Federated Learning for Secure and Privacy-Preserving Facial Recognition: Advances, Challenges, and Research Directions

Ajnas Muhammed(✉), João Marcos, and Nuno Gonçalves

Institute of Systems and Robotics, University of Coimbra,
3030-290 Coimbra, Portugal
{ajnas.muhammed,joao.marcos}@isr.uc.pt, nunogon@deec.uc.pt

Abstract. Federated learning is an innovative, decentralized machine learning paradigm that allows multiple devices or entities to collaboratively train a shared model without transferring data to a central server. By keeping data localized, this distributed approach ensures enhanced privacy and security for each participating node. Facial recognition, a rapidly evolving field, leverages deep learning techniques to achieve remarkable advancements, often surpassing human-level performance on certain datasets. However, the sensitive nature of facial data, which contains personally identifiable information, raises significant privacy and security concerns. Federated learning has emerged as a promising solution to address these privacy challenges in the facial recognition community. This paper presents a comprehensive review of existing literature on facial recognition frameworks utilizing federated learning. The reviewed techniques are systematically categorized to provide a structured analysis, emphasizing their contributions and relevance to the broader domain of federated learning-based facial recognition. Specifically, this work aims to summarize and analyze various federated learning-based facial recognition methods, their underlying techniques, and their objectives. Furthermore, it offers a high-level perspective on how different functionalities and design principles of federated learning have been applied in facial recognition applications. By doing so, this review identifies key challenges and highlights promising research directions for future advancements in the field.

Keywords: Federated Learning · Face Recognition · Privacy

1 Introduction

Federated Learning (FL) is a decentralized machine learning framework that enables multiple entities or devices to collaboratively train a shared model while keeping data localized. This approach is particularly valuable in scenarios where data privacy, security, and regulatory compliance are paramount, such as in

N. Gonçalves et al. (Eds.): IbPRIA 2025, LNCS 15937, pp. 199–213, 2026.
https://doi.org/10.1007/978-3-031-99565-1_16

healthcare, finance, biometrics, etc. By storing data locally and sharing only model updates, FL minimizes the risk of data breaches, enhances privacy and security, and ensures compliance with regulations like the European Union's General Data Protection Regulation (GDPR) [12]. Adhering to the principles defended by different regulations such as the European GDPR, is critical for any technology handling personal data, including FL systems. Several studies have explored how FL can be designed to meet these regulatory requirements [41]. Additionally, FL reduces the need for large-scale data transfers, conserving bandwidth and computational resources. These advantages have made FL a focal point of interest for both academic research and industry applications [9].

Face Recognition (FR) systems identify and authenticate individuals by analyzing their unique facial features. Over the years, FR has seen significant advancements due to improvements in machine learning algorithms, the availability of large datasets, and the increased computational power of modern hardware [14]. In recent times, particularly following the COVID-19 pandemic, FR has become a critical component of modern security systems [42], social media platforms [34], border control, and various other essential applications [37]. Despite its widespread adoption and numerous benefits, FR systems raise significant privacy concerns. Traditional FR approaches often rely on centralized data storage to collect and process sensitive facial information, which creates vulnerabilities to single-point attacks and data breaches [27]. Privacy preservation in FR is of utmost importance, as facial data constitutes personally identifiable information. Unauthorized access or misuse of this data can lead to serious consequences, including surveillance overreach, identity theft, and violations of personal freedom [16]. Consequently, there is a growing demand for techniques that can effectively preserve privacy while maintaining the accuracy and reliability of FR systems [27]. Addressing these challenges is critical to ensuring the ethical and secure deployment of FR technologies in various domains.

While FR systems have achieved remarkable accuracy and widespread application, they face significant challenges related to data privacy, security, and regulatory compliance. Recently, FR community has embraced FL as a promising solution to address pressing privacy concerns. FL facilitates decentralized data processing and model training, ensuring that sensitive facial data remains on local devices rather than being transferred to a central server. This approach not only enhances privacy but also improves security by significantly reducing the risk of data breaches. Furthermore, FL enables collaboration across diverse datasets, fostering the development of more robust and inclusive FR systems. By keeping data localized and sharing only model updates, FL aligns with stringent privacy regulations, making it a viable framework for building ethical and compliant FR technologies. The rapid advancements FR in FL, have created a critical need for a comprehensive survey paper in this domain.

Hence in this paper, we provide a comprehensive survey to systematically organize, analyze, and synthesize the growing body of research on FL-based FR. The primary objective of this survey is to provide a comprehensive overview as well as a bird's eye view of the recent advancements in FL-based FR. By review-

ing and classifying the latest research papers in this field, we aim to highlight the key contributions, identify the challenges, and suggest future directions for research. This survey will serve as a valuable resource for researchers, practitioners, and policymakers interested in the intersection of FL and FR. Table 1 shows the different papers reviewed in this survey with its year and key contributions.

Table 1. Details regarding the works reviewed in this survey.

Method	Year	Key contributions
[24]	2021	- Introduced FL in FR community - Employ gradient averaging to ensure model convergence
[32]	2022	- Proposed a gradient correction mechanism to improve the performance of FL-based FR systems - Addressed the issue of gradient divergence by adjusting the gradients
[2]	2021	- Integration of differential privacy into FL-based FR, ensures robustness and privacy
[10]	2021	- Proposed a technique that combines privacy, fairness, and explainability
[56]	2021	- Demonstrated that FL can be used for domain adaptation in FR while preserving data privacy
[49]	2021	- Used ensemble models in FL for improved transferability and privacy
[8]	2022	- Demonstrated the feasibility of deploying FL-based FR in industrial settings with limited resources
[30]	2022	- Enhanced model performance and privacy by reducing the impact of non-Independent and Identically Distributed (non-IID) data
[35]	2024	- Achieved better performance in FL-based FR tasks, particularly in imbalanced datasets using adaptive inter-class representation learning technique
[6]	2023	- Demonstrated the feasibility of decentralized FL for FR and highlights its advantages in terms of privacy and scalability
[23]	2024	- Introduced a self-supervised learning technique which reduces the need for labeled data and enhances model accuracy
[38, 48]	2024	- Proposed a privacy-preserving FL framework for FR using techniques like differential privacy and secure aggregation to protect sensitive data
[11]	2024	- Introduced a meta-learning approach to improve the model performance in imbalanced datasets and enhance generalization
[28]	2023	- Integrates FL-based FR with electrostatic discharge (ESD) testing
[13]	2024	- Introduced a Sybil-aware FL framework for FR - Detects and mitigates Sybil attacks (malicious clients)
[39]	2024	- Introduced a supervised FL framework for FR - Used labeled data from clients to train models and incorporates secure aggregation techniques to protect data privacy
[17]	2023	- Introduced a meta-learning to improve model generalization and reduces the need for retraining
[45]	2024	- Demonstrated the feasibility of using differential privacy in real-world FL-based FR
[26]	2024	- Explored the use of FL for FR in personal smart media cloud solutions while preserving privacy
[51]	2025	- Explored the trade-offs between privacy, accuracy, and fairness in FL-based FR - Used differential privacy, fairness-aware algorithms and similar methods to balance these trade-offs
[20]	2025	- Proposed an adaptive FL framework for privacy-preserving human-computer interaction and uses actor-critic selection to enhance fairness and privacy

2 Federated Learning for Face Recognition

This section categorizes various Federated Face Recognition (FFR) techniques into five distinct classes: 1). privacy-preserving techniques (Sect. 2.1), 2). algorithmic innovations (Sect. 2.2), 3). fairness and explainability (Sect. 2.3), 4). domain adaptation and transfer learning (Sect. 2.4), and 5). efficiency and scalability (Sect. 2.5). This classification is based on the primary focus areas, research directions, and contributions of the reviewed studies. Privacy-preserving techniques encompass methods and frameworks specifically designed to enhance data privacy in FFR systems. Algorithmic innovations highlight novel algorithms and optimization strategies aimed at improving the performance and robustness of FFR models. The fairness and explainability section explores approaches that address ethical considerations, bias mitigation, and transparency in FFR systems. Domain adaptation and transfer learning focus on techniques that enable FR models to generalize effectively across diverse domains and datasets. Finally, the efficiency and scalability section discusses strategies for optimizing communication, reducing latency, and enhancing computational efficiency in FFR systems. This structured classification provides a comprehensive framework for analyzing the advancements and challenges in FFR research.

In FFR, since the models trains locally on distributed devices and shares only model updates, not raw data, directly comparing performance of two FFR techniques is difficult. The performance of FFR depends on local data distribution, device quality, and aggregation communication protocols. As each federated system operates in a different environment with different data privacy constraints, device capabilities, and network conditions, makes the performance comparisons impossible. Similarly, lack of a standardized benchmark or centralized dataset for evaluation makes it difficult to compare two FFR systems. FFR systems have significant performance trade-offs compared to centralized (non-FFR) systems. Large, centralized datasets help non-FFR systems perform better. FL localize data for privacy and security, but limited data diversity and potential biases in local datasets may lower performance. Due to these inherent complexities and challenges associated with FFR, we have not included performance comparisons in this study.

2.1 Privacy Preservation in Federated Face Recognition

Privacy preservation techniques in FFR prioritize safeguarding sensitive facial data while ensuring optimal model performance. A technique for protecting the privacy of transferable face images using FL and ensemble models is presented in [49]. This technique can generate novel private facial images within a specific design framework. The experiment and results demonstrate that the generated facial images enhance transferability. They also assert that differential privacy or homomorphic encryption may be implemented to enhance protection and transferability.

PrivacyFace, a framework designed to enhance FFR by facilitating the exchange of auxiliary and privacy-agnostic information among clients is introduced in [30]. PrivacyFace comprises a practical Differentially Private Local Clus-

tering (DPLC) mechanism to extract sanitized clusters from local class centers, along with a consensus-aware recognition loss that enhances the discriminative quality of the results. Experimental results demonstrate that the technique introduces a minimal overhead while also achieving superior performance.

An additional FFR technique that incorporates FL with FR classifiers, both supervised and unsupervised, while ensuring user privacy, is presented in [38,48]. A GAN network has been employed on edge devices to generate counterfeit data. This process alleviates the transmitting burden on the available bandwidth and exposes potential vulnerabilities by disclosing specific information about the local user. The experimental results of the CelebA [53] dataset reveal that federated learning introduces significant advantages in both supervised and unsupervised face recognition tasks. Experimental findings indicate that the application of federated learning (FL) in both supervised and unsupervised facial recognition (FR) systems provides advantages such as enhanced privacy, as the original data is retained on edge devices, and the aggregated model demonstrates performance nearly equivalent to that of individual models, especially when the federated model does not employ a secure aggregator.

In [13], the authors introduced a novel defense mechanism against poisoning attacks in Federated Learning, termed Sybil-aware Federated Learning (SaFL), which mitigates the impact of sybils through an innovative time-variant aggregation scheme. In the experiment, they compared the method against a baseline without protection and two other widely used mitigation techniques. The experimental results demonstrate that the technique attains a high protection rate with minimal impact on performance, enhancing federated learning protection against targeted poisoning attacks.

Table 2. Summary of privacy preservation FFR.

Paper	Dataset(s) Used	Metrics	FL Framework	Key Contributions
[49]	VGGFace, VGGFace2 [7]	Privacy Metrics, Accuracy	Ensemble FL Framework	Uses ensemble models in FL to enhance privacy protection in FR
[30]	CASIA-WebFace [50], BUPT-Balancedface [43] IJB-B [47], IJB-C [29]	Privacy Metrics	FedAvg with Clustering	Introduces privacy-agnostic clustering to improve FFR
[38,48]	CelebA [53]	Privacy Metrics	FedAvg	Focuses on privacy-preserving FFR
[13]	VGGFace2 [7]	Accuracy, Sybil Attack Resilience	SaFL	Proposes Sybil-aware FL to enhance security in FR
[45]	CASIA-WebFace [50], LFW [18], IJB-B [47] IJB-C [29]	Accuracy, Differential Privacy Metrics	DP-FedFace	Introduces differential privacy in FFR
[20]	CIFAR-10 [1], CelebA [53], WISDM [46]	Accuracy, Privacy Metrics	Hierarchical FL Framework	Uses hierarchical FL with actor-critic selection for privacy-preserving FR

A privacy protection framework for facial recognition, termed DPFedFace, is presented for realistic scenarios in which each device contains solely the owner's facial data, as discussed in [45]. The technique safeguards user privacy by removing visually critical low-frequency components through human and model perception. They also implemented a learning privacy cost allocation mechanism that optimizes allocation strategies and incorporates noise into frequency domain features. The experimental results demonstrate that by implementing differential privacy and introducing noise, the technique attained strong privacy protection, thereby indicating that the attacker cannot reconstruct the face images from the uploaded class embeddings.

In [20], the authors presented a hierarchical federated learning framework to enhance client selection and model aggregation in human-computer interaction systems, specifically targeting facial recognition tasks. The framework employs an Actor-Critic based dynamic Client Selection mechanism, which modifies client participation in real-time during training, thereby optimizing the balance between exploration and exploitation to improve model accuracy and training efficiency. The multidimensional weighted aggregation method enhances the robustness and performance of the global model by considering factors such as data volume, loss values, and uncertainty in the context of heterogeneous data. Experimental findings on different datasets indicate that the proposed method attains accelerated convergence and superior accuracy relative to conventional random selection techniques. A summary of all privacy-preserving FFR techniques is presented in Table 2.

2.2 Algorithmic Innovations

This section explains different FFR technique with respect to algorithmic innovations. Algorithmic innovation class is chosen as FFR is a rapidly evolving field, and many papers introduce novel algorithms to improve model performance, robustness, and adaptability. Algorithmic innovations in FFR aim to improve the performance and convergence.

Gradient correction is a crucial technique in FL aimed at enhancing privacy while maintaining model performance. One notable approach is Federated Gradient Correction (FedGC) [32], which addresses the issue of gradient leakage. In this work, authors explore the idea of correcting gradients from the perspective of backward propagation and propose a softmax-based regularizer to correct gradients of class embeddings by precisely injecting a cross-client gradient term. Extensive experiments shows the superiority of FedGC, which can match the performance of conventional centralized methods.

Another technique called AdaFedFR [35], dynamically adjust inter-class representations to handle class imbalance. Authors state that this technique enhances generalization of face model and improve the efficiency of federated training under strict privacy-preservation. They utilize feature representations of public identities as learnable negative knowledge to optimize the local objective within the feature space, which further encourages the local model to learn

powerful representations and optimize personalized models for clients. Experimental results on different datasets demonstrate that the AdaFedFR outperforms the existing technique (at the time of their publication) on several prevalent FR benchmarks within less than three communication rounds, showing its communication-friendliness and great efficiency.

In [23], authors proposed a FFR framework via intra-subject Self-supervised learning (FedFS), a novel FL architecture tailored to train personalized FR models without imposing subjects. FedFS comprises two crucial components, adaptive soft label construction, and intra-subject self-supervised learning, that leverage aggregated features of the local and global models to cooperate with representations of an off-the-shelf model. Additionally, they also introduced a regularization loss to prevent overfitting and ensure the stability of the optimized model. The experiments of FedFS are conducted on different dataset, demonstrating superior performance compared to previous methods.

Table 3. Summary of Algorithmic innovations techniques.

Paper	Dataset(s) Used	Metrics	FL Framework	Key Contributions
[32]	LFW [18], CFP-FP [36] CASIA-WebFace [50], AgeDB-30 [31], SLLFW [25], CPLFW [54], CALFW [55], VGG2-FP [7], MegaFace [22], IJB-B [47]	Accuracy, Robustness	FedAvg with Gradient Correction	Proposed gradient correction to improve model robustness in non-IID data settings
[35]	MS-Celeb-1M [15], IJB-B [47], IJB-C [29], LFW [18], CFP-FP [36], AgeDB [31].	Accuracy, Inter-Class Representation	AdaFedFR	Proposes adaptive inter-class representation learning for FFR
[23]	DigiFace-1M [4] VGGFace [7]	Accuracy, Self-Supervised Learning Metrics	FedAvg with Self-Supervision	Combines self-supervised learning for FFR

An overview of algorithmic innovation techniques is provided in Table 3. The primary obstacles to these techniques include the difficulty of managing non-independent and identically distributed (non-IID) data, the enhancement of model convergence in distributed settings, and the resolution of class imbalance in FR datasets.

2.3 Fairness and Explainability

The following section provides an explanation of papers that discuss the explainability, fairness trade-offs, and biases in FFR models. Techniques of fairness and explainability guarantee that FFR systems are impartial and transparent. In an effort to mitigate bias in model predictions, methodologies such as adaptive federated learning with actor-critic selection (Privacy Safeguarding for Human-Computer Interaction) are applied. Model decisions are rendered interpretable

through the integration of explainable AI (XAI) methodologies, which is indispensable for healthcare and law enforcement applications. Nevertheless, the heterogeneity of data across clients and the difficulty of auditing decentralized models have made it difficult to achieve fairness in FL-based FR. A selected important techniques that elucidate fairness and explainability are as follows.

In [10], a project is initiated to the development of systems that can guarantee trustworthiness by delivering privacy, fairness, and explainability through design. To guarantee the privacy of the individual, they have implemented homomorphic encryption. Without sacrificing the accuracy of the final models, this method achieves fairness and explainability. The FairFace dataset is employed to verify the assertions of the authors through experiments.

A HessianFree Model Agnostic Meta Learning (HF-MAML) was introduced in an FFR by another technique [11]. In comparison to the current FFR models, they demonstrated that the proposed HF-MAML achieves higher scores in verification tests. Particularly, the verification scores display the greatest improvement in heterogeneous data partitions. An embedding regularization term is included into the loss function in order to achieve a balance between personalization and the development of a global model that is effective. It has been demonstrated that the performance of global model verification is enhanced when this term is combined with HF-MAML. HF-MAML and its embedding regularization extension are also employed to conduct a fairness analysis, which demonstrates that the standard deviation over the client evaluation scores can be reduced to enhance fairness [17].

The BUPT-Balancedface dataset, which is racially balanced, was used to train a deep learning FFR model in [51]. The private variant of the model was modified to incorporate differential privacy to ensure data confidentiality, while maintaining the emphasis on fairness. The accuracy, fairness, and privacy of both private and non-private models have been compared. The experimental findings suggest that the differential privacy reduces accuracy across demographics in an uneven manner and that adjusting the privacy budget enables a more appropriate balance of accuracy, fairness, and privacy. Furthermore, they conducted experiments to examine real-world bias by training the private model on the imbalanced CASIA-WebFace dataset. This dataset amplifies variability in accuracy and fairness disparities, demonstrating the influence of dataset composition on the interplay between privacy, accuracy, and fairness.

Table 4. Summary of techniques favoring fairness and explainability

Paper	Dataset(s) Used	Metrics	FL Framework	Key Contributions
[10]	FairFace [21]	Accuracy, Fairness, Privacy Metrics	Custom FL Framework	Combines privacy, fairness, and explainability in FFR
[11,17]	CelebA [53]	Accuracy, Imbalanced Data Metrics	Meta-Learning FL Framework	Addresses imbalanced data in FFR using meta-learning
[51]	RFW [44], BUPT-Balancedface dataset [43]	Accuracy, Privacy, Fairness Metrics	Custom FL Framework	Explores trade-offs between privacy, accuracy, and fairness in FFR

Table 4 provides a summary of the techniques for this category. The technique's primary obstacles in this category include, but are not limited to, guaranteeing equity among diverse demographic groups. Balancing fairness with model performance and ensuring that FFR systems are transparent and interpretable.

2.4 Domain Adaptation and Transfer Learning

FFR models are required to generalize across various domains, such as varying lighting conditions and demographics. Papers in this category concentrate on domain adaptation and transfer learning as a means of enhancing the performance of models in a variety of environments. Domain adaptation and transfer learning techniques improve the performance of FFR models in new environments. An efficient framework for conducting FL on AIoT devices was proposed by the authors in [8]. The method is particularly well-suited for AIoT devices due to its substantial reduction in the necessary computational cost and communication rounds. They demonstrate that the method has a negligible loss in utility when compared to its fully centralized counterpart through experiments on real-world datasets.

The technique in [56] enables models to adapt to new domains in the absence of labeled data. Ensemble models enhance transferability by integrating multiple models. In real-world scenarios with diverse data distributions, these techniques are indispensable for the deployment of FL-based FR systems. FedFR, a novel federated unsupervised domain adaptation approach for FR, was introduced in this paper. FedFR initially improves a hierarchical clustering algorithm to produce pseudo-labels for the target domain by utilizing the model that was pretrained in the source domain. Then, FedFR performs federated learning across domains by utilizing labeled data in the source domain and unlabeled data with pseudo labels in the target domain. The effectiveness and significance of FedFR are demonstrated through extensive experiments on different constructed benchmarks.

In [28], a novel approach to the integration of FFR with Electrostatic Discharging (ESD) testers is introduced. This integration employs the FL model with the pre-FedAvg algorithm, ensures the privacy and security of data by employing a distributed learning approach that preserves the high accuracy and efficiency of the ESD testing systems. The proposed approach's performance is assessed through metrics such as accuracy, precision-recall, and F1 score. The results indicate that this method outperforms the conventional method while simultaneously safeguarding data from a variety of vulnerabilities. The electronic manufacturing industries, quality control departments, and safety monitoring sections are among the key areas of this strategy.

Table 5 illustrates the summary of these techniques. Adapting models to another domains without compromising privacy, ensuring robustness in cross-domain FR tasks, and managing domain shifts in distributed environments are the primary challenges of these techniques.

Table 5. Summary of domain adaptation and transfer learning techniques

Paper	Dataset(s) Used	Metrics	FL Framework	Key Contributions
[8]	MS Celeb-1M [15], Private dataset	Accuracy, Latency, Communication Efficiency	Industrial FL Framework	Develops an FL framework for AIoT applications, focusing on FR in industrial settings
[56]	LFW [18], CelebA [53], CASIA-WebFace [50]	Accuracy, Domain Adaptation Metrics	FedAvg	Proposed unsupervised domain adaptation for FFR settings
[28]	Custom Dataset	Accuracy, Precision-recall, F1 score, Privacy metrics	Custom FL Framework	Integrates FL with ESD testing for privacy-preserving FR

2.5 Efficiency and Scalability

The computational and communication challenges of FFR are addressed by efficiency and scalability techniques. In resource-constrained environments, such as edge devices and IoT systems, these techniques are essential for the development of real-time FR applications. [24] introduces the initial work that employs FL for FR tasks. The authors of [2] proposed an FFR framework known as FedFace. FedFace employs the face images available on multiple clients to develop a generalizable and precise FR model. The face images stored at each client are not shared with other clients or the central host, and each client is a mobile device that contains the face images of the owner of the device (one identity per client). The experiments demonstrate the efficacy of FedFace in improving the verification performance of pre-trained face recognition systems on a variety of face verification benchmarks.

In [6], a decentralized FFR solution design and development are introduced. In this case, the devices that are used to capture the images lack the capacity to train models that can achieve a high level of accuracy on large amounts of data. Consequently, a method is proposed that performs the local detection of human faces and the extraction of characteristics using a pre-trained *FaceNet* model. These characteristics are transmitted to a robust processing unit, where a global model is developed and subsequently returned to the clients for recognition. The results of the experiment indicate that the solution is highly accurate and efficient.

Another approach recommends the use of FL as a method to safeguard the confidentiality of facial image data on edge devices, particularly in recognition systems [39]. The method places a substantial emphasis on decentralized training, thereby eliminating the need to transmit unprocessed image data to centralized servers. A secure aggregator consolidates the local models into a federated model, which is subsequently transmitted to individual devices via the primary server. The method proposed in this study offers two primary advantages. Initially, it guarantees the confidentiality of facial images by storing the original data on individual devices. Secondly, empirical evidence indicates that the federated model outperforms individual models in terms of average Equal Error Rate (EER) when a secure aggregator is not present.

Table 6. Summary of techniques highlighting efficiency and scalability

Paper	Dataset(s) Used	Metrics	FL Framework	Key Contributions
[2]	LFW [18], IJB-b [47], IJB-C [29]	Accuracy, Scalability	FedFace	Introduces FedFace, a framework for collaborative FR with improved scalability
[6]	LFW [18]	Accuracy, Decentralization Metrics	Decentralized FL Framework	Explores decentralized FL for FR, reducing reliance on a central server
[39]	LFW [18], IJB-C [29]	Accuracy, Security Metrics	Supervised FL Framework	Enhances security and efficiency in FFR using supervised learning

The summary of these techniques are given in Table 6. The main challenges with these techniques include, reducing communication overhead in FL, scaling FL-based FR systems to large datasets, ensuring real-time performance in industrial applications, among others.

3 Discussion and Future Research

The concept of trustworthy learning is gaining attention as FL advanced and continues to evolve [52]. Privacy, fairness, and explainability are needed to make FL systems ethical, transparent, and effective. This defines trustworthy learning. FL requires privacy to store sensitive data locally, preventing breaches. However, privacy is not enough. Fairness is also important because it prevents models from showing biases against certain groups or individuals. To address this issue, bias mitigation strategies and fairness-aware learning algorithms are being developed [19]. Explainability involves users' understanding of the models' decision-making process. To build trust and ensure model predictions can be interpreted and verified, this is necessary. To create effective, ethical, and trustworthy FL systems, researchers prioritize privacy, fairness, and explainability [5].

Due to device data distribution disparities, FFR data imbalance is common. Data that is not independent and identically distributed can affect model performance and generalization. Meta-learning, or "learning to learn," introduced to mitigate this problem. Model-agnostic meta-learning (MAML) and meta-gradient descent are being studied to improve FL model adaptability and robustness [3]. Meta-learning lets researchers create FL systems that manage data imbalance and perform well across devices and environments.

FFR is vulnerable to Sybil attacks and model poisoning due to its decentralization, making security a priority. Sybil attacks involve malicious entities creating many fake identities to compromise the global model and manipulate training. Researchers are developing Sybil-aware FL techniques to detect and mitigate Sybil attacks [13]. These methods use secure multi-party computation, robust aggregation, and anomaly detection algorithms to protect model updates. Sybil-aware FL improves FL system security to create more resilient and trustworthy models that can withstand malicious attacks and maintain performance and reliability.

Many critical areas will be the focus of future FFR research. Initially, FFR models require advanced methodologies to ensure impartiality and reduce biases

[33]. Second, FL systems must improve their scalability and efficiency to become widely adopted. This involves studying distributed computing, efficient communication protocols, and model compression [40]. Third, FFR models must be more explainable to build trust and transparency. Researchers are investigating ways to make FL model decision-making more understandable for users. Finally, FL security issues like Sybil attacks and model poisoning must be addressed. Advanced encryption, secure aggregation, and robust anomaly detection algorithms are being developed to improve FL system security and integrity.

4 Conclusion

Federated learning offers a robust solution to privacy concerns in face recognition. The reviewed papers demonstrate significant advancements in privacy preservation, algorithmic innovation, fairness and explainability, domain adaptation and transfer learning, and efficiency and scalability concepts in federated face recognition. While federated face recognition presents numerous advantages, it also comes with its own set of challenges. These include issues related to communication efficiency, data heterogeneity, and the need for robust security measures to prevent different attacks such as, model poisoning and Sybil attacks. Addressing these challenges requires ongoing research and innovation. Hence, future research on this area will be focus on integrating fairness, explainability, and addressing challenges such as data heterogeneity and communication efficiency. This survey aims to provide a comprehensive overview of the current state of research in this area, highlighting key contributions, challenges, and future directions. By doing so, we hope to inspire further research and innovation in the development of federated learning based face recognition systems.

Acknowledgments. 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).

Disclosure of Interests. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Abouelnaga, Y., Ali, O.S., Rady, H., Moustafa, M.: Cifar-10: KNN-based ensemble of classifiers. In: 2016 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 1192–1195. IEEE (2016)
2. Aggarwal, D., Zhou, J., Jain, A.K.: Fedface: collaborative learning of face recognition model. In: 2021 IEEE International Joint Conference on Biometrics (IJCB), pp. 1–8. IEEE (2021)
3. Aramoon, O., Chen, P.Y., Qu, G., Tian, Y.: Meta-federated learning. In: Federated Learning, pp. 161–179. Elsevier (2024)

4. Bae, G., et al.: 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 (2023)
5. Bárcena, J., Ducange, P., Marcelloni, F., Renda, A.: Increasing trust in AI through privacy preservation and model explainability: federated learning of fuzzy regression trees. Inf. Fusion **113**, 102598 (2025)
6. Brănescu, I., Ciobanu, R.I., Dobre, C., Mavromoustakis, C.: Decentralized machine learning for face recognition. In: 2023 22nd International Symposium on Parallel and Distributed Computing (ISPDC), pp. 1–8. IEEE (2023)
7. Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A.: 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 (2018)
8. Ding, Y., et al.: An efficient industrial federated learning framework for AIoT: a face recognition application. arXiv preprint arXiv:2206.13398 (2022)
9. Dritsas, E., Trigka, M.: Federated learning for IoT: a survey of techniques, challenges, and applications. J. Sens. Actuator Netw. **14**(1), 9 (2025)
10. Franco, D., Oneto, L., Navarin, N., Anguita, D.: Toward learning trustworthily from data combining privacy, fairness, and explainability: an application to face recognition. Entropy **23**(8), 1047 (2021)
11. Gansekoele, A., Hess, E., Bhulai, S.: Meta-learning for federated face recognition in imbalanced data regimes. In: 2024 2nd International Conference on Federated Learning Technologies and Applications (FLTA), pp. 24–31. IEEE (2024)
12. GDPR, G.: General data protection regulation. Regulation (EU) **679** (2016)
13. Ghafourian, M., Fierrez, J., Vera-Rodriguez, R., Tolosana, R., Morales, A.: Safl: sybil-aware federated learning with application to face recognition. In: 2024 IEEE International Conference on Image Processing Challenges and Workshops (ICIPCW), pp. 4050–4056. IEEE (2024)
14. Guo, G., Zhang, N.: A survey on deep learning based face recognition. Comput. Vis. Image Underst. **189**, 102805 (2019)
15. Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J.: MS-Celeb-1M: a dataset and benchmark for large-scale face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 87–102. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_6
16. Hasan, M.R., Guest, R., Deravi, F.: Presentation-level privacy protection techniques for automated face recognition–a survey. ACM Comput. Surv. **55**(13s), 1–27 (2023)
17. Hess, E.: The application and analysis of meta-learning in federated face recognition (2023)
18. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: 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 (2008)
19. Huang, W., et al.: Federated learning for generalization, robustness, fairness: a survey and benchmark. IEEE Trans. Pattern Anal. Mach. Intell. (2024)
20. Jiang, B., Liu, Z., Yue, G., Cui, X., Liu, Y., Wang, H.H.: Privacy safeguarding for human-computer interaction based on adaptive federated learning with actor-critic selection. IEEE Trans. Consum. Electron. (2025)
21. Kärkkäinen, K., Joo, J.: Fairface: face attribute dataset for balanced race, gender, and age. arXiv preprint arXiv:1908.04913 (2019)
22. Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4873–4882 (2016)

23. Kim, H., Choi, H., Kwak, Y.: Federated learning for face recognition via intra-subject self-supervised learning. arXiv preprint arXiv:2407.16289 (2024)
24. Kim, J., Park, T., Kim, H., Kim, S.: Federated learning for face recognition. In: 2021 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–2. IEEE (2021)
25. Knoche, M., Hormann, S., Rigoll, G.: Cross-quality LFW: a database for analyzing cross-resolution image face recognition in unconstrained environments. In: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), pp. 1–5. IEEE (2021)
26. Kumarage, B., Harankahadeniya, H., Induranga, A., Perera, I., Gunasekera, K.: Use of federated learning for personal smart media cloud solutions. In: 2024 Moratuwa Engineering Research Conference (MERCon), pp. 342–347. IEEE (2024)
27. Laishram, L., Shaheryar, M., Lee, J.T., Jung, S.K.: Toward a privacy-preserving face recognition system: a survey of leakages and solutions. ACM Comput. Surv. **57**(6), 1–38 (2025)
28. Lenin, S., Srivatsan, G., Basha, G.A., Aswin, Z.: Privacy-preserving integration of face recognition system and ESD tester using federated learning. In: 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), pp. 1246–1250. IEEE (2023)
29. Maze, B., et al.: Iarpa janus benchmark-c: face dataset and protocol. In: 2018 International Conference on Biometrics (ICB), pp. 158–165. IEEE (2018)
30. Meng, Q., Zhou, F., Ren, H., Feng, T., Liu, G., Lin, Y.: Improving federated learning face recognition via privacy-agnostic clusters. arXiv preprint arXiv:2201.12467 (2022)
31. Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., Zafeiriou, S.: 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 (2017)
32. Niu, Y., Deng, W.: Federated learning for face recognition with gradient correction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 1999–2007 (2022)
33. Panda, S., Dubey, R., Jena, B., Pareek, V., Tsai, L.W., Saxena, S.: Federated learning frameworks in smart e-healthcare: a systematic literature review with bias evaluation. Appl. Soft Comput. 112747 (2025)
34. de Paula, D., Alexandre, L.A.: Facial emotion recognition for sentiment analysis of social media data. In: Iberian Conference on Pattern Recognition and Image Analysis, pp. 207–217. Springer, Cham (2022)
35. Qiu, D., Lin, X., Wang, K., Chu, X., Yan, P.: Adafedfr: federated face recognition with adaptive inter-class representation learning. arXiv preprint arXiv:2405.13467 (2024)
36. Sengupta, S., Chen, J.C., Castillo, C., Patel, V.M., Chellappa, R., Jacobs, D.W.: Frontal to profile face verification in the wild. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–9. IEEE (2016)
37. Simmler, M., Canova, G.: Facial recognition technology in law enforcement: regulating data analysis of another kind. Comput. Law Secur. Rev. **56**, 106092 (2025)
38. Solomon, E., Woubie, A.: Federated learning method for preserving privacy in face recognition system. arXiv preprint arXiv:2403.05344 (2024)
39. Solomon, E., Woubie, A., Emiru, E.S., Abdelzaher, A.F.: Secure and efficient face recognition via supervised federated learning. In: 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN), pp. 291–296. IEEE (2024)

40. Soudan, B., Abbas, S., Kubba, A., Abu Waraga, O., Abu Talib, M., Nasir, Q.: Scalability and performance evaluation of federated learning frameworks: a comparative analysis. Int. J. Mach. Learn. Cybern. 1–15 (2025)
41. Truong, N., Sun, K., Wang, S., Guitton, F., Guo, Y.: Privacy preservation in federated learning: an insightful survey from the GDPR perspective. Comput. Secur. **110**, 102402 (2021)
42. Wang, L., Zhang, R.: Framework for facial recognition and reconstruction for enhanced security and surveillance monitoring using 3D computer vision. J. Electron. Imaging **32**(4), 042108 (2023)
43. Wang, M., Deng, W.: 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 (2020)
44. Wang, M., Deng, W., Hu, J., Tao, X., Huang, Y.: 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 (2019)
45. Wang, W., Li, S.: 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 (2024)
46. Weiss, G.M.: 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 (2019)
47. Whitelam, C., et al.: Iarpa janus benchmark-b face dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 90–98 (2017)
48. Woubie, A., Solomon, E., Attieh, J.: Maintaining privacy in face recognition using federated learning method. IEEE Access (2024)
49. Yang, J., Liu, J., Han, R., Wu, J.: Transferable face image privacy protection based on federated learning and ensemble models. Complex Intell. Syst. **7**(5), 2299–2315 (2021). https://doi.org/10.1007/s40747-021-00399-6
50. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)
51. Zarei, A., Hassanpour, A., Raja, K.: On privacy, accuracy, and fairness trade-offs in facial recognition. IEEE Access (2025)
52. Zhang, Y., et al.: A survey of trustworthy federated learning: issues, solutions, and challenges. ACM Trans. Intell. Syst. Technol. **15**(6), 1–47 (2024)
53. Zhang, Y., et al.: CelebA-spoof: large-scale face anti-spoofing dataset with rich annotations. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12357, pp. 70–85. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58610-2_5
54. Zheng, T., Deng, W.: Cross-pose LFW: a database for studying cross-pose face recognition in unconstrained environments. Beijing University of Posts and Telecommunications, Technical report **5**(7), 5 (2018)
55. Zheng, T., Deng, W., Hu, J.: Cross-age LFW: a database for studying cross-age face recognition in unconstrained environments. arXiv preprint arXiv:1708.08197 (2017)
56. Zhuang, W., Gan, X., Wen, Y., Zhang, X., Zhang, S., Yi, S.: Towards unsupervised domain adaptation for deep face recognition under privacy constraints via federated learning. arXiv preprint arXiv:2105.07606 (2021)