPAD)	First FL framework for fPAD; within 3.9 AUC points of the non-private upper bound
[66]	Oulu-NPU, CASIA-MFSD, Idiap Replay-Attack, MSU-MFSD, SiW, 3DMAD, HKBUMARsV2	HTER, EER, AUC	FedAvg + test-time entropy minimization	Adds privacy-preserving test-time adaptation; raises avg. AUC from 76.51% to 83.40%
[56]	Oulu-NPU, CASIA-MFSD, Idiap Replay-Attack, MSU-MFSD, SiW	HTER, EER, AUC	FedAvg on disentangled domain-invariant/specific parts (FedGPAD)	Federated domain disentanglement pushes avg. AUC to 89.25%, above the non-private upper bound
[50]	Oulu-NPU, CASIA-FASD, Idiap Replay-Attack, MSU-MFSD (leave-one-domain-out)	HTER	FedAvg + local models + SVM meta-classifier (Fed-StackFPAD)	Stacking ensemble of global/local probabilities cuts HTER by ~ 14 pts vs. FedGPAD
[60]	MCIO benchmark (MSU-MFSD, CASIA-MFSD, Idiap Replay-Attack, OULU-NPU)	Avg. HTER	Federated tokenizer/classifier + split-learned shared ViT encoder (FedSIS)	Block-sampled hybrid ViT; lowest avg. HTER of all compared methods, beating its own centralized counterpart
[49]	Merged FaceForensics++ / Deeperforensics-1.0 / WildDeepfake	Test Acc., Precision, F1	FedAvg (FFDL)	Dual-backbone (Xception + Inception) CCA feature fusion beats single-backbone federated baselines
[59]	FaceForensics++, Wild-Deepfake, Deeperforensics-1.0 + self-built Hybrid-domain split	Accuracy, AUC	FedAvg on VAE-residual features (FedForgery)	Classifies the reconstruction residual, not raw pixels; federated model trails both a non-federated baseline (DCL) and its own centralized ablation (explicit privacy cost)
[63]	FaceForensics++ (DeepFakes, Face2Face, FaceSwap, NeuralTextures)	AUC	FedAvg (Dynamic-Aware FL)	Inconsistency-Capture Module (global + local RoI branches); near-centralized at $K = 2$ clients, degrades sharply at $K = 5$
[58]	Custom Forgery Source Hybrid Dataset (FF++, WildDeepfake, CelebDF-v2, FMFCC-V)	Accuracy	FedAvg, shared + personalized branches (FedPR)	Personalized branch via AdaIN-style feature-statistic interpolation; matches/beats centralized ablation on 3 of 4 sub-datasets

Table 54 Cross-domain HTER (%) on the shared O/C/I/M leave-one-domain-out benchmark, consolidated across five security-category papers

Method	O&C&I $\rightarrow$ M	O&M&I $\rightarrow$ C	O&C&M $\rightarrow$ I	I&C&M $\rightarrow$ O	Avg.
FedPAD [71]	19.45	42.27	32.53	34.44	32.17
Federated TTA fPAD [66]	14.70	26.33	28.61	23.09	23.18
FedGPAD [56]	12.73	28.69	10.97	21.95	18.59
Fed-StackFPAD [50]	4.07	7.91	2.92	4.50	4.85
FedSIS [60]	–	–	–	–	3.91

Table 55 Cross-domain AUC (%) on the shared O/C/I/M leave-one-domain-out benchmark

Method	O&C&I $\rightarrow$ M	O&M&I $\rightarrow$ C	O&C&M $\rightarrow$ I	I&C&M $\rightarrow$ O	Avg.
FedPAD [71]	90.24	70.49	73.58	71.74	76.51
Fed. TTA fPAD [66]	90.57	77.77	82.07	83.21	83.40
FedGPAD [56]	91.25	80.58	95.34	89.85	89.25

Table 56 Cross-domain EER (%) on the shared O/C/I/M leave-one-domain-out benchmark

Method	O&C&I→M	O&M&I→C	O&C&M→I	I&C&M→O	Avg.
FedPAD [71]	17.43	36.95	26.54	34.45	28.84
Fed. TTA fPAD [66]	16.64	29.75	26.04	23.09	23.88
FedGPAD [56]	13.36	27.55	11.11	17.91	17.48

4 Discussion

This survey covers 50 recent studies between 2020 to 2026, focusing mainly on FFR. The entire paper is distributed across six categories. A few patterns stand out across all six categories.

- **The field is maturing, not just growing:** The initial works [23, 67, 71] propose and demonstrate the feasibility of FL in FR applications. However, later works assume feasibility and compete on harder problems such as generalization, fairness, and personalization under data heterogeneity. Two lineages make this visible: (1) FedPAD [71]→Federated TTA fPAD [66]→FedGPAD [56], and (2) FedFR [64]→pFedFace [53]→CL-FedFR [57].
- **Privacy, accuracy, and fairness trade-off:** This is a central tension across this survey. Almost every paper that touches more than one of these properties reports a measurable cost. A 35-point LFW accuracy swing across DP budgets in [43]; a $\sim 3000\times$ training-time cost for HE in [26]; F3's [61] explicit 0.4–17% accuracy cost for a 25–82% fairness gain; and FedForgery's [59] admitted AUC gap versus its own centralized version, all these results highlight this trade-off. Moreover, in this domain, no single framework fits this trade-off. Instead, each work makes its own local fix for data heterogeneity: clustering [31]), adaptive representations [32], split batch-normalization layers [53], server-side fairness aggregation [61], and prototype sharing [47]. The field would likely benefit from naming the different *kinds* of non-IID-ness these fixes target, rather than treating non-IID data as one single problem.
- **Personalization and generalization may not be simultaneously solvable:** CL-FedFR's [57] result is the clearest evidence that no single variant maximizes both personalized and generic accuracy at once. This looks like a genuine Pareto frontier, not a solved trade-off.
- **Shared benchmarks drive progress; their absence stalls it:** This is clearly evident from Sect. 3.7 results. The presentation-attack sub-cluster [50, 56, 60, 66, 71] converged on a shared O/C/I/M benchmark and improved average HTER from 32.17% to under 5% in five years, with several methods now matching or beating their non-federated reference point. The forgery-detection sub-cluster [49, 58, 59, 63] has no equivalent shared protocol and remains a set of isolated, non-comparable

results. This is a gap in benchmark standardization, not in algorithmic maturity.

- **Security splits into two distinct problems, and no paper addresses both:** Sect. 3.6 now makes this split explicit: security of the recognition *target* (presentation-attack and forgery detection) versus security of the training *process* (SaFL's [40] Sybil defense). A realistic adversary could plausibly attack both at once, but to the best of our knowledge, this domain has no work that defends against both.
- **No prior survey on FR treats FL as its main organizing axis:** The twelve-survey comparison (Table 1) confirms this. Only five of twelve FR surveys engage with FL at all, and none treat it centrally. [22] is the closest methodological peer, though it organizes by architecture and database rather than by training paradigm. This establishes that this survey's framing is novel, not that its taxonomy is the only valid one.
- **Limitations of this review:** The PRISMA protocol excludes non-English and purely theoretical work. Two records identified during the initial search, [91] and [92], could not be full-text verified and are excluded from the 50 studies reviewed here. Karra and Kakhandki [51] is retained despite evaluating facial expression rather than identity recognition, with this caveat noted at first mention. [45]'s comparison uses only five identities, which limits how far its conclusions generalize.

5 Future work

This section gives a list of concrete future works based on the discussion and the gaps mentioned in Sect. 4.

- **Build shared benchmarks outside presentation-attack detection.** The O/C/I/M protocol's organic adoption is exactly what made the comparison Sect. 3.7 possible. Forgery detection, assessing fairness, and evaluation of efficiency continue to depend on incompatible, ad-hoc datasets. A community-organized benchmark suite, similar to MLPerf or GLUE in other domains, would enable future studies to present directly comparable metrics rather than compelling subsequent reviews to either make uncertain comparisons or omit them entirely.
- **Explore Vertical Federated Learning (VFL) for FR.** This topic is completely ignored in the reviewed corpus.

A practical use-case is obvious: cross-institutional implementations where various clients possess complementary attributes for shared identities, rather than separate identity sets.

- **Apply formal cryptographic guarantees more rigorously.** HE, secret sharing, and MPC are implemented in only a handful of publications [26] and noted merely as a future direction in others [29]. These approaches are considerably less stringent than the (ϵ, δ) examination seen in DP literature, even though face embeddings are high-dimensional and linked to an identity that cannot be reissued.
- **Study multi-modal biometric fusion under federation.** The integration of FR with iris, vein, or gait recognition has not been addressed in any examined security-related literature, despite its significance for border control and high-security applications. The integration of many modalities introduces novel inquiries that this survey's corpus fails to tackle: per-modality client distribution, per-modality privacy budgets, and cross-modal spoofing.
- **Design a framework that defends both the training process and the recognition target.** No reviewed paper combines Sybil/poisoning defenses [40] with presentation-attack or forgery defenses (Sect. 3.6) in one system. This combination remains open.
- **Integrate federated unlearning with cancelable-template revocation.** Federated unlearning [55] and cancelable-template revocation [21] are currently separate lines of work. Combining them would let a client's exit from the system satisfy both data-erasure and template-revocation guarantees at the same time.
- **Report personalization results as trade-off curves, not single numbers.** CL-FedFR's [57] finding suggests personalized and generic accuracy trade off against each other. Future personalization work should report this trade-off directly, rather than reporting one best-configuration number.
- **Validate at realistic scale.** Several reviewed papers validate at a scale well below real deployment: [45]'s five-identity partition and synthetic client shards used in several fairness and personalization papers. Larger, multi-institution pilot studies are needed to confirm these methods transfer beyond simulated federation.
- **Address label noise and cross-client data quality.** Practical federated biometric deployments suffer from label noise and uneven data quality across clients, yet no reviewed study treats these as a first-class problem; the algorithmic-innovation category (Sect. 3.2) addresses non-IID distribution and class imbalance almost exclusively. Work in adjacent federated biometrics, such as FedRDA [93] and FedLN [94], offers robust-aggregation

and noise-correction strategies that have not yet been adapted to FFR, and represents a natural bridge toward federated biometrics more broadly.

6 Conclusion

This survey reviewed 50 studies (2020–2026) at the intersection of FL and FR, categorized into six categories, and validated through a comparison of twelve surveys (Table 1) that no previous research prioritizes FL as its central framework. A singular tension dominates various categories as the core concept of the field: privacy, accuracy, and fairness are often in conflict, with nearly every study addressing multiple dimensions reporting a quantifiable trade-off (e.g., the 35-point variation in DP accuracy [43], and the accepted gap between centralized and federated approaches in FedForgery [59]), and no standard framework currently encompasses all three aspects. The security category demonstrates the significance of shared benchmarks: the presentation-attack sub-cluster, having established a single protocol, improved from 32% to below 5% average HTER over five years, whereas forgery detection, devoid of a comparable standard, continues to yield disparate, non-comparable outcomes—a weakness in infrastructure rather than in methodological quality. Open problems worth dedicated future work include vertical federated learning for face recognition, rigorous cryptographic guarantees, multi-modal biometric fusion under federation, joint process-level and target-level security, and unified federated unlearning with cancelable-template revocation. Taken together, these findings position federated learning as a promising but unevenly developed foundation for privacy-preserving face recognition, and this survey aims to serve as a structured reference for what is solved, partially solved, and still open in this fast-moving field.

Author contributions Ajnas Muhammed prepared the manuscript All authors reviews the manuscript.

Funding This study has received funding from the European Union's Horizon Europe research and innovation programme ACHILLES under Grant Agreement No 101189689, and FCT - Fundação para a Ciência e a Tecnologia, I.P., under the projects UIDB/00048/2020 (DOI 10.54499/UIDB/00048/2020).

Data availability No datasets were generated or analysed during the current study.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share

adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- McMahan B, Moore E, Ramage D, Hampson S, Arcas BA (2017) Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics, pp 1273–1282. PMLR
- GDPR G (2016) General data protection regulation. Regulation (EU) 679
- Bonta R (2022) California consumer privacy act (ccpa). Retrieved from State of California Department of Justice: <https://oag.ca.gov/privacy/ccpa>
- Truong N, Sun K, Wang S, Guitton F, Guo Y (2021) Privacy preservation in federated learning: an insightful survey from the gdpr perspective. *Comput Secur* 110:102402
- Dritsas E, Trigka M (2025) Federated learning for IoT: a survey of techniques, challenges, and applications. *J Sens Actuator Netw* 14(1):9
- Guo G, Zhang N (2019) A survey on deep learning based face recognition. *Comput Vis Image Underst* 189:102805
- Wang L, Zhang R (2023) Framework for facial recognition and reconstruction for enhanced security and surveillance monitoring using 3d computer vision. *J Electron Imaging* 32(4):042108–042108
- Paula D, Alexandre LA (2022) Facial emotion recognition for sentiment analysis of social media data. In: Iberian conference on pattern recognition and image analysis. Springer, pp 207–217
- Simmmler M, Canova G (2025) Facial recognition technology in law enforcement: regulating data analysis of another kind. *Comput Law Secur Rev* 56:106092
- Laishram L, Shaheryar M, Lee JT, Jung SK (2025) Toward a privacy-preserving face recognition system: a survey of leakages and solutions. *ACM Comput Surv* 57(6):1–38
- Hasan MR, Guest R, Deravi F (2023) Presentation-level privacy protection techniques for automated face recognition—a survey. *ACM Comput Surv* 55(13s):1–27
- Zeng L, Shen P, Zhu X, Tian X, Chen C (2025) A review of privacy-preserving biometric identification and authentication protocols. *Comput Secur*, 104309
- Wang M, Li S, Zhang X, Feng G (2025) Facial privacy in the digital era: a comprehensive survey on methods, evaluation, and future directions. *Comput Sci Rev* 58:100785
- Du H, Shi H, Zeng D, Zhang X, Mei T (2022) The elements of end-to-end deep face recognition: a survey of recent advances. *ACM Comput Surv* 54(10s):212–121242. <https://doi.org/10.1145/3507902>
- Wang M, Deng W (2021) Deep face recognition: a survey. *Neurocomputing* 429:215–244. <https://doi.org/10.1016/J.NEUCOM.2020.10.081>
- Leyva R, Epiphaniou G, Maple C (2026) Attack vectors for face recognition systems: a comprehensive review. *ACM Comput Surv* 58(1):19–11937. <https://doi.org/10.1145/3736753>
- Roux QL, Bourbao E, Teglia Y, Kallas K (2024) A comprehensive survey on backdoor attacks and their defenses in face recognition systems. *IEEE Access* 12:47433–47468. <https://doi.org/10.1109/ACCESS.2024.3382584>
- Kotwal K, Marcel S (2026) Review of demographic fairness in face recognition. *IEEE Trans Biom Behav Identity Sci* 8(1):20–45. <https://doi.org/10.1109/TBIOM.2025.3601217>
- Yucer S, Tektas F, Moubayed NA, Breckon TP (2025) Racial bias within face recognition: a survey. *ACM Comput Surv* 57(4):105–110539. <https://doi.org/10.1145/3705295>
- Hahn VK, Marcel S (2023) Biometric template protection for neural-network-based face recognition systems: a survey of methods and evaluation techniques. *IEEE Trans Inf Forensics Secur* 18:639–666. <https://doi.org/10.1109/TIFS.2022.3228494>
- Hayat A, Ahad MAR, Kumar V, Bhateja AK (2025) Deepshield-fed: Securing face templates via deep cancelable transforms and federated learning. In: IEEE international joint conference on biometrics, IJCB 2025, Osaka, Japan, September 8–11, 2025, pp. 1–10. IEEE. <https://doi.org/10.1109/IJCB65343.2025.11410798>
- Nemavhola A, Chibaya C, Viriri S (2025) A systematic review of CNN architectures, databases, performance metrics, and applications in face recognition. *Information* 16(2):107. <https://doi.org/10.3390/INFO16020107>
- Kim J, Park T, Kim H, Kim S (2021) Federated learning for face recognition. In: 2021 IEEE international conference on consumer electronics (ICCE), pp 1–2. IEEE
- Niu Y, Deng W (2022) Federated learning for face recognition with gradient correction. *Proc AAAI Conf Artif Intell* 36:1999–2007
- Aggarwal D, Zhou J, Jain AK (2021) Fedface: collaborative learning of face recognition model. In: 2021 IEEE international joint conference on biometrics (IJCB). IEEE, pp 1–8
- Franco D, Oneto L, Navarin N, Anguita D (2021) Toward learning trustworthily from data combining privacy, fairness, and explainability: an application to face recognition. *Entropy* 23(8):1047
- Zhuang W, Gan X, Wen Y, Zhang X, Zhang S, Yi S (2021) Towards unsupervised domain adaptation for deep face recognition under privacy constraints via federated learning. *arXiv preprint arXiv:2105.07606*
- Zhuang W, Gan X, Zhang X, Wen Y, Zhang S, Yi S (2022) Federated unsupervised domain adaptation for face recognition. In: IEEE international conference on multimedia and expo, ICME 2022, Taipei, Taiwan, July 18–22, pp 1–6. IEEE. <https://doi.org/10.1109/ICME52920.2022.9859587>
- Yang J, Liu J, Han R, Wu J (2021) Transferable face image privacy protection based on federated learning and ensemble models. *Complex Intell Syst* 7(5):2299–2315
- Ding Y, Wu X, Li Z, Wu Z, Tan S, Xu Q, Pan W, Yang Q (2022) An efficient industrial federated learning framework for aiot: a face recognition application. *arXiv preprint arXiv:2206.13398*
- Meng Q, Zhou F, Ren H, Feng T, Liu G, Lin Y (2022) Improving federated learning face recognition via privacy-agnostic clusters. *arXiv preprint arXiv:2201.12467*
- Qiu D, Lin X, Wang K, Chu X, Yan P (2024) Adafedfr: federated face recognition with adaptive inter-class representation learning. *arXiv preprint arXiv:2405.13467*
- Brănescu I, Ciobanu R-I, Dobre C, Mavromoustakis C (2023) Decentralized machine learning for face recognition. In: 2023 22nd International symposium on parallel and distributed computing (ISPDC). IEEE, pp 1–8
- Kim H, Choi H, Kwak Y (2024) Federated learning for face recognition via intra-subject self-supervised learning. *arXiv preprint arXiv:2407.16289*
- Solomon E, Woubie A (2024) Federated learning method for preserving privacy in face recognition system. *arXiv preprint arXiv:2403.05344*
- Woubie A, Solomon E, Attieh J (2024) Maintaining privacy in face recognition using federated learning method. *IEEE Access*
- Gansekoele A, Hess E, Bhulai S (2024) Meta-learning for federated face recognition in imbalanced data regimes. In: 2024 2nd

- international conference on federated learning technologies and applications (FLTA). IEEE, pp 24–31
38. Hess E (2023) The application and analysis of meta-learning in federated face recognition
 39. Lenin S, Srivatsan G, Basha GA, Aswin Z (2023) Privacy-preserving integration of face recognition system and esd tester using federated learning. In: 2023 International conference on circuit power and computing technologies (ICCPCT). IEEE, pp 1246–1250
 40. Ghafourian M, Fierrez J, Vera-Rodriguez R, Tolosana R, Morales A (2024) Safi: sybil-aware federated learning with application to face recognition. In: 2024 IEEE international conference on image processing challenges and workshops (ICIPCW). IEEE, pp 4050–4056
 41. Solomon E, Woubie A, Emiru ES, Abdelzaher AF (2024) Secure and efficient face recognition via supervised federated learning. In: 2024 33rd IEEE international conference on robot and human interactive communication (ROMAN). IEEE, pp 291–296
 42. Wang W, Li S (2024) Dp-fedface: Privacy-preserving facial recognition in real federated scenarios. In: Proceedings of the 33rd ACM international conference on information and knowledge management, pp 4123–4127
 43. Zarei A, Hassanpour A, Raja K (2025) On privacy, accuracy, and fairness trade-offs in facial recognition. IEEE Access
 44. Jiang B, Liu Z, Yue G, Cui X, Liu Y, Wang HH (2025) Privacy safeguarding for human-computer interaction based on adaptive federated learning with actor-critic selection. IEEE Trans Consumer Electron
 45. Nusa S, Juwono Y, Ma'Rufah L, Armawan I, Shiddiqi AM (2024) A comparative study of federated learning methods for face recognition on non-iid dataset. In: 2024 IEEE international conference on artificial intelligence and mechatronics systems (AIMS), pp 1–6. IEEE
 46. Vyas H, Torrens PM (2024) Tackling “leaky” facial and object recognition operations over federated learning. In: 2024 4th Interdisciplinary conference on electrics and computer (INTCEC). IEEE, pp 1–5
 47. Kim T, Kim J, Teoh ABJ (2026) A prototype-based framework for open-set heterogeneous federated face recognition. Expert Syst Appl 299:130020. <https://doi.org/10.1016/J.ESWA.2025.130020>
 48. Peng J, Sun H, Yang D, El-Latif AAA, Rodrigues JJPC (2026) Prisedfedfr: privacy-secure face recognition model training via federated learning and random projection. Expert Syst Appl 297:129383. <https://doi.org/10.1016/J.ESWA.2025.129383>
 49. Gautam V, Kaur G, Malik M, Pawar A, Singh A, Singh KK, Askar SS, Abouhawwash M (2025) FFDL: feature fusion-based deep learning method utilizing federated learning for forged face detection. IEEE Access 13:5366–5379. <https://doi.org/10.1109/ACCESS.2024.3523257>
 50. Gündogar MF, Erdem ÇE, Korçak Ö (2025) Fed-stackfpad: federated learning for face presentation attack detection with stacking to tackle data heterogeneity. IEEE Access 13:190354–190370. <https://doi.org/10.1109/ACCESS.2025.3629353>
 51. Karra SR, Kakhandki AL (2025) Towards robust adaptive federated learning: unsupervised multidomain face recognition with greater cane rat algorithm and weighted fused extreme learning machine. Iran J Comput Sci 8(4):2083–2099. <https://doi.org/10.1007/S42044-025-00302-3>
 52. Makhanov N, Amirgaliyev B, Islamgozhayev T, Yedilkhan D (2025) Federated self-supervised few-shot face recognition. J Imaging 11(10):370. <https://doi.org/10.3390/JIMAGING11100370>
 53. Gao Y, Su A, Zhao C, Song J (2025) pfedface: Personalized federated learning for face recognition. In: (2025) IEEE international conference on acoustics, speech and signal processing, ICASSP 2025, Hyderabad, India, April 6–11, 2025, pp 1–5. IEEE. <https://doi.org/10.1109/ICASSP49660.2025.10889072>
 54. Solomon E (2025) Unsupervised federated learning for face recognition in decentralized environments. In: 8th International conference on information and computer technologies, ICTT 2025, Hawaii-Hilo, HI, USA, March 14–16, 2025, pp 101–106. IEEE. <https://doi.org/10.1109/ICTT64582.2025.00022>
 55. Qi F, Liu A, Zhang Z, Xu C (2025) FORGET ME: federated unlearning for face generation models. In: Gurrin C, Schoeffmann K, Zhang M, Rossetto L, Rudinac S, Dang-Nguyen D, Cheng W, Chen P, Benois-Pineau J. (eds) Proceedings of the 33rd ACM international conference on multimedia, MM 2025, Dublin, Ireland, October 27–31, 2025, pp 11288–11297. ACM. <https://doi.org/10.1145/3746027.3754935>
 56. Shao R, Perera P, Yuen PC, Patel VM (2024) Federated generalized face presentation attack detection. IEEE Trans Neural Netw Learn Syst 35(1):103–116. <https://doi.org/10.1109/TNNLS.2022.3172316>
 57. Dube DC, Erdem ÇE, Korçak Ö (2024) Cl-fedfr: curriculum learning for federated face recognition. In: Radeva P, Furnari A, Bouatouch K, Sousa AA (eds) Proceedings of the 19th international joint conference on computer vision, imaging and computer graphics theory and applications, VISIGRAPP 2024, Volume 2: VISAPP, Rome, Italy, February 27–29, pp 845–852. SCITEPRESS. <https://doi.org/10.5220/0012574000003660>
 58. Liu D, Dang Z, Peng C, Wang N, Hu R, Gao X (2024) Federated face forgery detection learning with personalized representation. CoRR
 59. Liu D, Dang Z, Peng C, Zheng Y, Li S, Wang N, Gao X (2023) Fedforgery: generalized face forgery detection with residual federated learning. IEEE Trans Inf Forensics Secur 18:4272–4284. <https://doi.org/10.1109/TIFS.2023.3293951>
 60. Alkhunaizi N, Srivatsan K, Almalik F, Almakky I, Nandakumar K (2023) Feddis: Federated split learning with intermediate representation sampling for privacy-preserving generalized face presentation attack detection. In: IEEE international joint conference on biometrics, IJCB 2023, Ljubljana, Slovenia, September 25–28, 2023, pp. 1–11. IEEE. <https://doi.org/10.1109/IJCB57857.2023.10448785>
 61. Kanaparthi S, Padala M, Damle S, Sarvadevabhatla RK, Gujar S (2023) F3: fair and federated face attribute classification with heterogeneous data. In: Kashima H, Idé T, Peng W (eds) Advances in knowledge discovery and data mining—27th Pacific-Asia conference on knowledge discovery and data mining, PAKDD 2023, Osaka, Japan, May 25–28, 2023, Proceedings, Part I. Lecture Notes in Computer Science, vol. 13935, pp 483–494. Springer. https://doi.org/10.1007/978-3-031-33374-3_38
 62. Liu L, Zhang Y, Gao H, Yu X, Cheng J (2022) Fedfv: federated face verification via equivalent class embeddings. Multimed Syst 28(5):1833–1843. <https://doi.org/10.1007/S00530-022-00927-5>
 63. Hu Z, Xie H, Yu L, Gao X, Shang Z, Zhang Y (2022) Dynamic-aware federated learning for face forgery video detection. ACM Trans Intell Syst Technol 13(4):57–15725. <https://doi.org/10.1145/3501814>
 64. Liu C, Wang C, Chien S, Lai S (2022) Fedfr: joint optimization federated framework for generic and personalized face recognition. In: Thirty-sixth AAAI conference on artificial intelligence, AAAI 2022, thirty-fourth conference on innovative applications of artificial intelligence, IAAI 2022, the twelfth symposium on educational advances in artificial intelligence, EAAI 2022 virtual event, February 22–March 1, 2022, pp 1656–1664. AAAI Press. <https://doi.org/10.1609/AAAI.V36I2.20057>
 65. Roh J, Fang Y (2022) Robust smart home face recognition under starving federated data. In: 6th International conference on universal village, UV 2022, Boston, MA, USA, October 22–25, 2022, pp 1–11. IEEE. <https://doi.org/10.1109/UV56588.2022.10185525>

66. Shao R, Zhang B, Yuen PC, Patel VM (2021) Federated test-time adaptive face presentation attack detection with dual-phase privacy preservation. In: 16th IEEE international conference on automatic face and gesture recognition, FG 2021, Jodhpur, India, December 15–18, 2021, pp 1–8. IEEE. <https://doi.org/10.1109/FG52635.2021.9666952>
67. Bai F, Wu J, Shen P, Li S, Zhou S (2021) Federated face recognition. CoRR [arXiv:abs/2105.02501](https://arxiv.org/abs/2105.02501)
68. Kanaparthi S, Padala M, Damle S, Gujar S (2021) Fair federated learning for heterogeneous face data. CoRR [arXiv:abs/2109.02351](https://arxiv.org/abs/2109.02351)
69. Oza P, Patel VM (2021) Federated learning-based active authentication on mobile devices. 2021 IEEE international joint conference on biometrics (IJCB). IEEE, pp 1–8
70. Triastcyn A, Faltings B (2020) Federated generative privacy IEEE Intell Syst 35(4):50–57
71. Shao R, Perera P, Yuen PC, Patel VM (2020) Federated face presentation attack detection. CoRR [arXiv:abs/2005.14638](https://arxiv.org/abs/2005.14638)
72. Yang J, Liu J, Wu J (2020) Facial image privacy protection based on principal components of adversarial segmented image blocks. IEEE Access 8:103385–103394
73. Zhang Y, Yin Z, Li Y, Yin G, Yan J, Shao J, Liu Z (2020) Celebapspoo: large-scale face anti-spoofing dataset with rich annotations. In: Computer vision—ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part XII 16, pp 70–85. Springer
74. Wang M, Deng W (2020) Mitigating bias in face recognition using skewness-aware reinforcement learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9322–9331
75. Weiss GM (2019) Wisdm smartphone and smartwatch activity and biometrics dataset. UCI machine learning repository WISDM smartphone and smartwatch activity and biometrics dataset data set 7(133190–133202):5
76. Yi D, Lei Z, Liao S, Li SZ (2014) Learning face representation from scratch. arXiv preprint [arXiv:1411.7923](https://arxiv.org/abs/1411.7923)
77. Shahreza HO, Rathgeb C, Osorio-Roig D, Hahn VK, Marcel S, Busch C (2022) Hybrid protection of biometric templates by combining homomorphic encryption and cancelable biometrics. In: 2022 IEEE international joint conference on biometrics (IJCB). IEEE, pp 1–10
78. Abdellatef E, Soliman RF, Omran EM, Ismail NA, Elrahman SEA, Ismail KN, Rihan M, Amin M, Eisa AA, El-Samie FEA (2022) Cancelable face and iris recognition system based on deep learning. Opt Quant Electron 54(11):702
79. Sardar A, Umer S, Pero C, Nappi M et al (2020) A novel cancelable facehashing technique based on non-invertible transformation with encryption and decryption template. IEEE Access 8:105263–105277
80. Wu L, Meng L, Zhao S, Wei X, Wang H (2021) Privacy-preserving cancelable biometric authentication based on rdm and ecc. IEEE Access 9:90989–91000
81. Bae G, La Gorce M, Baltrušaitis T, Hewitt C, Chen D, Valentin J, Cipolla R, Shen J (2023) Digiface-1m: 1 million digital face images for face recognition. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 3526–3535
82. Cao Q, Shen L, Xie W, Parkhi OM, Zisserman A (2018) Vggface2: a dataset for recognising faces across pose and age. In: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), pp 67–74. IEEE
83. Maze B, Adams J, Duncan JA, Kalka N, Miller T, Otto C, Jain AK, Niggel WT, Anderson J, Cheney J (2018) Iarpa janus benchmark-c: Face dataset and protocol. In: 2018 international conference on biometrics (ICB). IEEE, pp 158–165
84. Guo Y, Zhang L, Hu Y, He X, Gao J (2016) Ms-celeb-1m: a dataset and benchmark for large-scale face recognition. In: Computer vision—ECCV 2016: 14th European conference, Amsterdam, The Netherlands, October 11–14, 2016, proceedings, Part III 14, pp 87–102. Springer
85. Wang M, Deng W, Hu J, Tao X, Huang Y (2019) Racial faces in the wild: reducing racial bias by information maximization adaptation network. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 692–702
86. Whitelam C, Taborsky E, Blanton A, Maze B, Adams J, Miller T, Kalka N, Jain AK, Duncan JA, Allen K (2017) Iarpa janus benchmark-b face dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 90–98
87. Huang GB, Mattar M, Berg T, Learned-Miller E (2008) Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In: Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition
88. Sengupta S, Chen J-C, Castillo C, Patel VM, Chellappa R, Jacobs DW (2016) Frontal to profile face verification in the wild. In: 2016 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1–9
89. Moschoglou S, Papaioannou A, Sagonas C, Deng J, Kotsia I, Zafeiriou S (2017) Agedb: the first manually collected, in-the-wild age database. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 51–59
90. Kärkkäinen K, Joo J (2019) Fairface: Face attribute dataset for balanced race, gender, and age. arXiv preprint [arXiv:1908.04913](https://arxiv.org/abs/1908.04913)
91. Yang S, Wang M, Deng W, Hu J (2024) Fedsc: federated generalized face anti-spoofing via shuffled codebook. In: Antonacopoulos A, Chaudhuri S, Chellappa R, Liu C, Bhattacharya S, Pal U (eds) Pattern recognition - 27th international conference, ICPR 2024, Kolkata, India, December 1–5, 2024, Proceedings, Part III. Lecture Notes in Computer Science, vol. 15303, pp 226–242. Springer. https://doi.org/10.1007/978-3-031-78122-3_15
92. Shang E, Yang Z, Liu H, Du J, Wang X (2022) Fedfr: evaluation and selection of loss functions for federated face recognition. In: Gao H, Wang X, Wei W, Dagiuklas T (eds) Collaborative computing: networking, applications and worksharing—18th EAI international conference, CollaborateCom 2022, Hangzhou, China, October 15–16, 2022, Proceedings, Part I. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 460, pp 95–114. Springer. https://doi.org/10.1007/978-3-031-24383-7_6
93. Xie Y, Ren H, He H, Fei S, Guo J, Sun L (2025) Fedrda: hierarchical noise detection for federated finger vein recognition. IEEE Trans Inf Forensics Secur
94. Tsouvalas V, Saeed A, Ozcelebi T, Meratnia N (2024) Labeling chaos to learning harmony: federated learning with noisy labels. ACM Trans Intell Syst Technol 15(2):1–26
95. Bakas, S., Li, X., Shah, P., & Roth, H. R. (2026). Federated learning in healthcare: From research to real-world deployment. Annual Review of Biomedical Engineering, 28. <https://doi.org/10.1146/annurev-bioeng-080125-041414>
96. Muhammed, Ajnas, João Marcos, and Nuno Gonçalves. "Federated Learning for Secure and Privacy-Preserving Facial Recognition: Advances, Challenges, and Research Directions." Iberian Conference on Pattern Recognition and Image Analysis. Cham: Springer Nature Switzerland, 2025.