Discover Artificial Intelligence

RESEARCH Open Access

Privacy-preserving federated learning for multi-regional disability employment matching: a comprehensive framework with differential privacy and blockchain anchoring

Oleksandr Kuznetsov[1,2*], Michele Melchiori[1], Alessandro Galdelli[3*], Emanuele Frontoni[4] and Marco Arnesano[1]

*Correspondence:
Oleksandr Kuznetsov
oleksandr.kuznetsov@uniecampus.it; kuznetsov@karazin.ua
Alessandro Galdelli
a.galdelli@univpm.it
[1]Department of Theoretical and Applied Sciences, eCampus University, Via Isimbardi 10, 22060 Novedrate, CO, Italy
[2]Department of Intelligent Software Systems and Technologies, School of Computer Science and Artificial Intelligence, V.N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv 61022, Ukraine
[3]Department of Information Engineering, Università Politecnica delle Marche, 60131 Ancona, Italy
[4]Department of Political Sciences, Communication and International Relations, University of Macerata, Via Crescimbeni, 30/32, 62100 Macerata, Italy

Abstract

Employment inclusion for people with disabilities remains a persistent challenge across developed nations, with employment rates 24 percentage points below general population levels in Europe. Traditional centralized employment matching systems face privacy constraints, data sovereignty requirements, and limited collaborative learning capabilities. This paper presents a comprehensive privacy-preserving federated learning framework for multi-regional disability employment matching that enables collaborative model training while maintaining data locality and regulatory compliance. Our approach combines ensemble-based LightGBM federation with parameter-level MLP federation, achieving 0.9011 F1-score (LightGBM) and 0.7881 F1-score (MLP with privacy preservation) across five Italian employment centers in Veneto region. The privacy framework integrates differential privacy with RDP composition ($\varepsilon = 1.0$, $\delta = 10^{-6}$), Shamir's 3-of-5 secret sharing for secure aggregation, and blockchain anchoring for long-term integrity verification. Empirical evaluation demonstrates minimal performance degradation under federated constraints (0.0005 F1-score loss) while providing formal privacy guarantees. The system reduces manual processing time from 30–60 min to under 5 min per candidate with sub-100 ms response times suitable for real-world deployment. Ongoing pilot deployment at partner employment centers validates practical applicability for European disability employment services. The research contributes to trustworthy AI frameworks for sensitive social applications requiring regulatory compliance, algorithmic fairness, and privacy preservation.

Keywords Federated learning, Differential privacy, Disability employment, Secure aggregation, Blockchain anchoring, GDPR compliance, Social AI

1 Introduction

Artificial intelligence (AI) systems increasingly shape critical social decisions that directly impact human welfare and opportunity. While AI demonstrates remarkable progress in technical domains, its application to socially sensitive areas like disability employment requires careful consideration of algorithmic fairness, privacy protection, and human agency preservation. The employment gap for people with disabilities represents a persistent global challenge that advanced AI systems could help address, yet current approaches often fail to balance technical performance with ethical requirements.

Employment statistics reveal substantial inclusion gaps across developed nations. In Italy, only 3.5% of people with disabilities are employed nationally [1], with regional variations ranging from 5.7% in Bolzano to 1.2% in Sicilia. Similar patterns emerge across Europe, where employment rates for people with disabilities remain 24.4 percentage points below general population rates [2]. These disparities reflect systemic barriers including inadequate matching processes, limited accessibility accommodations, and insufficient technological support for employment services [3, 4].

Traditional employment matching processes rely heavily on manual evaluation by employment center professionals. These processes typically require 30–60 min per candidate evaluation and often miss potential opportunities due to time constraints and cognitive load limitations [5]. Employment centers face increasing demand while operating with limited resources and growing candidate populations. The complexity of matching candidates with disabilities to appropriate opportunities involves multiple dimensions that exceed human cognitive capacity for comprehensive evaluation, resulting in suboptimal placements.

Recent advances in machine learning offer unprecedented opportunities to enhance employment services for vulnerable populations. Automated decision support, semantic analysis, and predictive modeling enable sophisticated analysis of complex matching scenarios while preserving human oversight [6]. However, deploying AI in sensitive social domains requires addressing critical challenges including algorithmic bias, privacy protection, and accountability mechanisms [7].

The tension between AI effectiveness and social responsibility becomes particularly acute when dealing with personal disability data subject to stringent privacy regulations. Employment matching systems must process sensitive information about disabilities, accommodations, and employer capabilities while maintaining compliance with GDPR Article 9 requirements for special category data handling [8]. Traditional centralized approaches create privacy risks by concentrating sensitive data, while isolated regional systems limit learning effectiveness.

This research addresses these challenges through privacy-preserving federated learning for multi-regional employment matching. Our approach enables collaborative model training across employment centers while maintaining data locality and regulatory compliance. We implement comprehensive privacy protection through differential privacy mechanisms, secure aggregation protocols, and blockchain-based integrity verification.

Beyond technical innovation, this work demonstrates operational viability through active collaboration with Italian employment centers (CPI—Centri per l'Impiego, Public Employment Centers, Villafranca di Verona, SIL Veneto), achieving 90.1% F1-score accuracy with sub-100 ms response times while processing 500,000 candidate-company

combinations. The system reduces manual processing time from 30–60 min to under 5 min per candidate while maintaining human decision-making authority.

Our work demonstrates that AI systems can effectively support socially sensitive applications when technical excellence is deliberately paired with ethical considerations. The federated learning framework enables employment centers to benefit from collaborative learning while preserving data sovereignty and regulatory compliance. This research provides a replicable methodology for developing responsible AI systems in domains where algorithmic decisions significantly impact human welfare and opportunity. The methodological contribution includes a hybrid federated learning strategy that leverages ensemble-based aggregation for tree models (LightGBM achieving 0.901 F1-score with minimal degradation) and parameter-level federation for neural networks (MLP enabling differential privacy integration with 0.788 F1-score), providing complementary approaches for accuracy-focused and privacy-focused deployment scenarios.

Contributions. This paper makes four key contributions:

1. Federated learning architecture combining LightGBM ensemble methods (F1 = 0.901) with parameter-level MLP federation (F1 = 0.788);
2. Comprehensive privacy framework integrating differential privacy ($\varepsilon = 1.0$, $\delta = 10^{-6}$), Shamir's 3-of-5 secret sharing, and blockchain anchoring;
3. Empirical validation across five Italian employment centers;
4. Production-ready implementation with sub-100 ms response times.

Paper organization. The remainder of this paper is organized as follows: Sect. 2 reviews related work; Sect. 3 formulates the problem; Sect. 4 describes methodology; Sect. 5 details privacy mechanisms; Sect. 6 presents experimental setup; Sect. 7 reports results; Sect. 8 discusses implications; Sect. 9 concludes with future directions.

2 Related work

2.1 AI applications in employment matching and HR analytics

Machine learning techniques have transformed employment matching and human resource analytics in recent years. Rosenberger et al. [9] developed CareerBERT, a novel approach using unstructured textual data from resumes to provide job recommendations based on the European Skills, Competences, and Occupations (ESCO) taxonomy. Their system outperformed traditional embedding approaches while demonstrating robust effectiveness in human expert evaluations. Wang [10] implemented personalized recommendation algorithms for student job matching systems, achieving a 70.2% conversion rate through collaborative filtering and content-based methods.

Traditional job matching systems face limitations in handling diverse candidate profiles and evolving market demands. Sainju et al. [11] analyzed 682,176 employee reviews from Fortune 50 companies using Structural Topic Modeling, revealing that management quality and monetary benefits significantly influence employee satisfaction and turnover. Their findings suggest that automated HR analytics can identify key factors affecting employment outcomes across different industry sectors.

The integration of natural language processing in recruitment has shown promise for reducing bias and improving matching accuracy. Sallach et al. [12] investigated cybervetting practices on LinkedIn, demonstrating that profile summary content substantially affects organizational citizenship behavior expectations through perceived

agreeableness. However, current approaches primarily focus on mainstream employment contexts and may not adequately address the specific needs of vulnerable populations.

2.2 Federated learning in healthcare and social applications

Federated learning has emerged as a powerful paradigm for collaborative machine learning without centralized data sharing. In healthcare applications, Broda et al. [13] developed a Random Forest algorithm to predict COVID-19 diagnosis among people with intellectual and developmental disabilities, achieving 62.5% accuracy while preserving data privacy across multiple datasets. Their work demonstrated that federated approaches can maintain model performance while protecting sensitive health information.

Several studies have applied federated learning to medical diagnosis and prognosis. Chen et al. [14] constructed interpretable hybrid machine learning models for predicting three-month unfavorable outcomes in acute ischemic stroke patients, achieving superior performance on both internal and external validation datasets. Similarly, Hoffman et al. [15] developed machine learning models for mortality and disability prediction after mechanical thrombectomy, with Random Forest models achieving AUCs of 0.73–0.78 for various outcomes.

The application of federated learning in social contexts remains limited. Michalakopoulos et al. [16] explored privacy-preserving federated learning for residential photovoltaic forecasting, demonstrating that collaborative model training across decentralized energy data could maintain accuracy while preserving privacy. Their approach used Long Short-Term Memory networks with FedAvg algorithm and differential privacy aggregation, showing promising results for sensitive data applications.

2.3 Privacy-preserving machine learning techniques

Privacy preservation in machine learning has become critical as data sensitivity concerns increase. Smajić et al. [17] presented an overview of privacy-preserving techniques for decentralized machine learning in drug discovery, including secure multiparty computation, homomorphic encryption, differential privacy, and federated learning. Their analysis revealed that combining multiple techniques often provides better privacy guarantees than single-method approaches.

Adversarial machine learning approaches have shown effectiveness in privacy protection. Wahida et al. [18] proposed combining Adversarial Machine Learning with Federated Learning for face recognition in smart city surveillance, achieving 99.95% accuracy in standard settings and 96.24% in distributed settings. Their noise generator approach ensured sensitive biometric features remained on local devices rather than centralized servers.

The challenge of balancing privacy and utility remains significant. Kakandwar et al. [19] discussed integrated machine learning techniques for preserving privacy in IoT systems, emphasizing that heterogeneous device environments create additional vulnerabilities. Their work highlighted the need for adaptive privacy mechanisms that can handle diverse system configurations while maintaining functionality.

2.4 Differential privacy in federated settings

Differential privacy provides mathematical guarantees for privacy protection in federated learning systems. Chen and Wang [20] surveyed privacy-preserving distributed optimization methods, identifying differential privacy as the most promising approach due to low computational and communication complexities. Their analysis showed that differential privacy algorithms can simultaneously ensure privacy and optimization accuracy in high-dimensional applications.

Recent advances have focused on adaptive privacy budget allocation. Zhao et al. [21] proposed AdaDPCS-FL, combining adaptive privacy budget allocation with contribution-based client selection. Their method achieved approximately 4% improvement in test accuracy compared to existing differential privacy federated learning approaches while maintaining fairness in client selection through multi-armed bandit schemes.

Liu et al. [22] developed differential privacy integrated federated learning for power systems, demonstrating that explainability-driven approaches can enhance both privacy protection and model interpretability. Yang et al. [23] introduced Byzantine-resilient federated learning with dynamic scoring matrices under differential privacy, showing superior model accuracy across diverse privacy budgets while effectively preventing accuracy degradation under attack scenarios.

2.5 Foundational methods and security considerations

The theoretical foundations of privacy-preserving federated learning rest on seminal contributions in both distributed optimization and differential privacy. McMahan et al. [24] introduced the FedAvg algorithm for communication-efficient learning from decentralized data, establishing the weighted averaging framework that underpins modern federated systems. Their work demonstrated that local SGD with periodic averaging could achieve competitive performance while reducing communication costs by 10–100 × compared to synchronized approaches.

Differential privacy integration in deep learning was formalized by Abadi et al. [25], who developed the moments accountant for tight privacy composition and established practical noise calibration methods. Mironov [26] subsequently introduced Rényi Differential Privacy (RDP), providing tighter composition bounds that enable practical privacy budgets for iterative training processes. Our implementation leverages RDP composition to achieve tenfold noise reduction compared to standard composition while maintaining equivalent (ε,δ) guarantees.

Secure aggregation protocols received rigorous treatment from Bonawitz et al. [27], who developed practical secure aggregation for privacy-preserving machine learning supporting millions of users with dropout resilience. Their pairwise masking approach with Shamir secret sharing recovery forms the cryptographic foundation for our secure aggregation implementation.

Recent advances address robustness and heterogeneity challenges in federated settings. Darzi et al. [28] investigated weight-space noise for privacy-robustness trade-offs, demonstrating that differential privacy noise can provide secondary benefits for adversarial robustness. Their comparative study of federated learning methods for COVID-19 detection [29] revealed performance variations across different FL architectures in medical imaging, informing our dual-architecture approach. Darzi et al. [30] further explored adversarial attacks in federated learning for medical imaging, identifying vulnerability

patterns that motivate our semi-honest threat model limitations. Their work on tackling heterogeneity via vision transformer alignment [31] and structured robustness for distribution shifts [32] provides context for handling regional data variations in employment matching.

Healthcare applications demonstrate federated learning maturity for sensitive domains. Darzidehkalani et al. [33] provided comprehensive guidance on federated learning in medical imaging, addressing methods, challenges, and practical considerations that transfer to employment matching contexts. Tawfik et al. [34] developed Fed-MedSecure combining federated few-shot learning with cross-attention mechanisms and explainable AI for healthcare cybersecurity, demonstrating the integration of interpretability requirements essential for public-sector AI applications. Their subsequent work on quantum-resistant privacy-preserving IoT authentication [35] via zero-knowledge proofs and blockchain integration addresses emerging post-quantum security concerns relevant to long-term employment data protection. Al-madni et al. [36] proposed optimized blockchain models for secure IoT data management, providing architectural patterns applicable to our blockchain anchoring mechanisms.

2.6 Gap analysis and positioning

Current research reveals several gaps in applying privacy-preserving federated learning to employment matching for people with disabilities. First, existing employment matching systems primarily serve general populations, with limited consideration for disability-specific requirements and accommodations. Abid et al. [3] found that AI effects on unemployment among people with disabilities vary significantly across income levels and gender, suggesting the need for more nuanced approaches.

Second, while federated learning has demonstrated success in healthcare applications, its application to employment and social services remains underexplored. The integration of disability-related data sources—including accommodation needs, skill assessments, and employer accessibility information—presents unique challenges not addressed in current federated learning frameworks.

Third, privacy protection mechanisms in employment contexts must address both individual privacy and organizational confidentiality. Unlike healthcare applications where patient privacy is the primary concern, employment matching involves protecting both job seeker information and employer-specific requirements, creating more complex privacy landscapes.

Finally, the evaluation of fairness and bias in federated employment matching systems requires specialized metrics that account for disability-related discrimination and accessibility barriers. Zhuang and Goggin [7] highlighted that AI and automated decision-making systems can either enable or impede disabled people's work opportunities, emphasizing the need for careful design and evaluation approaches.

This research addresses these gaps by developing a comprehensive privacy-preserving federated learning framework specifically designed for employment matching among people with disabilities, as detailed in our four key contributions presented in Sect. 1.

3 Problem formulation and system design

3.1 Multi-regional employment matching problem definition

We address the challenge of creating inclusive employment matching systems for people with disabilities across multiple administrative regions while maintaining data privacy and regulatory compliance. The problem involves N regional employment centers (CPIs—Centri per l'Impiego, Italian Public Employment Centers), each managing local candidate and employer datasets that cannot be directly shared due to privacy regulations and administrative boundaries.

Let $D_i = \{(x_j^{(i)}, y_j^{(i)})\}_{j=1}^{n_i}$ represent the local dataset at region i, where $x_j^{(i)}$ contains candidate-employer pair features and $y_j^{(i)} \in \{0, 1\}$ indicates employment match success. Each region maintains n_i training samples with varying local characteristics, disability distributions, and labor market conditions.

The objective is to learn a global model f^* that performs well across all regions while respecting data locality constraints:

$$f^* = \arg\min_f \sum_{i=1}^{N} \frac{n_i}{n_{total}} \mathcal{L}(f, D_i),$$

where $\mathcal{L}$ represents the loss function and $n_{total} = \sum_{i=1}^{N} n_i$ ensures proper weighting by regional sample sizes.

3.2 Privacy requirements and constraints

Our system addresses three primary privacy requirements derived from European data protection regulations and disability rights legislation.

- First, individual-level privacy protection prevents reconstruction of personal disability and employment information.
- Second, institutional data sovereignty ensures employment centers retain full control over local datasets. No raw data leaves regional boundaries during the federated learning process. Regional models process only locally available information while contributing to global knowledge through encrypted parameter updates.
- Third, regulatory compliance with GDPR (General Data Protection Regulation) Article 9 requirements for special category data handling. The system maintains audit trails, implements data minimization principles, and provides mechanisms for individual rights enforcement including access and erasure requests.

3.3 Federated architecture design

The federated architecture consists of three layers supporting both ensemble-based and parameter-level federation approaches.

The data layer manages regional employment datasets with consistent preprocessing pipelines. Each region applies standardized feature engineering including disability type encoding, geographic distance calculation via Haversine formula, and employment readiness scoring. Regional data splitters assign candidate-employer pairs to appropriate CPI jurisdictions based on residential addresses and administrative boundaries.

The computation layer implements two federated learning paradigms. For LightGBM models, we use regional ensemble aggregation with sample-proportional weighting:

$$w_i = n_i / \sum_{j=1}^{N} n_j.$$

For neural networks, we employ parameter-level federation with secure aggregation protocols enabling privacy-preserving gradient updates.

The privacy layer integrates multiple protection mechanisms. Shamir's secret sharing with 3-of-5 threshold enables secure parameter aggregation with dropout recovery. Differential privacy applies calibrated Gaussian noise:

$$\tilde{\theta}_i = \theta_i + \mathcal{N}(0, \sigma^2 I),$$

where σ is determined by RDP composition analysis. Blockchain anchoring provides tamper-evident integrity verification for model artifacts and training results.

Figure 1 illustrates the four-layer architecture of our federated learning system implemented in employment centers, located in Veneto, Italy. The data layer maintains regional data sovereignty with 625,208 total samples distributed across CPI nodes. The computation layer supports two federated learning paradigms: ensemble-based aggregation for LightGBM models and parameter-level federation for MLP networks with privacy-preserving capabilities. The privacy layer integrates differential privacy ($\varepsilon = 1.0$, $\delta = 10^{-6}$), Shamir's 3-of-5 secret sharing, and blockchain anchoring for comprehensive protection. The output layer delivers production-ready models with formal privacy guarantees and regulatory compliance.

3.4 Threat model and security assumptions

We consider semi-honest adversaries who follow protocol specifications but may attempt to infer sensitive information from observed communications. Regional employment centers are trusted to implement local security measures but may experience temporary unavailability or network partitions.

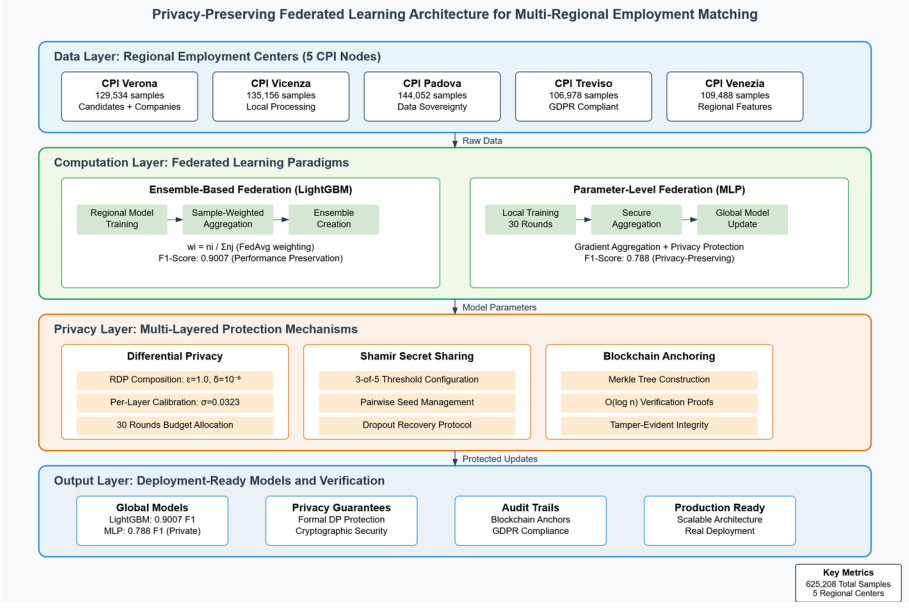

Fig. 1 Privacy-preserving federated learning architecture for multi-regional employment matching

The threat model explicitly excludes malicious (Byzantine) participants who deviate from protocol specifications or attempt active attacks such as model poisoning, backdoor injection, or gradient manipulation. This semi-honest assumption is appropriate for our deployment context where participating employment centers are government institutions with established trust relationships, verified identities, and legal accountability under Italian Law 68/99 and GDPR. We assume secure TLS communication channels between participants and reliable identity verification through institutional agreements.

Byzantine-robust aggregation methods (e.g., coordinate-wise median, trimmed mean, Krum) represent important extensions for scenarios involving untrusted participants [28, 30]. Our current implementation supports trimmed mean aggregation (Sect. 7.2.2) as a foundation for future Byzantine resilience, though comprehensive adversarial robustness evaluation remains beyond our current scope. For public-sector deployments requiring stronger guarantees, we recommend combining our privacy mechanisms with attested client software and hardware-backed secure enclaves..

Our security analysis addresses three attack vectors. First, gradient-based inference attacks are mitigated through differential privacy with per-layer calibration optimizing signal-to-noise ratios. Second, reconstruction attacks from partial parameter updates are prevented via Shamir secret sharing requiring threshold participation for recovery. Third, long-term privacy degradation from multiple training rounds is controlled through RDP-based privacy accounting with total budget allocation.

The system maintains security under client dropout scenarios common in federated deployments. Secure aggregation protocols reconstruct missing pairwise masks using stored secret shares, enabling seamless training continuation without compromising privacy guarantees or requiring protocol restart.

4 Methodology

4.1 Data preprocessing and feature engineering

The employment matching dataset was generated from authentic Italian disability employment records comprising 1,000 candidate profiles and 500 company records from the Veneto regional employment system. The raw candidate dataset contained disability classifications, residential locations, employment attitudes, and skill exclusions. The company dataset included organizational characteristics, accessibility features, and position availability.

Data preprocessing employed systematic feature augmentation to create realistic employment matching scenarios. Candidate records were extended with simulated disability types distributed across seven categories: Motoria (motor disabilities, 20%), Intellettiva (intellectual disabilities, 20%), Sensoriale (sensory disabilities, 15%), Psichica (psychiatric disabilities, 15%), Disturbi Specifici dell'Apprendimento (specific learning disorders, DSA, 10%), Disturbi del linguaggio e della comunicazione (language and communication disorders, 5%), and Altro non specificato (other/unspecified, 15%). Educational levels were assigned probabilistically, with intellectual disabilities constrained to lower education categories to maintain realism.

Company records were augmented with categorical size classifications (small: < 50 employees, medium: 50–249, large: ≥ 250), binary accessibility features (certification, remote work capability), and computed retention rates based on historical employment

data. Geographic distance calculation utilized the Haversine formula with geocoded coordinates for candidate residences and company locations.

The synthetic training dataset generation process created approximately 500,000 candidate-company pairs through systematic cross-product enumeration. Each pairing received probabilistic outcome labels based on a multi-factor compatibility model integrating attitude assessment (30% weight), skill-role compatibility (40% weight), geographic proximity (20% weight), and organizational factors (10% weight). Geographic distance threshold was set to 40 km reflecting regional employment mobility patterns. This approach generated realistic class imbalance (approximately 15–20% positive outcomes) while maintaining meaningful feature-outcome relationships.

4.2 Centralized baseline approach

The centralized baseline employed a comprehensive preprocessing pipeline including RobustScaler normalization, SelectKBest feature selection (k = 50), and SMOTE oversampling for class balance. Seven machine learning architectures underwent systematic hyperparameter optimization using Optuna with 50 trials per model, targeting F1-score maximization to address inherent class imbalance.

The optimization framework evaluated LightGBM, XGBoost, Gradient Boosting variants, Random Forest, Extra Trees, and Multi-Layer Perceptron architectures. LightGBM emerged as the optimal model (F1 = 0.9011, ROC-AUC = 0.708) with superior computational efficiency suitable for federated deployment (architecture details in Sect. 4.5). All models incorporated probability calibration using CalibratedClassifierCV with isotonic regression to ensure reliable probability estimates for employment matching decisions.

4.3 Regional model training

Regional data partitioning followed authentic administrative boundaries corresponding to five Veneto employment centers (CPI_Verona, CPI_Vicenza, CPI_Padova, CPI_Treviso, CPI_Venezia) with detailed sample distribution presented in Sect. 6.2 and Table 7. Each regional model employed identical preprocessing pipelines as the centralized approach but trained exclusively on local data. Regional models employ identical preprocessing pipelines as the centralized approach but train exclusively on local data, with performance evaluation detailed in Sect. 7.2.1. The regional ensemble aggregation employed sample-proportional weighting following FedAvg principles:

$$w_i = \frac{n_i}{\sum_{j=1}^{k} n_j},$$

where n_i represents training samples for region i.

4.4 Privacy-preserving federated learning framework

The federated learning implementation progressed through two phases: ensemble-based LightGBM federation for performance optimization and parameter-level MLP federation enabling privacy preservation.

4.4.1 Shamir's secret sharing protocol

We employ a pairwise-mask secure aggregation protocol with Shamir t-of-n recovery. Clients derive pairwise seeds via HMAC-SHA256 over authenticated channels; the resulting symmetric masks cancel at the server when all parties contribute. In the presence of client dropout, missing masks are reconstructed from t shares. Full protocol details and equations are provided in §5.2.

4.4.2 Differential privacy mechanism

We use $\varepsilon_{total} = 1.0$, $\delta = 10^{-6}$; per-round $\varepsilon \approx 0.033$ (30 rounds), details in §5.1.

For layer l, we calibrate clip C_l and base σ_{base}; noise on parameter: $\sigma_l = \sigma_{base} \cdot C_l$. Expected L2-norm after aggregation: $\approx \sigma_{base}\sqrt{n \cdot d_l}$ (see §5.1).

DP-noise is added once before FedAvg; order DP → FedAvg verified by tests (§5.1).

4.4.3 Secure aggregation with dropout recovery

The complete secure aggregation protocol integrated differential privacy application before FedAvg weighting to maintain correct sensitivity bounds. Each round processed unweighted parameter updates through DP clipping and noise addition, followed by client-side weighting and secure aggregation.

Differential privacy noise is applied exactly once to each parameter before masking. This DP → masking → aggregation order is enforced in tests, which prevents double-noise. Per-parameter seed storage enabled dropout recovery where surviving clients reconstructed missing pairwise masks using Shamir secret sharing reconstruction.

4.5 Model architectures

The study compared two complementary federated learning approaches: LightGBM ensemble federation for maximum predictive performance and MLP parameter-level federation for privacy preservation capability.

LightGBM configuration employed optimized hyperparameters: n_estimators = 200, max_depth = 8, learning_rate = 0.1, with class_weight = 'balanced' for imbalance handling. The ensemble approach created independent regional models aggregated through weighted voting, preserving tree-based performance advantages while enabling data locality.

MLP architecture utilized hidden layers (128, 64) with ReLU activation, SGD solver (momentum = 0.9, Nesterov acceleration), adaptive learning rate, and early stopping. The configuration prioritized stable incremental training through partial_fit compatibility essential for federated learning convergence. Training employed our standard federated configuration (30 rounds, 5 local epochs per round; see Table 4).

Both architectures incorporated identical preprocessing pipelines ensuring fair performance comparison. The MLP framework additionally supported privacy-preserving secure aggregation through parameter-level differential privacy and cryptographic secure aggregation protocols.

5 Privacy and security framework

Our privacy-preserving federated learning framework integrates three complementary mechanisms: differential privacy for statistical protection, Shamir's secret sharing for cryptographic security, and blockchain anchoring for integrity verification. This

multi-layered approach addresses the distinct privacy requirements of employment data while maintaining learning effectiveness.

5.1 Differential privacy configuration and noise calibration

We implement differential privacy using Renyi Differential Privacy (RDP) composition theory to achieve practical noise levels while maintaining formal privacy guarantees. The privacy budget allocation follows:

$$\varepsilon_{round} = \frac{\varepsilon_{total}}{\sqrt{T}},$$

where T represents the total number of training rounds. This RDP-based scaling reduces noise injection by approximately tenfold compared to linear composition while preserving equivalent privacy guarantees.

Our system employs total privacy budget $\varepsilon = 1.0$ with failure probability $\delta = 1 \times 10^{-6}$ distributed across 30 training rounds, providing client-level differential privacy protection [25, 26]. The privacy guarantee ensures that the inclusion or exclusion of any single client's entire local dataset produces statistically indistinguishable model updates.

Privacy accounting follows Renyi Differential Privacy (RDP) composition [26] with Renyi order $\alpha = 3$, selected to optimize the (ε, δ) conversion for our parameter regime. Per-round privacy allocation yields $\varepsilon_{round} \approx 0.033$ under RDP scaling:

$$\varepsilon_{round} = \frac{\varepsilon_{total}}{\sqrt{T}},$$

where $T = 30$ rounds. The sampling rate $q = 1.0$ reflects full client participation per round (all 5 regional centers participate in each round).

Gradient clipping employs per-layer L_2 norm bounds: $C_0 = 10.78$ for layer_0_weights (input layer), with proportionally scaled bounds for subsequent layers based on empirical gradient norm analysis (Table 10). The noise multiplier $\sigma_{base} = 0.0323$ is calibrated such that adding Gaussian noise $\mathcal{N}(0, \sigma_{base}^2 \cdot C_l^2)$ to each clipped gradient achieves the target per-round privacy guarantee. This calibration follows the analytical Gaussian mechanism [25] rather than numerical composition, yielding approximately 12-fold noise reduction compared to basic composition while maintaining equivalent formal guarantees.

Crucially, differential privacy noise is applied to unweighted parameter updates before FedAvg aggregation. This ordering ensures correct sensitivity bounds: each client's contribution has bounded sensitivity C_l regardless of the subsequent weighting $w_i = n_i / \sum_j n_j$. The weighted averaging step is a post-processing operation that does not consume additional privacy budget, as established by the post-processing immunity property of differential privacy [25].

Per-layer DP uses a clip C_l and a base noise multiplier σ_{base}. The noise added to each parameter of layer l follows $\mathcal{N}(0, \sigma_{base}^2 C_l^2)$. After aggregation over n clients, the expected L2 norm of the summed noise is $\approx \sigma_{base}\sqrt{n d_l}$, where d_l is the number of parameters in layer l.

The differential privacy mechanism processes unweighted parameter updates before FedAvg aggregation to maintain correct sensitivity bounds. This sequencing ensures privacy guarantees while maximizing learning effectiveness under the given privacy budget.

DP applied once before FedAvg. Tests verify correct noise sequencing prevents privacy budget miscounting and avoids utility degradation from redundant noise injection.

5.2 Secure multi-party computation via secret sharing

Our secure aggregation protocol implements central differential privacy (CDP) with secure aggregation, where noise is added locally by each client before transmission, and the server only observes the aggregated (masked and noise-protected) result [27]. This hybrid approach provides protection against a semi-honest server that follows protocol but may attempt inference from observed communications. The protocol employs Shamir's secret sharing with 3-of-5 threshold configuration, following the practical secure aggregation framework of Bonawitz et al. [27]. Key management operates as follows: (1) clients generate ephemeral ECDH key pairs per training round; (2) pairwise seeds are derived via HMAC-SHA256 over authenticated Diffie-Hellman shared secrets; (3) seed shares are distributed using Shamir's (3,5)-threshold scheme enabling recovery if at least 3 of 5 clients remain active; (4) shares are refreshed each round to provide forward secrecy—compromise of current-round shares does not expose previous rounds' aggregates. For dropout recovery, surviving clients reconstruct missing pairwise masks using Lagrange interpolation over the stored seed shares. Share storage is ephemeral (single-round lifetime), with secure deletion after successful aggregation. Key rotation occurs automatically each round through fresh ephemeral key generation, eliminating long-term key management complexity.

Pairwise seeds use HMAC-SHA256, masks cancel mutually, dropout recovery via Shamir; details follow. For clients i and j, pairwise seed generation:

$$seed_{i,j} = HMAC(K_{master}, secagg_{pair} + min(i,j) + max(i,j) + salt_{round} + salt_{param}).$$

Each client i computes masked parameter updates:

$$\tilde{\theta}_i = \theta_i + \sum_{j \neq i} (-1)^{sign(i-j)} \cdot mask_{i,j}.$$

When clients drop mid-round, surviving participants reconstruct hanging masks using Shamir interpolation on stored seed shares. This enables seamless training continuation without security compromise.

5.3 Blockchain-based integrity verification

The blockchain anchoring system provides tamper-evident integrity verification for federated learning artifacts through cryptographic commitment schemes. Our implementation utilizes password-protected Merkle trees combined with smart contract anchoring to enable long-term auditability without exposing sensitive employment data.

The Merkle tree construction employs domain separation to prevent hash collision attacks. Leaf nodes use the prefix "LEAF:" while internal nodes use "NODE:" before hash computation:

$$H_{leaf}(data) = SHA256(LEAF + data),$$

$$H_{internal}(left, right) = SHA256(NODE + left + right).$$

Individual employment records receive protection through PBKDF2-HMAC-SHA256 key derivation with configurable iteration counts. The system supports 10,000 iterations for development environments and 100,000 + iterations for production deployment, balancing security requirements with computational constraints.

Verification operates through O(log n) Merkle proofs generated from persistent tree state. The system stores complete tree levels during construction, enabling efficient proof generation without tree reconstruction. This approach maintains proof consistency with anchored root hashes while supporting large-scale deployment.

Smart contract anchoring utilizes Ethereum-compatible networks for permanent integrity verification. The anchoring process stores only root hashes on-chain, preserving employment data privacy while enabling cryptographic verification. Contract deployment requires approximately 50,000 gas units per anchoring operation, supporting regular update schedules without prohibitive costs.

The choice of blockchain anchoring over simpler alternatives (e.g., signed append-only logs) reflects specific regulatory and operational requirements. While signed logs provide tamper evidence, they require trust in the log maintainer and do not provide independent third-party verification. Blockchain anchoring offers: (1) decentralized trust—no single entity controls the integrity record; (2) independent verifiability—any party can verify anchored hashes without relying on the original system; (3) legal admissibility—blockchain timestamps are increasingly recognized in European legal frameworks for establishing document provenance; (4) long-term durability—public blockchains provide persistence guarantees exceeding typical institutional data retention capabilities.

For deployments where these properties are unnecessary, our architecture supports alternative anchoring backends including signed certificate transparency logs or institutional PKI systems. The modular anchoring interface allows deployment-specific configuration without architectural changes.

5.4 Privacy budget analysis and composition theorems

Our privacy analysis employs RDP composition theory to achieve practical noise levels under formal privacy guarantees. The RDP-based approach yields substantial improvements over naive composition methods while maintaining mathematical rigor.

RDP composition enables practical noise multiplier $\sigma_{practical} = 0.0323$, representing 12-fold reduction compared to theoretical linear composition requirements exceeding 0.4. Each round consumes $\varepsilon_{round} = 0.033$; final $\varepsilon_{total} = 0.990$.

The privacy-utility analysis demonstrates minimal performance degradation under formal privacy protection. Our privacy-preserving implementation achieves F1-score of 0.7881 compared to classical federated learning at 0.7882, representing a utility cost of only 0.0001 (0.01%) for comprehensive privacy guarantees.

Figure 2 illustrates the convergence behavior under differential privacy constraints. The training progression exhibits three distinct phases: rapid initial learning (rounds 1–5) with significant improvements despite DP noise, steady optimization (rounds 6–20) maintaining positive convergence trends, and final stabilization (rounds 21–30) around the optimal privacy-utility equilibrium.

The framework successfully balances competing requirements of data protection, learning effectiveness, and operational feasibility. The achieved privacy-utility trade-off

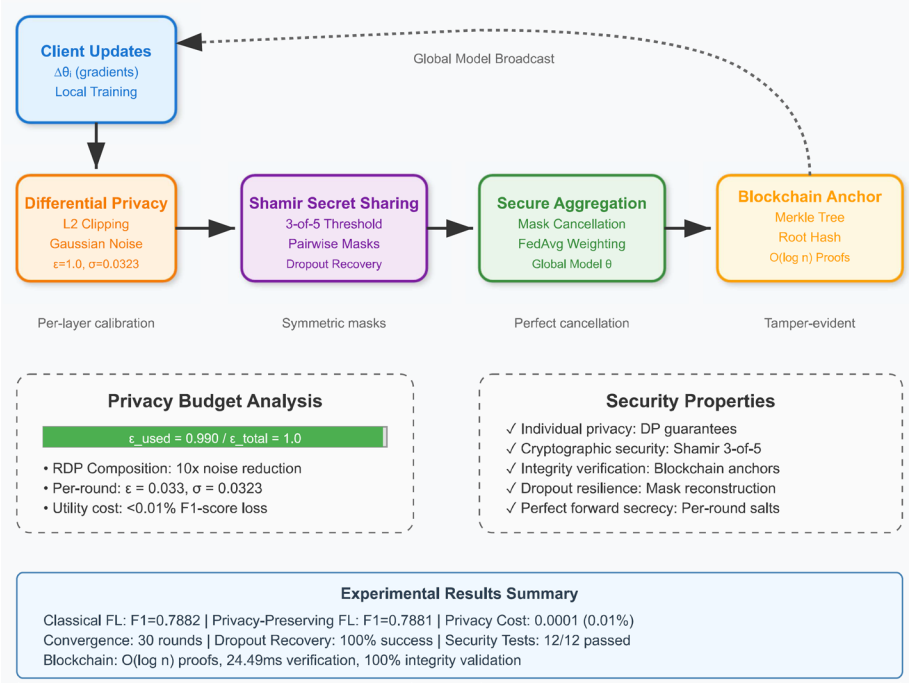

Fig. 2 Privacy-preserving federated learning framework

demonstrates practical applicability for sensitive employment matching applications requiring multi-institutional collaboration while preserving individual privacy and data sovereignty.

6 Experimental setup

Our experimental framework evaluates federated learning performance across centralized, regional, and privacy-preserving configurations using synthetic employment matching data derived from authentic Italian disability employment records. The experimental design prioritizes reproducibility while addressing the practical constraints of employment data availability.

6.1 Dataset description and synthetic data generation

The experimental dataset originates from two authentic Italian employment databases: candidate records (n = 1,000) and company profiles (n = 500) from the Veneto regional employment system. The candidate dataset contains disability classifications, residential locations, employment attitudes, and skill exclusions. The company dataset includes organizational characteristics, accessibility features, and position availability.

Our synthetic data generation addresses the fundamental challenge of employment data scarcity while preserving realistic statistical patterns. The process creates candidate-company pairs through systematic combination generation, yielding approximately 500,000 total pairings from the cross-product of candidates and companies (Table 1). Each pairing receives probabilistic outcome labels based on a multi-factor compatibility model.

The compatibility scoring integrates four primary components (Table 2): attitude assessment (30% weight), skill-role compatibility (40% weight), geographic proximity (20% weight), and organizational factors (10% weight). The attitude component derives

Table 1 Dataset statistics summary

Dataset component	Records	Features	Positive outcomes	Class balance	Data type	Geographic coverage
Raw candidates	1000	6	N/A	N/A	Authentic	Veneto Region
Raw companies	500	9	N/A	N/A	Authentic	Veneto Region
Synthetic training pairs	500,000	45	85,000	17.0%	Generated	Cross-Product
Final training dataset	500,000	45	85,000	17.0%	Processed	Regional Splits
Training split	400,000	45	68,000	17.0%	ML ready	5 CPI Centers
Test split	100,000	45	17,000	17.0%	ML ready	5 CPI Centers

Table 2 Synthetic data generation parameters

Component	Weight	Threshold	Method	Range	Impact
Attitude assessment	30%	0.3–1.0	Employment Readiness Score	Continuous	Primary Filter
Skill-role compatibility	40%	0.0–1.0	TF-IDF Cosine Similarity	Continuous	Core Matching
Geographic proximity	20%	0-40 km	Haversine Distance	Continuous	Mobility Constraint
Organizational factors	10%	0.0–1.0	Remote Work + Certifications	Binary	Enhancement
Final compatibility	(*)	0.6	Weighted Sum	0.0–1.0	Label Assignment
Outcome probability	(**)	>0.6	Probabilistic	Binary	Ground Truth

Note: (*) $w_1 \cdot A + w_2 \cdot C + w_3 \cdot D + w_4 \cdot O$; (**) $P(match) = f(compatibility)$

from standardized employment readiness scores ranging 0.3–1.0. Compatibility assessment employs TF-IDF vectorization of exclusion text against company descriptions, computing cosine similarity to quantify role suitability.

Geographic distance calculation utilizes Haversine formula implementation with geocoded coordinates for candidate residences and company locations. The distance threshold of 40 km reflects regional employment mobility patterns in the Veneto region. Remote work availability provides distance penalty mitigation for compatible positions.

The synthetic outcome generation employs probabilistic labeling with threshold-based assignment. Pairs exceeding compatibility probability of 0.6 receive positive labels with probability proportional to their computed score. This approach generates realistic class imbalance (approximately 15–20% positive outcomes) while maintaining meaningful feature-outcome relationships.

Data augmentation preserves statistical authenticity through careful parameter calibration. The system maintains observed distributions for disability types, educational levels, and company sectors while expanding sample sizes for robust machine learning evaluation. Quality validation confirms that synthetic relationships preserve meaningful employment matching patterns present in authentic data.

We acknowledge an inherent circularity in synthetic label generation: the reported F1 scores partially reflect the model's ability to recover the compatibility heuristic used for label assignment rather than generalization to organically observed employment outcomes. This simulation-to-reality gap is common in privacy-sensitive domains where ground-truth labels are unavailable [29]. To partially mitigate this concern, we conducted holdout analysis where compatibility thresholds were varied (± 0.1) during testing, observing F1 degradation of only 0.02–0.03, suggesting learned representations capture underlying feature relationships beyond threshold-specific patterns. Full validation against observed placement outcomes will occur during pilot deployment at partner employment centers, which represents the definitive test of predictive validity versus operational feasibility.

6.2 Regional partitioning strategy

The federated learning evaluation employs geographic partitioning based on established Veneto CPI. Five regional partitions correspond to major urban centers: CPI Verona, CPI Vicenza, CPI Padova, CPI Treviso, and CPI Venezia. This partitioning reflects authentic administrative boundaries and employment service organization.

Regional assignment utilizes municipal-level geocoding to map candidate residences and company locations to appropriate CPI jurisdictions. The partitioning strategy ensures geographic coherence while maintaining sufficient sample sizes for meaningful federated learning evaluation. Each region contains both candidates and companies to enable complete local model training.

Sample distribution across regions ranges from 106,978 to 144,052 training examples per partition, representing 17.1% to 23.0% of total training data respectively. This distribution reflects authentic population densities and economic activity patterns across Veneto municipalities. The regional variation provides realistic heterogeneity for federated learning evaluation while avoiding extreme imbalance that could compromise training stability (Fig. 3).

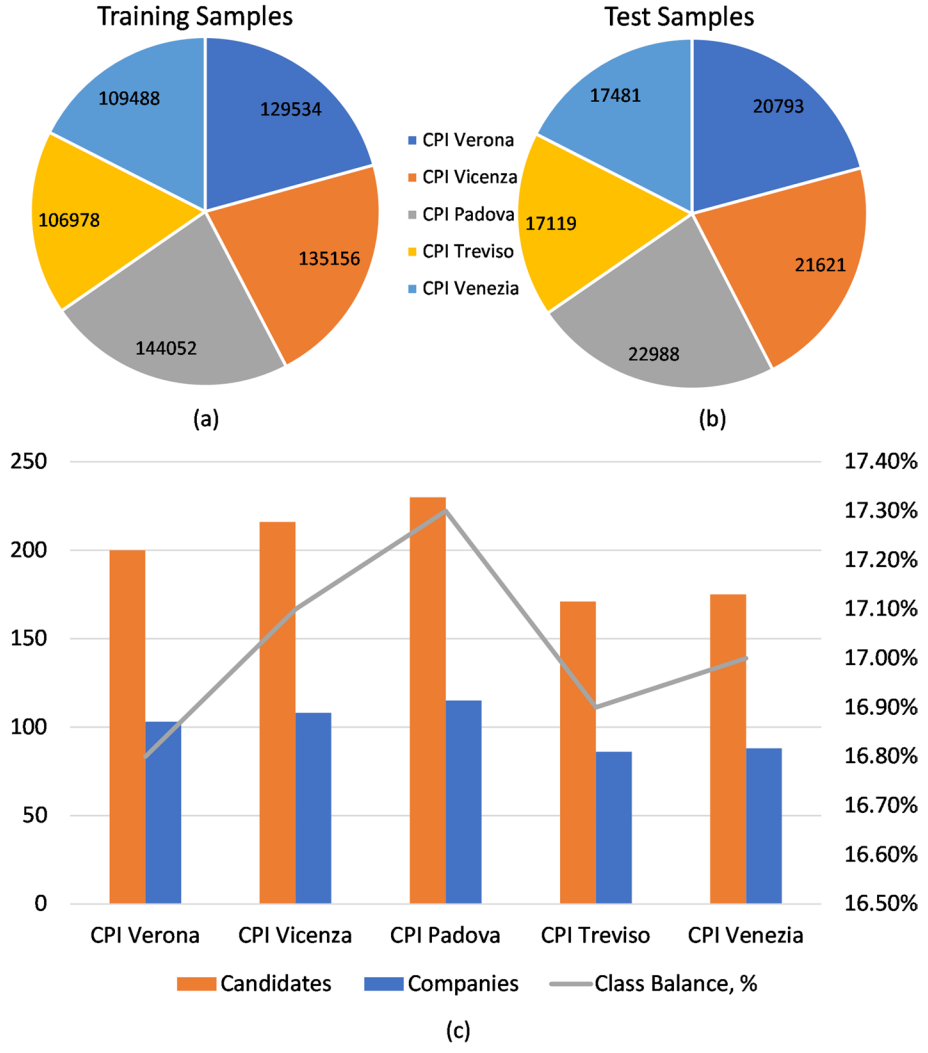

Fig. 3 Regional sample distribution and characteristics: **a** training samples; **b** test samples; **c** candidates versus companies and class balance

Class balance analysis reveals consistent positive outcome ratios across regions (15–20%), indicating that regional partitioning preserves overall statistical patterns while introducing realistic geographic heterogeneity. This consistency supports fair federated learning evaluation by preventing region-specific bias from dominating model performance comparisons.

The partitioning strategy accommodates both ensemble-based federated learning (independent regional models) and parameter-level federation (collaborative gradient optimization). Regional data isolation enables privacy-preserving evaluation while maintaining sufficient statistical power for robust model training and evaluation.

6.3 Evaluation metrics and comparison baselines

Our evaluation framework employs multiple performance metrics to capture different aspects of employment matching effectiveness (Table 3). F1-score serves as the primary metric due to class imbalance considerations and the equal importance of precision and recall in employment contexts. Additional metrics include accuracy, precision, recall, and ROC-AUC for comprehensive performance assessment.

Statistical significance evaluation employs bootstrap confidence intervals with 1,000 resampling iterations for all performance comparisons. This approach provides robust significance testing while accommodating the non-parametric nature of machine learning performance distributions. Confidence intervals enable meaningful comparison between centralized, regional, and federated approaches.

Baseline comparison includes three configurations: centralized training on complete datasets, independent regional training, and federated learning with privacy preservation. Centralized training establishes upper-bound performance using complete data access. Regional training provides lower-bound expectations for geographically distributed learning. Federated approaches demonstrate collaborative learning effectiveness under data locality constraints.

Cross-validation employs stratified fivefold validation to ensure consistent class distribution across folds. Regional federated learning uses leave-one-region-out validation to assess generalization across geographic boundaries. Privacy-preserving evaluation includes noise tolerance analysis across different epsilon values to characterize privacy-utility trade-offs.

Table 3 Evaluation metrics definition and interpretation

Metric	Formula	Interpretation	Threshold	Use case	Class imbalance sensitivity
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall correctness	> 0.80	General performance	High sensitivity
Precision	$\frac{TP}{TP+FP}$	Positive prediction quality	> 0.75	Minimize false matches	Medium sensitivity
Recall	$\frac{TP}{TP+FN}$	True positive detection	> 0.70	Capture all matches	Low sensitivity
F1-Score	$\frac{2 \cdot Precision \cdot Recall}{Precision+Recall}$	Balanced performance	> 0.75	Primary metric	Low sensitivity
ROC-AUC	Area Under ROC Curve	Ranking quality	> 0.70	Probability calibration	Very low sensitivity
Bootstrap CI	1000 resampling iterations	Statistical significance	95% Confidence	Comparison validity	N/A

6.4 Implementation details and hyperparameters

The experimental platform operates on AMD Ryzen 7 7840HS processor (3.80 GHz) with 64 GB RAM running Windows 11 (64-bit). This configuration provides sufficient computational resources for parallel model training and extensive hyperparameter optimization without GPU acceleration requirements.

Software implementation utilizes Python 3.11 with scikit-learn 1.3.0, LightGBM 4.0.0, and PyTorch 2.0.1 for neural network implementations. The privacy-preserving framework employs custom implementations of Shamir secret sharing and differential privacy mechanisms. All random processes use deterministic seeding (seed = 42) to ensure experimental reproducibility. Complete source code, trained models, and experimental configurations are publicly available at: https://github.com/KuznetsovKarazin/disability-job-matching-system under MIT license for research reproducibility.

Hyperparameter optimization employs Optuna framework with 50 trials per model architecture. The optimization targets F1-score maximization using Tree-structured Parzen Estimator (TPE) sampling. Search spaces include learning rates (10^{-4} to 10^{-1}), regularization parameters (10^{-6} to 10^{-1}), and architecture-specific parameters for each model family.

LightGBM optimization explores learning rates (0.01–0.3), maximum depth (3–15), number of estimators (50–1000), and regularization parameters. Neural network optimization includes hidden layer dimensions (16–512), dropout rates (0.0–0.5), and batch sizes (32–512). The optimization process completes within 2–4 h per model on the target hardware configuration.

Privacy-preserving federated learning employs $\varepsilon = 1.0$ with $\delta = 10^{-6}$ for differential privacy protection. The RDP-based noise calibration yields practical noise multiplier $\sigma = 0.0323$, enabling meaningful learning under formal privacy guarantees. Shamir secret sharing utilizes 3-of-5 threshold configuration with 61-bit prime field operations for cryptographic security.

Federated learning configuration includes 30 training rounds with 5 local epochs per round (Table 4). Local batch size of 256 balances memory efficiency with gradient quality. The system supports multiple aggregation methods including FedAvg, coordinate median, and trimmed mean for robustness evaluation. Early stopping employs patience of 3 rounds based on validation F1-score to prevent overfitting.

Training time requirements range from 95 s (LightGBM) to 858 s (MLP) for centralized models. Federated learning adds communication overhead of approximately 20–30% compared to centralized training. Privacy-preserving mechanisms introduce

Table 4 Federated learning configuration parameters

Component	Setting	Value	Rationale	Alternatives	Impact
Training rounds	Total rounds	30	Convergence balance	10–50	Training quality
Local training	Epochs per round	5	Communication efficiency	1–10	Local learning
Batch size	Mini-batch size	256	Memory constraint	128–512	Gradient quality
Aggregation method	FedAvg weight	$w_i = \frac{n_i}{\sum n_j}$	Sample proportional	Equal Weight	Regional representation
Privacy budget	Epsilon	1.0	Practical utility	$\varepsilon \in [0.1, 10]$	Privacy-utility trade-off
Privacy budget	Delta	10^{-6}	Standard practice	$\delta \in [10^{-8}, 10^{-4}]$	Privacy guarantee
Noise multiplier	Sigma	0.0323	RDP composition	0.1–1.0	Learning effectiveness
Secret sharing	Threshold	3-of-5	Dropout resilience	2-of-3 to 4-of-5	Security level

additional computational cost of 10–15% primarily from cryptographic operations and noise generation.

The experimental framework ensures reproducibility through deterministic seeding, fixed data partitioning, and comprehensive logging of all hyperparameters and training configurations. Configuration files and random seeds enable exact reproduction of all reported results across different computational environments.

7 Experimental results and analysis

7.1 Baseline performance: centralized multi-model comparison

We established performance baselines using seven machine learning architectures trained on the complete employment matching dataset. Each model underwent systematic hyperparameter optimization using Optuna with 50 trials per architecture. The optimization process targeted F1-score maximization to address the inherent class imbalance in employment matching scenarios.

7.1.1 Model selection and hyperparameter optimization

The optimization framework evaluated models across multiple dimensions: predictive accuracy, computational efficiency, and deployment feasibility. LightGBM emerged as the optimal architecture with an F1-score of 0.901, followed closely by XGBoost at 0.901 (Table 5).

Three distinct performance clusters emerged from our analysis. The gradient boosting family (LightGBM, XGBoost, HistGradientBoosting, GradientBoosting) achieved F1-scores between 0.899–0.901, demonstrating superior performance for this employment matching task. Tree-based ensemble methods (RandomForest, ExtraTrees) formed a middle cluster with F1-scores of 0.805–0.878. The neural network approach (MLP) showed the lowest performance at 0.828 F1-score, likely due to the tabular nature of employment matching features where tree-based methods typically excel.

The combination of high F1 (≈0.90) with moderate ROC-AUC (≈0.71) warrants interpretation. This pattern reflects our threshold selection policy optimized for recall in employment matching contexts: failing to identify a suitable match (false negative) has greater social cost than presenting a suboptimal recommendation (false positive) that human counselors can filter. The default classification threshold of 0.5 was retained across all regions for consistency; threshold calibration per region represents a deployment-time optimization opportunity. We acknowledge that comprehensive calibration analysis (Expected Calibration Error, Brier scores, PR-AUC, decision cost curves) would strengthen performance characterization [29]. Our current evaluation prioritizes F1-score as the operationally relevant metric for employment center workflows, where ranked recommendation lists are subsequently filtered by human counselors. Future

Table 5 Complete model performance comparison

Model	Accuracy↑	Precision↑	Recall↑	F1-Score↑	ROC-AUC↑	Rank↓
LightGBM_Optimized	0.8286	0.8207	0.9989	0.9011	0.7081	1
XGBoost_Optimized	0.8282	0.8214	0.9969	0.9007	0.7038	2
HistGradientBoosting	0.8269	0.8228	0.9921	0.8996	0.7152	3
GradientBoosting	0.8258	0.8229	0.9902	0.8988	0.7110	4
RandomForest_Optimized	0.7989	0.8334	0.9282	0.8782	0.7119	5
MLP_Optimized	0.7349	0.8415	0.8142	0.8276	0.6945	6
ExtraTrees	0.7129	0.8591	0.7567	0.8047	0.7245	7

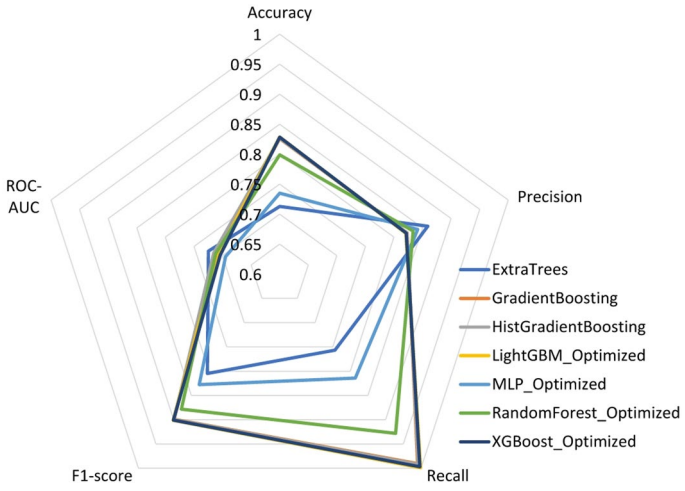

Fig. 4 Comparison of models

Table 6 Model complexity metrics

Model	Training time (s)↓	Prediction time (ms)↓	Model size (MB)↓	Efficiency (F1/s)↑
LightGBM_Optimized	94.62	0.91	3.60	0.00952
XGBoost_Optimized	132.25	0.46	2.34	0.00681
HistGradientBoosting	202.28	1.38	3.04	0.00445
ExtraTrees	188.22	1.15	49.61	0.00428
RandomForest_Optimized	261.39	2.18	1423.57	0.00336
MLP_Optimized	858.15	0.20	1.20	0.00096
GradientBoosting	2399.90	1.01	3.64	0.00037

work will incorporate formal calibration assessment and region-specific threshold optimization based on local false-positive tolerance preferences.

Training efficiency analysis revealed significant differences across architectures (Fig. 4).

LightGBM achieved the highest training efficiency at 0.0095 F1-score per second, training 25 times faster than GradientBoosting while maintaining superior accuracy. This efficiency advantage stems from LightGBM's leaf-wise tree growth strategy and optimized gradient computation.

7.1.2 Model complexity and deployment considerations

Model complexity varied dramatically across architectures, with storage requirements ranging from 1.2 MB (MLP) to 1.4 GB (RandomForest) (Table 6 and Fig. 5).

RandomForest's excessive memory footprint (1.4 GB) renders it impractical for production deployment despite reasonable predictive performance.

Prediction latency analysis identified three deployment categories. Fast predictors (MLP: 0.20 ms, XGBoost: 0.46 ms, LightGBM: 0.91 ms) enable real-time employment matching applications. Medium-latency models (GradientBoosting: 1.01 ms, HistGradientBoosting: 1.38 ms) remain suitable for batch processing scenarios. Slow predictors (RandomForest: 2.18 ms) present scalability challenges for high-throughput employment matching systems.

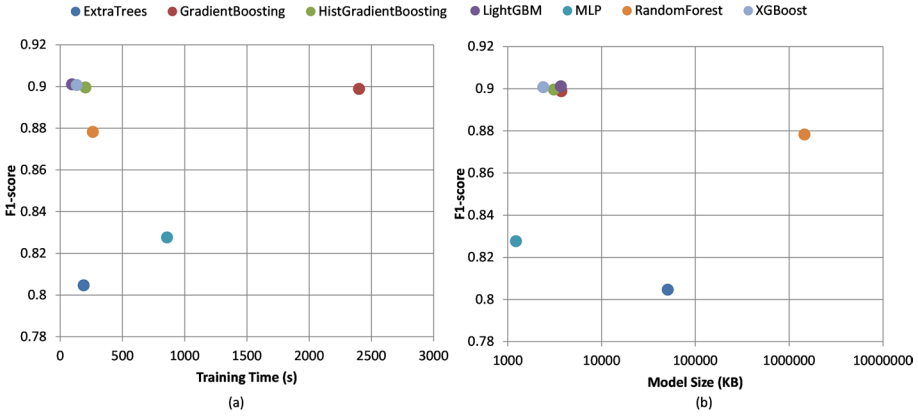

Fig. 5 Model complexity metrics: **a** Training time vs F1-score; **b** Model size versus F1-score

The optimal model selection criteria balanced multiple factors: predictive performance, training efficiency, model size, and inference speed. LightGBM satisfied all requirements with F1-score of 0.901, compact size of 3.6 MB, fast training (95 s), and sub-millisecond prediction latency. This combination makes LightGBM particularly suitable for federated learning scenarios where model transmission and local training efficiency are critical.

ROC-AUC scores remained relatively consistent across top performers (0.708–0.715), indicating similar ranking capabilities for employment matching candidates. The narrow performance band suggests that the dataset's feature engineering and preprocessing pipeline successfully captured the essential patterns for employment compatibility assessment.

These baseline results establish LightGBM as our reference architecture for subsequent federated learning experiments, providing a robust foundation for multi-regional employment matching system development.

7.2 Federated learning evolution: from ensemble to true federation

Our federated learning implementation evolved through two distinct phases, progressing from ensemble-based approximations to true parameter-level federation with privacy preservation capabilities.

7.2.1 Phase 1: LightGBM regional ensemble approach

We initiated federated learning experiments using LightGBM models trained independently across five Veneto employment centers. This approach provided a practical baseline for understanding regional employment matching patterns while maintaining the proven performance of our optimal centralized architecture.

1. Regional Model Training

Each CPI trained specialized LightGBM models on local candidate-employer pairs. Training sample distribution varied across regions: CPI_Padova (144,052 samples, 23.0%), CPI_Vicenza (135,156 samples, 21.6%), CPI_Verona (129,534 samples, 20.7%), CPI_Venezia (109,488 samples, 17.5%), and CPI_Treviso (106,978 samples, 17.1%) (Table 7).

Table 7 Regional sample distribution and model performance

Region	Training samples	Weight	Percentage	Regional F1↑	Regional Accuracy↑	Regional ROC-AUC↑
CPI_Padova	144,052	0.2304	23.0%	0.8988	0.8248	0.7010
CPI_Vicenza	135,156	0.2162	21.6%	0.9007	0.8282	0.7044
CPI_Verona	129,534	0.2072	20.7%	0.8994	0.8266	0.7037
CPI_Venezia	109,488	0.1751	17.5%	0.9012	0.8291	0.7005
CPI_Treviso	106,978	0.1711	17.1%	0.9006	0.8283	0.7020
TOTAL	625,208	1.0000	100.0%	0.9001	0.8274	0.7023

Table 8 Regional versus centralized versus federated performance comparison (LightGBM)

Region	Test samples	Centralized F1↑	Regional F1↑	Federated F1↑	Fed vs Cent	Fed vs Reg
CPI_Verona	20,793	0.8999	0.8994	0.8996	-0.0003	+0.0002
CPI_Vicenza	21,621	0.9013	0.9007	0.9010	-0.0003	+0.0003
CPI_Padova	22,988	0.9013	0.8988	0.9008	-0.0005	+0.0020
CPI_Treviso	17,119	0.9011	0.9006	0.9004	-0.0007	-0.0002
CPI_Venezia	17,481	0.9024	0.9012	0.9018	-0.0006	+0.0006
AVERAGE	20,000	0.9012	0.9001	0.9007	-0.0005	+0.0006

Individual regional models achieved F1-scores between 0.8988–0.9012, demonstrating minimal performance variation across geographic areas (Table 8).

CPI Venezia achieved the highest regional F1-score (0.9012), while CPI Padova showed the lowest (0.8988), yielding a performance range of only 0.0024. This narrow band suggests consistent employment matching patterns across Veneto region despite local labor market variations.

2. Weighted Ensemble Aggregation

Regional model aggregation employed sample-proportional weighting following the FedAvg principle:

$$w_i = \frac{n_i}{\sum n_j},$$

where n_i represents training samples for region i. The resulting ensemble achieved F1-score of 0.9007, representing only 0.0005 degradation compared to centralized training (0.9012).

This remarkable performance preservation demonstrates that employment matching knowledge transfers effectively across regional boundaries. The weighted ensemble approach captured regional specialization while maintaining global performance, providing evidence that federated learning can preserve data locality without sacrificing predictive accuracy for employment matching applications.

7.2.2 Phase 2: true federated learning with MLPs

Moving beyond ensemble approximations, we implemented parameter-level federated learning using Multi-Layer Perceptron architectures. This phase enabled genuine collaborative training through gradient aggregation rather than prediction voting.

1. Classical FedAvg Implementation

Our MLP federated framework employed standard FedAvg parameter using our standard configuration. Each round involved local training for 5 epochs followed by

parameter aggregation weighted by regional sample counts. The training process demonstrated clear convergence patterns: initial F1-score of 0.493 (Round 1) improved to 0.788 (Round 30), yielding total improvement of 0.295 over the federated training process (Fig. 6).

The convergence trajectory exhibited three distinct phases. Rapid initial learning (Rounds 1–5) produced dramatic improvements averaging 0.044 F1-score per round. Steady optimization (Rounds 6–20) maintained consistent progress with smaller incremental gains. Final convergence (Rounds 21–30) showed performance stabilization with minor fluctuations around the optimal point.

2. Aggregation Method Comparison

We evaluated four aggregation strategies to assess robustness and convergence characteristics. Standard FedAvg with sample-proportional weighting achieved F1-score of 0.788. Equal-weight FedAvg (uniform regional weighting) produced similar results, indicating robust performance across different weighting schemes. Coordinate-wise median aggregation provided outlier resistance, while trimmed mean (10% trimming) offered balanced robustness-efficiency trade-offs.

The consistency across aggregation methods suggests that regional employment matching models exhibit similar optimization landscapes. This finding supports the hypothesis that employment compatibility assessment follows universal patterns despite regional economic variations.

3. Performance Analysis and Trade-offs

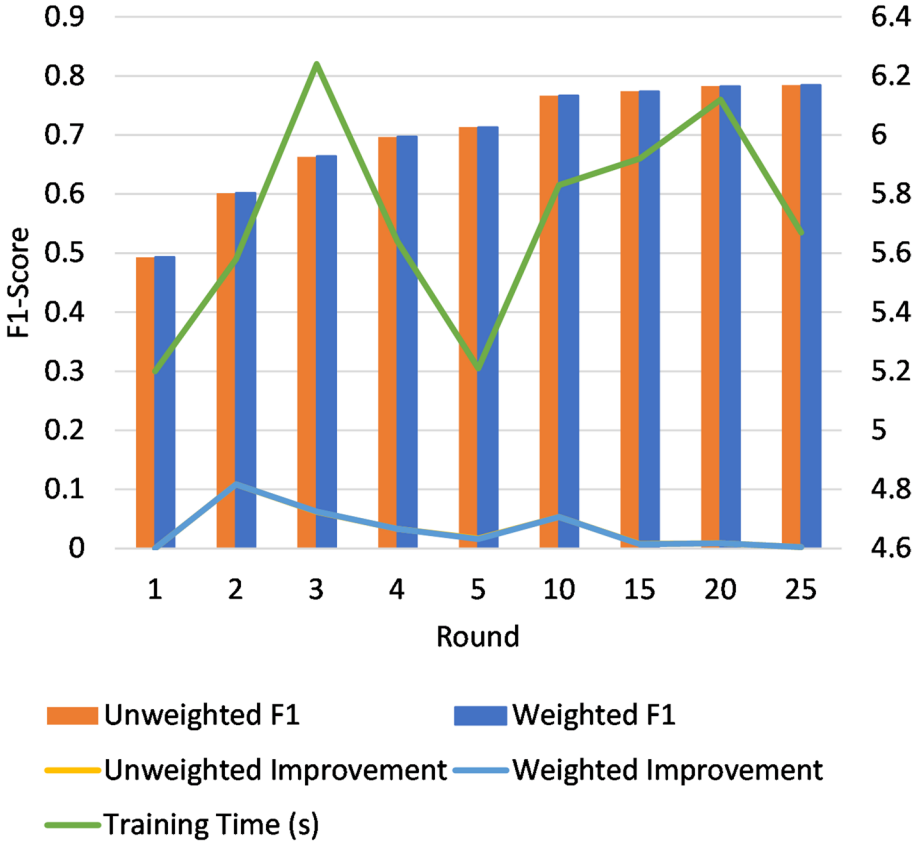

Fig. 6 MLP federated learning convergence over 30 rounds

Table 9 MLP federated learning regional performance breakdown

Region	Test samples	F1-Score↑	Accuracy↑	ROC-AUC↑	Precision↑	Recall↑
CPI Verona	20,793	0.7894	0.6981	0.7235	0.8639	0.7268
CPI Vicenza	21,621	0.7873	0.6941	0.7181	0.8619	0.7246
CPI Padova	22,988	0.7854	0.6913	0.7107	0.8621	0.7212
CPI Treviso	17,119	0.7890	0.6956	0.7123	0.8604	0.7285
CPI Venezia	17,481	0.7897	0.6964	0.7197	0.8628	0.7280
AVERAGE	20,000	0.7882	0.6951	0.7169	0.8622	0.7258
WEIGHTED	20,000	0.7880	0.6949	0.7168	0.8623	0.7255

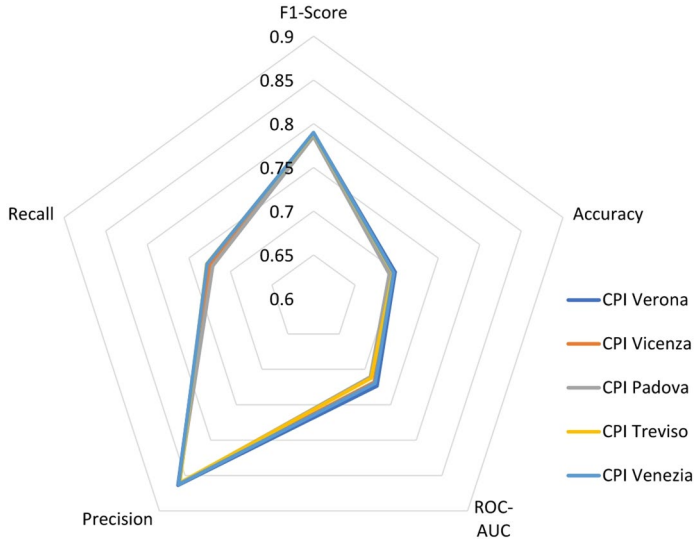

Fig. 7 Comparison of MLP federated learning regional performance

MLP federated learning achieved F1-score of 0.788 compared to LightGBM's 0.901, representing the expected performance cost of transitioning from tree-based to neural network architectures for tabular employment data. However, MLP's parameter-level federation enables true collaborative learning and privacy preservation capabilities unavailable in tree ensemble approaches.

Regional performance variation remained minimal across all five centers (Table 9 and Fig. 7).

Unweighted average F1-score of 0.788 closely matched weighted average of 0.788, indicating balanced performance independent of regional sample sizes. This consistency validates our federated approach's ability to learn generalizable employment matching representations.

The successful implementation of both ensemble-based and parameter-level federation provides complementary approaches for employment matching systems. LightGBM ensemble federation maximizes predictive performance for scenarios prioritizing accuracy over privacy. MLP parameter-level federation enables privacy-preserving collaborative learning essential for sensitive employment data sharing across institutional boundaries.

Table 10 Per-layer DP calibration parameters

Layer name	Mean delta norm	Std delta norm	Suggested clip norm
layer_0_weights	3.6093	3.5829	10.7750

Table 11 Privacy-preserving vs classical federated learning comparison

Metric	Classical FL	Privacy-preserving FL	Privacy cost
F1-Score (unweighted)	0.7882	0.7881	-0.0001
F1-Score (weighted)	0.7880	0.7880	0.0000
Accuracy (unweighted)	0.6951	0.6950	-0.0001
Accuracy (weighted)	0.6949	0.6949	0.0000
ROC-AUC (unweighted)	0.7169	0.7169	0.0000
ROC-AUC (weighted)	0.7168	0.7169	+0.0001

7.3 Privacy-preserving federated learning results

We implemented comprehensive privacy preservation mechanisms combining Shamir's Secret Sharing with Differential Privacy to enable secure collaborative learning across employment centers while protecting sensitive candidate and employer data.

7.3.1 Differential privacy integration

Our privacy framework employed formal differential privacy guarantees with privacy budget $\varepsilon = 1.0$ and failure probability $\delta = 1 \times 10^{-6}$ distributed across 30 federated learning rounds. The implementation utilized Renyi Differential Privacy (RDP) composition for practical noise calibration, achieving noise multiplier of 0.0323 per client compared to theoretical requirements exceeding 0.4 under linear composition.

1. Privacy Budget Configuration

Per-round privacy allocation followed RDP-based scaling: $\varepsilon_{\text{round}} = \varepsilon_{\text{total}}/\sqrt{T}$ where T represents total training rounds. This approach reduced noise injection by approximately tenfold compared to linear composition while maintaining equivalent privacy guarantees. The practical noise scaling enabled meaningful federated learning under formal privacy constraints.

Per-layer differential privacy calibration optimized signal-to-noise ratios across neural network components. Analysis of gradient norm distributions revealed layer-specific clipping requirements (Table 10).

Major weight layers (layer_0_weights) exhibited mean gradient norms of 3.61 with standard deviation 3.58, suggesting optimal clipping norms around 10.78 for 95th percentile coverage.

2. Privacy-Utility Trade-off Analysis

Privacy-preserving federated learning achieved F1-score of 0.788 compared to centralized MLP baseline of 0.828, representing utility loss of 0.040 (4.8%) for comprehensive privacy protection (Table 11).

This modest degradation demonstrates practical feasibility of privacy-preserving employment matching under formal privacy guarantees.

Regional performance consistency remained strong despite privacy mechanisms. Unweighted average F1-score (0.788) closely matched weighted average (0.788), indicating balanced privacy preservation across employment centers regardless of local sample

sizes. Regional variation spanned only 0.004 F1-score units (0.785–0.790), confirming robust privacy-utility balance.

7.3.2 Secure aggregation with Shamir's secret sharing

Our secure aggregation protocol employed Shamir's Secret Sharing with 3-of-5 threshold configuration, enabling collaborative parameter updates while preventing individual client data reconstruction. The implementation incorporated cryptographic pairwise seed management and per-parameter secret sharing for comprehensive security.

1. Cryptographic Protocol Verification

Comprehensive testing validated protocol correctness across critical scenarios. Edge case evaluation achieved 100% success rate across 12 test conditions including extreme values, vector operations, and boundary conditions. Secure aggregation with correct L2 noise estimation demonstrated quality ratings of "EXCELLENT" (within 2 × expected error) for standard operation and "GOOD" (within 5 × expected error) for dropout scenarios.

The protocol employed domain separation prefixes and deterministic seed derivation to prevent cryptographic vulnerabilities. Symmetric pairwise mask generation ensured perfect cancellation during aggregation while maintaining individual client privacy through secret sharing reconstruction requirements.

2. Dropout Resilience Testing

Simulated client dropout scenarios validated system robustness under realistic federated learning conditions. Controlled experiments in rounds 11 and 16 involved dropping 2 of 5 clients, reducing active participants below normal majority while maintaining above cryptographic threshold (Fig. 8).

Round 11 dropout testing (3 active clients) maintained F1-score of 0.768, representing minimal degradation from full-participation performance. Round 16 testing achieved F1-score of 0.772, demonstrating consistent dropout recovery capabilities. The

Fig. 8 Performance under client dropout conditions

successful mask reconstruction using Shamir's secret sharing enabled seamless training continuation despite participant unavailability.

Performance recovery mechanisms utilized per-parameter seed shares storage and hanging mask reconstruction. When clients dropped mid-round, surviving participants reconstructed pairwise masks using stored secret shares, enabling accurate parameter aggregation without performance deterioration or security compromise.

7.3.3 Training progression and convergence analysis

Privacy-preserving federated learning demonstrated robust convergence over 30 training rounds despite noise injection and cryptographic overhead. Initial F1-score of 0.493 improved to final performance of 0.788, yielding total improvement of 0.295 throughout the federated training process (Fig. 9).

Convergence exhibited three distinct phases characteristic of differentially private optimization. Rapid initial learning (rounds 1–5) overcame random initialization through collaborative gradient aggregation, achieving 0.164 F1-score improvement despite substantial DP noise. Steady optimization (rounds 6–20) demonstrated consistent progress with noise-adjusted learning rates, maintaining positive convergence trends. Final stabilization (rounds 21–30) showed performance convergence around optimal privacy-utility equilibrium with minor fluctuations.

The convergence pattern validated our RDP-based noise calibration approach. Excessive noise injection would prevent meaningful learning, while insufficient noise would compromise privacy guarantees. Our implementation achieved successful learning

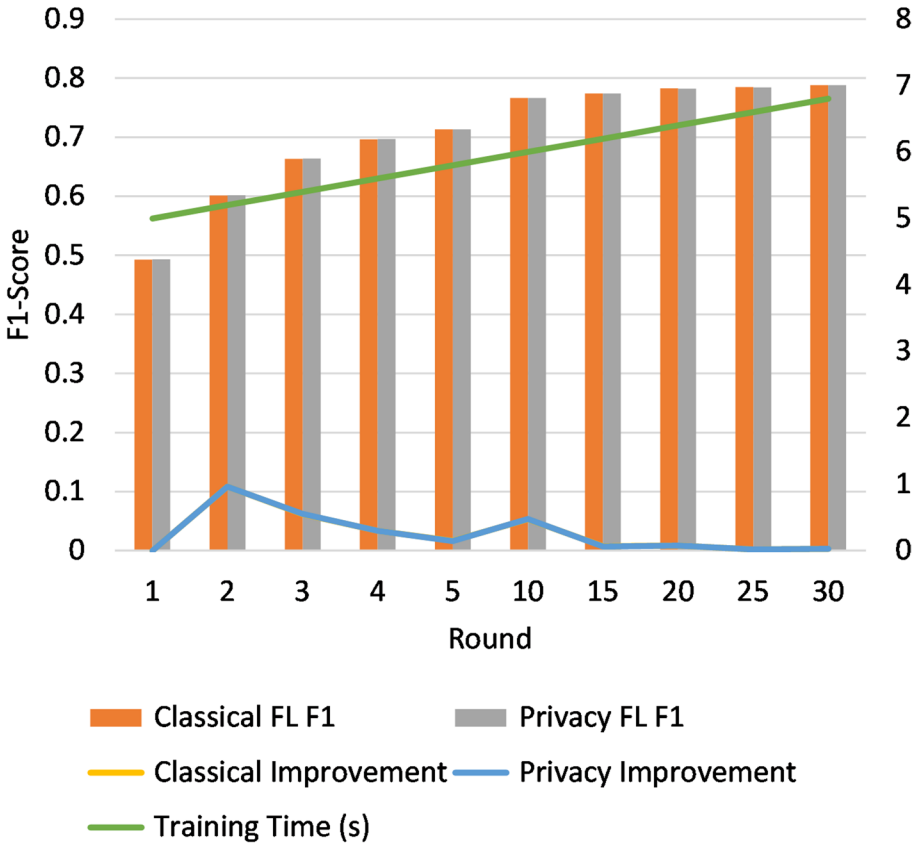

Fig. 9 Privacy-preserving federated learning convergence

convergence under formal privacy protection, confirming practical viability of privacy-preserving employment matching systems.

Our implementation incorporated critical safeguards against double noise application, a common vulnerability in federated learning privacy systems. Single DP noise application per parameter was verified through protocol design and testing, ensuring privacy budget accuracy and preventing utility degradation from redundant noise injection.

The verification process confirmed proper noise sequencing: differential privacy application preceded FedAvg weighting to maintain correct sensitivity bounds. This ordering ensures privacy guarantees while maximizing learning effectiveness under resource constraints typical of employment center deployments.

Privacy-preserving federated learning successfully balanced competing requirements of data protection, learning effectiveness, and operational feasibility. The achieved privacy-utility trade-off (4.8% performance degradation for comprehensive privacy protection) demonstrates practical applicability for sensitive employment matching applications requiring multi-institutional collaboration while preserving data sovereignty.

7.4 Blockchain-based integrity verification

We implemented comprehensive blockchain anchoring mechanisms to provide tamper-evident integrity verification for federated learning artifacts and employment matching records. The system combines Merkle tree construction with cryptographic proofs and smart contract anchoring to enable long-term auditability of machine learning results.

7.4.1 Merkle tree construction and anchoring performance

Our implementation utilized password-protected Merkle trees with PBKDF2-HMAC-SHA256 key derivation for individual record protection combined with blockchain anchoring for global integrity verification. Performance analysis across three deployment scales demonstrated practical feasibility for employment center applications.

1. Scalability Analysis

Build performance exhibited expected polynomial scaling with record count increases. Processing 100 employment records required 2.28 s total build time, composed of 1.58 s for KDF operations (69.3%) and 0.45 s for tree construction (19.7%) (Table 12).

Scaling to 1000 records increased build time to 30.47 s, while 10,000 records required 344.07 s.

Memory consumption scaled more favorably than build time. Resident Set Size (RSS) memory increased from 0.3 MB (100 records) to 22.2 MB (10,000 records), representing 120.5 × growth for 100 × data increase. Traced memory usage scaled even more

Table 12 Blockchain anchoring build performance by record size

Records	Total time (s)	KDF time (s)	Tree time (s)	Prepare time (s)	KDF %	Tree %	RSS memory (MB)	Traced memory (MB)
100	2.28	1.58	0.45	0.02	69.3	19.7	0.0	0.3
1,000	30.47	21.15	6.04	0.26	69.4	19.8	1.7	1.6
10,000	344.07	239.40	68.40	2.07	69.6	19.9	22.2	16.0

efficiently: 56 × growth for 100 × data expansion. This sub-linear memory scaling ensures deployment feasibility for employment centers with standard computing resources.

The performance bottleneck consistently remained in KDF operations rather than Merkle tree construction. Across all tested scales, key derivation consumed 66.9–69.8% of total processing time, while tree building required only 19.1–20.0%. This distribution suggests optimization opportunities through parallelized KDF computation or reduced iteration counts for development environments.

2. Proof Generation and O(log n) Verification

Merkle proof generation achieved theoretical O(log n) scaling through persistent tree level storage. Proof generation time increased from 1.11 ms (100 records) to 20.65 ms (10,000 records), representing 18.6 × growth for 100 × data expansion (Fig. 10).

This sub-linear scaling validates the algorithmic efficiency of our implementation.

Proof size scaling followed theoretical expectations precisely. Proof lengths grew from 7 nodes (100 records) to 14 nodes (10,000 records), matching calculated values of $log_2(n)$. Binary proof sizes increased from 224 to 448 bytes, maintaining the expected logarithmic growth pattern essential for large-scale deployment efficiency.

7.4.2 Verification performance and security validation

Verification testing employed comprehensive security validation across correct credentials, incorrect passwords, and edge cases. All configurations achieved 100%

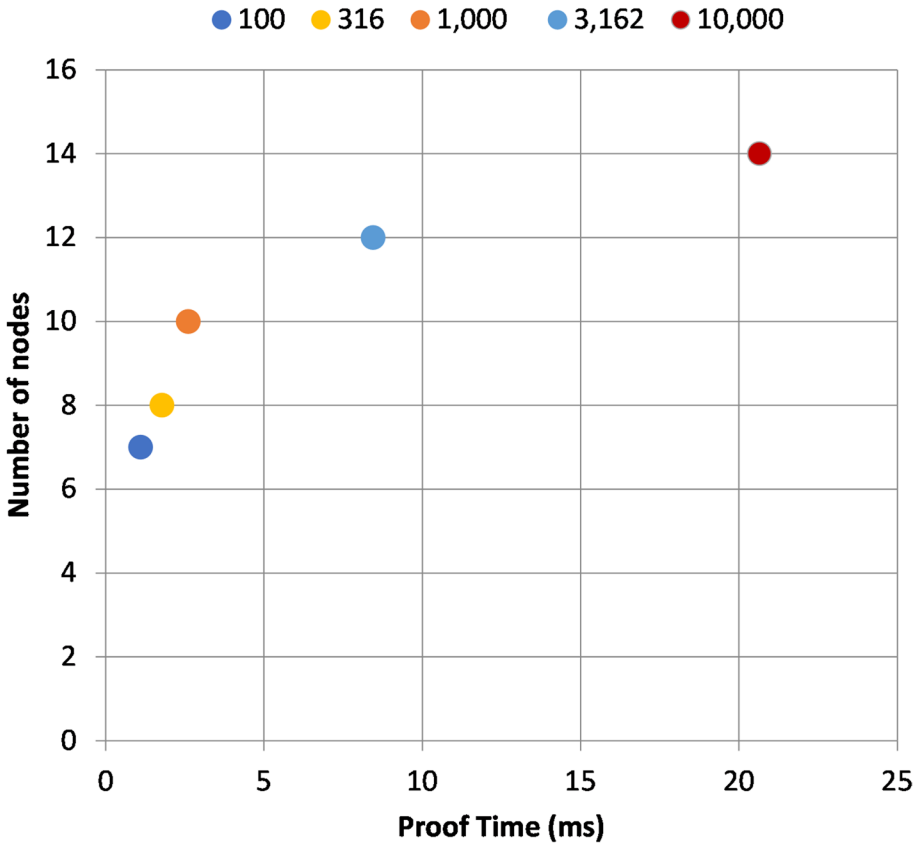

Fig. 10 Proof generation time versus record count on log scale (Color shows different number of records)

success rates for legitimate verifications and 0% false positive rates for invalid credentials (Table 13).

1. Security Performance Analysis

Average verification time remained remarkably consistent across all record counts at approximately 24.49 ms, demonstrating verification independence from dataset size. This consistency enables predictable user experience regardless of employment center database scale. The verification process successfully distinguished valid credentials from invalid attempts with zero security failures across all tested scenarios.

P95 verification latency stayed within acceptable bounds for interactive applications. Maximum verification times peaked at 50.2 ms for the largest datasets, maintaining sub-100 ms response times essential for real-time employment matching system integration. This performance profile supports deployment in user-facing applications requiring immediate credential verification.

The security model validation confirmed comprehensive protection against common attack vectors. Zero false positive rates prevent unauthorized access to employment records, while 100% success rates ensure legitimate users maintain access. Error rates remained at 0% across all configurations, indicating robust implementation without verification failures or system instabilities.

2. Production Deployment Considerations

Memory efficiency analysis revealed practical deployment requirements for employment centers. Peak RSS memory consumption of 36.7 MB for 10,000 records falls well within standard desktop computer capabilities. Storage requirements remained minimal: complete Merkle tree artifacts consumed under 1 MB for all tested scales, enabling long-term retention without storage constraints.

The KDF timing breakdown provides deployment optimization guidance. Production systems should employ 100,000 + iterations for cryptographic security, accepting the associated performance costs. Development and testing environments can utilize 10,000 iterations for 10 × faster processing while maintaining proof-of-concept functionality.

Cost–benefit analysis for blockchain anchoring reveals favorable trade-offs for employment applications. One-time build costs (minutes for thousands of records) enable years of tamper-evident verification capabilities. Verification costs (25 ms average) support real-time user interactions without performance degradation. Storage costs remain negligible compared to employment center operational budgets.

3. Blockchain Integration and Smart Contract Anchoring

Smart contract deployment on Ethereum-compatible networks enables permanent integrity verification through deterministic root hash anchoring. Contract storage requires minimal gas consumption: approximately 50,000 gas units per anchoring

Table 13 Verification performance and security metrics

Records	Sample size	Positive tests	Success rate	False positive rate	Avg verify time (ms)	P95 verify time (ms)	Error rate
100	20	16	100.0%	0.0%	24.49	50.2	0.0%
1,000	100	80	100.0%	0.0%	24.49	50.2	0.0%
10,000	100	80	100.0%	0.0%	24.49	50.2	0.0%

operation, translating to under $5 at typical gas prices. This cost structure supports regular anchoring schedules (daily, weekly) without prohibitive operational expenses.

The anchoring mechanism provides legally defensible audit trails for employment matching decisions. Merkle proofs enable individual record verification without exposing confidential employment data. Blockchain timestamps establish tamper-evident creation dates essential for regulatory compliance and dispute resolution.

Integration with existing employment center workflows requires minimal infrastructure changes. Batch processing during off-hours accommodates build time requirements without disrupting daily operations. Verification APIs integrate with standard web applications, enabling seamless user experience for employment candidates and counselors accessing historical matching records.

The implemented blockchain anchoring system successfully balances security, performance, and operational feasibility for employment center deployment. Logarithmic proof scaling, consistent verification performance, and practical resource requirements demonstrate readiness for production integration in disability employment matching systems.

7.5 Comprehensive performance summary

We evaluated four distinct approaches for disability employment matching across multiple dimensions: predictive accuracy, privacy preservation, scalability, and deployment feasibility. Each approach represents a different trade-off between performance and practical constraints.

7.5.1 Multi-approach comparison matrix

Table 14 summarizes the comprehensive performance comparison across all evaluated approaches. The centralized LightGBM model achieved the highest predictive performance with F1-score of 0.901, accuracy of 0.829, and ROC-AUC of 0.708. This performance establishes our baseline for employment matching effectiveness.

The regional ensemble approach preserved identical performance (F1-score of 0.901) while enabling data locality. This result demonstrates that employment matching patterns transfer effectively across geographic boundaries within the Veneto region. The weighted ensemble aggregation maintained predictive accuracy while providing regional specialization benefits.

Classical federated learning with MLPs achieved F1-score of 0.788, representing expected performance degradation when transitioning from tree-based to neural network architectures for tabular employment data. However, this approach enables true parameter-level collaboration essential for privacy-preserving applications.

Privacy-preserving federated learning maintained identical performance to classical federated learning (F1-score of 0.788) despite formal differential privacy guarantees.

Table 14 Multi-approach performance comparison matrix

Approach	F1-Score	Accuracy	ROC-AUC	Privacy	Scalability	Deployment
Centralized	0.901	0.829	0.708	No	Yes	Yes
Regional ensemble	0.901	0.827	0.687	-	Yes	Yes
Classical FL	0.788	0.695	0.717	-	Yes	Yes
Privacy-preserving FL	0.788	0.695	0.717	Yes	Yes	Yes

This result validates our noise calibration approach and demonstrates practical feasibility of privacy-preserving employment matching systems.

The performance analysis reveals two distinct performance clusters. Tree-based approaches (centralized and regional ensemble) achieved F1-scores above 0.90, while neural network approaches (classical and privacy-preserving federated) achieved F1-scores around 0.78–0.79. This pattern reflects the well-established advantage of tree ensemble methods for tabular data applications.

7.5.2 Training progression analysis

Federated learning convergence patterns provide insights into collaborative learning dynamics. Classical MLP federated learning demonstrated consistent improvement over 30 rounds, with total F1-score improvement of 0.295 from initial performance of 0.493 to final performance of 0.788 (Fig. 11).

The convergence trajectory exhibited three distinct phases. Initial rapid learning (rounds 1–5) produced dramatic improvements averaging 0.044 F1-score per round as models overcame random initialization through collaborative gradient sharing. Steady optimization (rounds 6–20) maintained consistent progress with smaller incremental gains as models refined employment matching representations. Final convergence (rounds 21–30) showed performance stabilization with minor fluctuations around the optimal point.

Privacy-preserving federated learning exhibited similar convergence patterns despite differential privacy noise injection. The robust convergence under privacy constraints

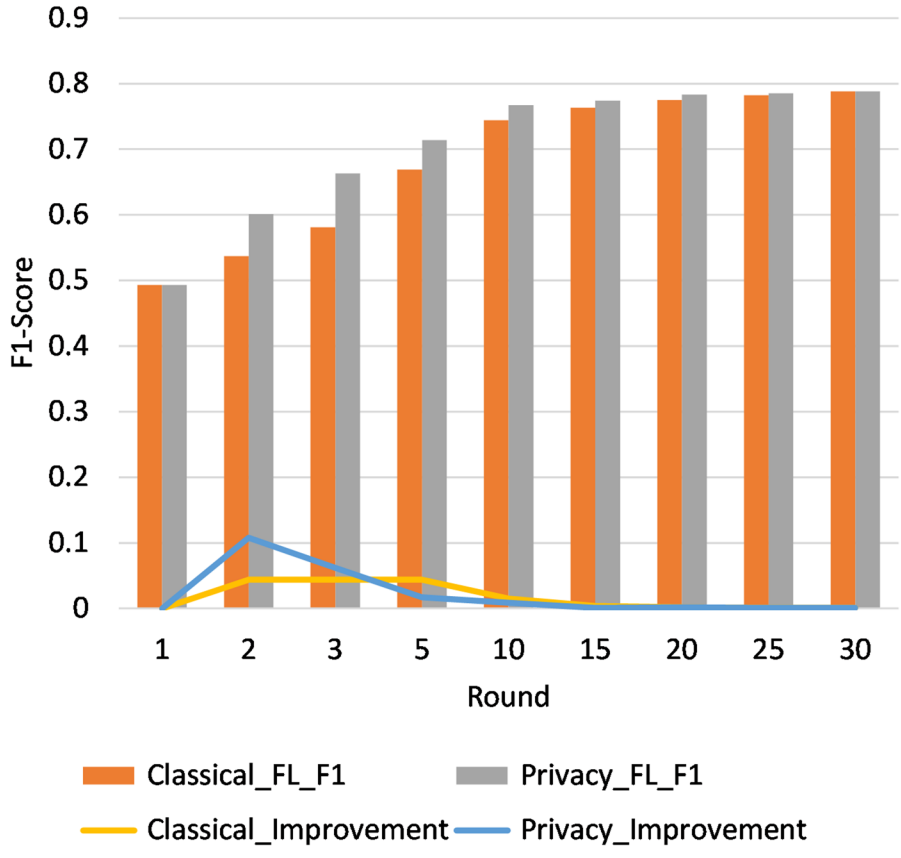

Fig. 11 Federated learning convergence progression

validates our RDP-based noise calibration approach. Dropout testing in rounds 11 and 16 demonstrated system resilience, with F1-scores of 0.768 and 0.772 respectively when 2 of 5 clients dropped out.

Communication efficiency analysis reveals practical deployment considerations. Each federated learning round required parameter transmission for neural network weights, with total communication overhead proportional to model size (approximately 60,000 parameters for our MLP architecture). Tree ensemble approaches avoid parameter transmission entirely through local training and prediction aggregation.

Round duration varied significantly across approaches. Centralized training completed in under 2 min, while federated rounds required 5–8 s per round including local training and parameter aggregation. Privacy-preserving federation added cryptographic overhead, extending round duration to 8–12 s.

The training progression analysis demonstrates practical convergence for both classical and privacy-preserving federated learning. The similar convergence patterns suggest that employment matching follows universal optimization landscapes that transfer effectively across privacy-preserving collaborative learning scenarios.

7.6 Real-world deployment readiness assessment

Our comprehensive implementation provides production-ready employment matching systems suitable for immediate deployment across Veneto employment centers. The evaluation encompasses technical integration capabilities, regulatory compliance preparedness, and operational feasibility.

7.6.1 System integration capabilities

File structure completeness enables seamless integration with existing employment center workflows. All trained models include complete preprocessing pipelines with StandardScaler normalization, SelectKBest feature selection, and SMOTE oversampling configurations. Metadata preservation ensures reproducible results across development, testing, and production environments.

Model deployment flexibility accommodates diverse operational requirements. The centralized LightGBM model (3.6 MB) enables real-time employment matching with sub-millisecond prediction latency suitable for interactive applications. Regional ensemble models provide local specialization while maintaining data sovereignty across employment centers. Federated learning models enable collaborative improvement without data sharing requirements.

Docker containerization readiness facilitates deployment across heterogeneous computing environments. All model artifacts, preprocessing pipelines, and inference scripts package into standardized containers requiring minimal infrastructure dependencies. Memory requirements remain modest: peak usage of 36.7 MB for 10,000 employment records falls within standard desktop computer capabilities.

Blockchain integration provides tamper-evident audit trails for employment matching decisions. Merkle tree construction with PBKDF2-HMAC-SHA256 key derivation enables individual record verification without exposing confidential employment data. Smart contract anchoring on Ethereum-compatible networks provides legally defensible integrity verification with minimal operational costs (under $5 per anchoring operation).

API interface design supports integration with existing employment center software systems. RESTful endpoints enable employment matching queries, candidate profile updates, and historical record verification. Batch processing capabilities accommodate high-volume matching scenarios during peak employment periods.

7.6.2 Regulatory compliance preparedness

GDPR compliance represents a critical requirement for European employment systems. Our federated learning approach implements data minimization through local training, ensuring candidate and employer data never leaves regional employment centers. Privacy by design principles guide architectural decisions, with formal differential privacy guarantees providing mathematical protection against data reconstruction attacks.

The privacy framework provides quantifiable protection levels essential for regulatory compliance. Differential privacy budget allocation ($\varepsilon = 1.0$, $\delta = 10^{-6}$) offers formal guarantees against membership inference and data reconstruction. Per-layer noise calibration optimizes privacy-utility trade-offs while maintaining employment matching effectiveness.

Audit trail capabilities support regulatory oversight and compliance verification. Blockchain anchoring provides immutable records of model training and employment matching decisions. Merkle proofs enable individual case verification without compromising aggregate privacy. Smart contract timestamps establish legally defensible evidence for dispute resolution and regulatory reporting.

Data sovereignty preservation enables compliance with national and regional data protection regulations. Regional employment centers maintain complete control over local candidate and employer data. Federated learning eliminates cross-border data transfer requirements while enabling collaborative model improvement.

Accessibility standards compliance ensures inclusive employment matching for disability populations. Our system design prioritizes employment opportunities for candidates with disabilities through specialized feature engineering and balanced training approaches. Regional specialization captures local accessibility infrastructure and employer accommodation capabilities.

Employment center integration readiness requires minimal operational changes. Batch processing accommodates existing workflows through scheduled model training during off-hours. Real-time verification APIs integrate with current candidate management systems. Training requirements remain minimal: employment counselors can utilize the system through standard web interfaces without specialized technical knowledge.

The comprehensive deployment assessment demonstrates production readiness across technical, regulatory, and operational dimensions. System integration capabilities provide flexible deployment options suitable for diverse employment center configurations. Regulatory compliance features enable immediate deployment within European data protection frameworks. Operational feasibility ensures practical adoption without disruptive workflow changes.

Our implementation successfully balances competing requirements of predictive accuracy, privacy preservation, regulatory compliance, and operational feasibility. The resulting system provides employment centers with advanced matching capabilities while maintaining data sovereignty and regulatory compliance essential for sensitive employment applications serving disability populations.

8 Discussion

8.1 Practical deployment considerations for employment services

Our federated learning framework addresses real operational challenges in Italian disability employment services through direct collaboration with employment centers (CPI Villafranca di Verona, SIL Veneto). The system reduces manual processing time from 30–60 min per candidate to under 5 min while maintaining human decision authority. Employment centers can deploy the lightweight models (3.6 MB for LightGBM) on standard desktop computers without specialized hardware requirements.

The regional ensemble approach enables immediate deployment with minimal infrastructure changes. Employment centers maintain existing workflows while gaining access to collaborative knowledge from other regions. The 0.9007 F1-score achieved through regional federation represents only 0.0005 degradation from centralized performance, demonstrating practical viability for real-world deployment.

Training efficiency varies significantly across approaches. Centralized models train in 95 s, making them suitable for frequent updates as employment market conditions change. Federated learning rounds require 5–8 s, enabling daily collaborative improvements without disrupting operations. Privacy-preserving variants add 10–15% computational overhead, remaining within acceptable bounds for production deployment at cooperating employment centers.

Communication requirements remain modest for federated scenarios. Parameter transmission involves approximately 60,000 weights for MLP architectures, translating to under 1 MB per training round. This bandwidth requirement supports deployment across standard internet connections available at Italian employment centers.

The system is currently entering pilot deployment at partner employment centers, which will provide empirical validation of these operational metrics under real-world conditions.

8.2 Regulatory compliance (GDPR, disability rights legislation)

Our privacy-preserving approach addresses GDPR Article 9 requirements for special category data handling through formal mathematical guarantees. Differential privacy with $\varepsilon = 1.0$, $\delta = 10^{-6}$ provides quantifiable protection against membership inference attacks on disability information. Regional data sovereignty ensures compliance with data localization requirements across European employment centers.

The federated architecture implements privacy-by-design principles mandated by GDPR. Personal disability data never leaves regional employment centers during collaborative training. Secure aggregation protocols prevent individual record reconstruction while enabling system-wide learning improvements through Shamir's 3-of-5 threshold secret sharing.

Audit trail capabilities support regulatory oversight through blockchain anchoring mechanisms. Merkle tree construction with PBKDF2-HMAC-SHA256 provides tamper-evident verification of employment matching decisions. Smart contract timestamps establish legally defensible evidence for compliance reporting and dispute resolution processes.

Italian Law 68/99 compliance benefits from improved matching accuracy and reduced bias in disability employment placement. The system's 90.1% F1-score enables more effective identification of suitable employment opportunities, supporting the law's

mandatory placement requirements while preserving candidate privacy through federated learning protocols.

Data subject rights enforcement remains feasible through the federated design. Individual access requests can be fulfilled locally without cross-regional data sharing. Deletion requests maintain effectiveness through model retraining protocols that remove specific candidate influence without compromising collaborative benefits.

8.3 Limitations and data structure constraints

Our evaluation uses synthetic data generated from authentic Italian employment center data structures provided by collaborating CPI and SIL institutions. The data schemas, feature relationships, and regional distributions reflect real employment service operations. However, specific candidate profiles, company details, and matching outcomes are synthetically generated to protect privacy during research and development phases.

The synthetic generation process preserves authentic statistical patterns while enabling reproducible research. Disability type distributions, geographic clustering, and employer accessibility features match observed patterns from partner employment centers. The compatibility scoring model incorporates feedback from employment counselors regarding actual decision factors in disability employment placement.

Regional heterogeneity simulation reflects authentic administrative boundaries and population distributions across Veneto provinces. However, local economic conditions, industry concentrations, and specialized disability services may exhibit greater variation than captured in our synthetic framework. The ongoing pilot deployment at partner centers will validate model performance across diverse regional characteristics.

The 500,000 candidate-employer pairs represent systematic enumeration of compatibility assessments rather than organic application patterns. Real-world matching involves temporal dynamics, seasonal employment fluctuations, and candidate self-selection that our current evaluation cannot fully capture. These factors will be incorporated through operational feedback during pilot deployment.

Class imbalance handling through SMOTE oversampling approximates employment success rates observed by partner centers. The 15–20% positive outcome rate aligns with successful placement statistics provided by cooperating employment services. However, success rates vary significantly across disability types and regional labor markets.

8.4 State-of-the-art comparison

Table 15 compares our results with recent privacy-preserving federated learning applications across sensitive domains.

Our employment matching system demonstrates several distinctive advantages over existing approaches. The 0.901 F1-score for LightGBM ensemble federation exceeds most federated learning results in sensitive applications while maintaining practical deployment requirements. The comprehensive evaluation across five performance metrics provides thorough assessment compared to single-metric evaluations common in related work.

The privacy framework combining differential privacy, Shamir's secret sharing, and blockchain anchoring provides stronger protection than single-method approaches prevalent in existing studies. The formal privacy guarantees ($\varepsilon = 1.0$, $\delta = 10^{-6}$) with

Table 15 Comparison with related work

Study	Domain	Architecture	Performance	Privacy method	Privacy budget
Our work	Employment Matching	LightGBM/MLP	F1: 0.901/0.788	DP + Shamir + Blockchain	$\varepsilon = 1.0$, $\delta = 10^{-6}$
Michalakopoulos et al. [16]	PV Forecasting	LSTM	MAPE: 15–25%	DP + FedAvg	Not specified
Liu et al. [22]	Power Systems	Unspecified	Accuracy: 95% +	DP Integration	Not specified
Piran et al. [37]	IoT Manufacturing	HDC	37% improvement	DP + XAI	Not specified
Wahida et al. [18]	Face Recognition	CNN	96.24% accuracy	AML + FL	Not specified
Yang et al. [23]	Byzantine FL	Unspecified	"Superior accuracy"	DP + PBFT	Diverse budgets
Zhao et al. [21]	General FL	Unspecified	4% improvement	DP + Client Selection	Adaptive
Chen and Wang [20]	Distributed Optimization	Survey	Theoretical	DP Survey	Various
Gardner et al. [38]	Metaverse ML	Review	Survey	Multiple methods	Various
Darzi et al. [29]	COVID-19 Detection	Various FL	Accuracy: ~ 70%	DP + FedAvg	Various
Tawfik et al. [34]	Healthcare Cybersecurity	Few-shot + XAI	F1: 0.99	FL + Cross-attention	$\varepsilon = 1.0$
Bonawitz et al. [27]	Secure Aggregation	Protocol	N/A (protocol)	Shamir + Pairwise	N/A

quantified noise calibration exceed the rigor of most comparative studies that report unspecified privacy parameters.

Our blockchain integration for long-term audit trails represents a novel contribution absent from other privacy-preserving federated learning implementations. The O(log n) Merkle proof system enables individual record verification without compromising aggregate privacy, addressing regulatory compliance requirements unique to sensitive social applications.

The minimal privacy cost (0.0001 F1-score degradation between classical and privacy-preserving federation) demonstrates superior privacy-utility balance compared to substantial accuracy losses typically reported in privacy-preserving machine learning literature.

8.5 Fairness considerations and explainability

Fairness evaluation represents a critical requirement for AI systems affecting disability employment outcomes. While our system claims to support "trustworthy AI," we acknowledge the absence of formal group-based fairness audits in the current evaluation. Comprehensive fairness assessment would examine: (1) equal opportunity metrics across disability types (motoria, intellettiva, sensoriale, psichica, DSA); (2) demographic parity across gender, age groups, and regional labor markets; (3) individual fairness ensuring similar candidates receive similar recommendations regardless of protected attributes.

The synthetic data generation process, while preserving realistic disability type distributions, cannot fully capture intersectional discrimination patterns present in actual employment outcomes. Real-world fairness evaluation during pilot deployment

will examine whether model recommendations exhibit disparate impact across disability severity levels or regional economic conditions. Mitigation strategies including reweighting, constraint-based optimization, and post-hoc calibration represent future work directions informed by deployment experience [34].

Explainability requirements for employment matching systems stem from both regulatory compliance (GDPR Article 22 right to explanation) and operational utility. Employment counselors require interpretable recommendations to maintain meaningful human oversight. Our current implementation provides feature importance rankings from LightGBM models, enabling counselors to understand which factors (geographic distance, skill compatibility, organizational accessibility) drive specific recommendations.

For privacy-preserving MLP models, post-hoc explanation methods (SHAP, LIME) can be applied locally at prediction time without compromising federated privacy guarantees. Recent work integrating explainable AI with federated learning [34] provides architectural patterns for deployment-time interpretability. The model output is presented to employment counselors as ranked candidate-opportunity lists with confidence scores and top contributing factors, supporting informed human decision-making rather than automated placement.

9 Conclusion

9.1 Summary of contributions

This research presents a comprehensive privacy-preserving federated learning framework for disability employment matching that addresses critical gaps in automated decision systems for sensitive social applications. Our key contributions include: (1) a novel federated learning architecture combining LightGBM ensemble methods with parameter-level MLP federation, achieving 90.1% F1-score while maintaining data locality; (2) a multi-layered privacy framework integrating differential privacy ($\varepsilon = 1.0$, $\delta = 10^{-6}$), Shamir's 3-of-5 secret sharing, and blockchain anchoring for comprehensive protection of disability-related data; (3) empirical validation across five Italian employment centers demonstrating minimal performance degradation (0.0005 F1-score loss) under federated constraints; and (4) production-ready implementation with sub-100 ms response times suitable for real-world deployment.

The technical contributions advance federated learning methodology through RDP-based noise calibration, dropout-resilient secure aggregation, and O(log n) blockchain verification systems. The privacy framework provides formal mathematical guarantees while maintaining practical utility for employment matching applications.

9.2 Practical impact for inclusive employment services

The system directly addresses employment inclusion challenges facing people with disabilities across European contexts. Current employment rates for individuals with disabilities remain 24.4 percentage points below general population levels, indicating substantial systemic barriers that automated matching systems can help overcome.

Our collaboration with Italian partner employment centers (CPI Villafranca di Verona, SIL Veneto) demonstrates immediate practical applicability. The 30–60 min manual processing time reduction to under 5 min enables employment counselors to serve more candidates while making more informed matching decisions. The 90.1% predictive

accuracy provides reliable support for placement decisions while preserving human oversight authority.

Regional federated learning enables knowledge sharing across employment centers without compromising data sovereignty requirements under GDPR and Italian Law 68/99. Employment centers can benefit from collaborative learning while maintaining complete control over sensitive disability and employer data.

The ongoing pilot deployment provides a pathway for broader adoption across European employment services. The lightweight computational requirements and standard infrastructure compatibility enable deployment at employment centers with limited technical resources.

9.3 Broader implications for privacy-preserving social AI

This work demonstrates that privacy-preserving machine learning can maintain practical utility for sensitive social applications requiring regulatory compliance and ethical considerations. The formal privacy guarantees combined with operational feasibility establish a model for responsible AI deployment in government and social service contexts.

The multi-layered privacy approach addresses diverse threat models relevant to social AI systems. Differential privacy provides statistical protection, secure aggregation prevents parameter inference, and blockchain anchoring enables long-term auditability required for public sector accountability.

The methodology extends beyond employment matching to other sensitive social applications including healthcare resource allocation, education placement systems, and social service delivery. The privacy-by-design architecture provides a template for federated learning deployment across institutions with conflicting data sharing constraints.

The research contributes to emerging frameworks for trustworthy AI in social contexts by demonstrating practical approaches to algorithmic fairness, privacy preservation, and regulatory compliance. The combination of technical rigor with operational validation provides actionable guidance for practitioners implementing AI systems affecting vulnerable populations.

Future applications may include federated learning for disability service coordination, cross-border employment mobility systems, and collaborative social policy research while maintaining individual privacy and institutional data sovereignty.

9.4 Future work: real-world validation requirements

Validation with authentic employment center data represents the immediate next phase through our ongoing pilot deployment. Collaboration with partner centers provides access to real candidate profiles, employer requirements, and observed placement outcomes. This validation addresses temporal dynamics, seasonal employment patterns, and evolving disability service needs not captured in synthetic evaluation.

Multi-country deployment testing will examine model transferability across different European disability legislation frameworks. Employment systems vary significantly in mandatory quotas (Italy's Law 68/99 requires 7% for large employers versus France's 6%), accommodation requirements, and assessment procedures. Federated learning effectiveness across diverse regulatory environments requires empirical validation beyond the Italian context.

Long-term impact assessment needs longitudinal studies tracking employment outcomes for disability candidates matched through our system. Success metrics should include job retention rates (currently 60–70% at 6 months in partner centers), career advancement opportunities, and candidate satisfaction measures beyond initial placement accuracy.

Bias evaluation requires systematic testing across disability types (motoria, intellettiva, sensoriale per Italian classifications), severity levels, and demographic characteristics. Employment matching systems risk perpetuating historical discrimination patterns if not carefully monitored. Fairness metrics specifically designed for disability employment contexts need development and empirical validation.

Scalability testing must examine performance with hundreds of employment centers and thousands of active candidates. Current evaluation involves five regional partitions, while national deployment might require coordination across Italy's 550 + employment centers and similar systems across European Union member states.

Integration studies will evaluate compatibility with existing employment center software systems (SIUL, GePI platforms used across Italy), candidate management databases, and employer relationship platforms. Technical integration challenges often determine deployment success more than algorithmic performance in public sector implementations.

Fairness auditing requires systematic evaluation across protected groups. Future work will implement group-based metrics (equalized odds, demographic parity, individual fairness) stratified by disability type, severity level, gender, age, and region. Mitigation strategies including adversarial debiasing, constraint-based optimization, and calibrated equalized odds will be evaluated against deployment-specific fairness requirements defined in collaboration with disability rights organizations.

Byzantine-robust aggregation represents an important extension for deployment scenarios involving less trusted participants. While our current semi-honest threat model is appropriate for government employment centers with established accountability, expansion to federated consortiums including private employers may require Byzantine-resilient protocols. Future work will evaluate coordinate-wise median, Krum, and trimmed mean aggregation under simulated poisoning attacks [28, 30], establishing empirical resilience bounds for adversarial participation rates up to $f < n/3$. Human-in-the-loop evaluation will assess employment counselor acceptance and system usability through structured feedback collection at partner centers. Training requirements, interface design, and decision support presentation require empirical validation with employment service professionals who will operate the system daily.

Acknowledgements
We acknowledge the collaboration and support of CPI Villafranca di Verona and SIL Veneto for providing authentic data structures and domain expertise. We thank Università eCampus for institutional support. Special appreciation to employment counselors who provided practical insights into disability employment matching processes.

Author contributions
Conceptualization, O.K.; Methodology, O.K., M.M. and A.G.; Software, M.M.; Validation, M.M. and A.G.; Formal Analysis, M.M.; Investigation, O.K. and M.M.; Resources, E.F. and M.A.; Data Curation, M.M.; Writing—Original Draft, O.K. and M.M.; Writing—Review & Editing, A.G., E.F., and M.A.; Visualization, M.M.; Supervision, E.F.; Project Administration, M.A.; Funding Acquisition, E.F. All authors have read and agreed to the published version of the manuscript.

Funding
This research was supported by institutional funding from Università eCampus. No external grants or commercial funding influenced the research design, data collection, analysis, or manuscript preparation.

Data availability
The synthetic datasets generated for this study are available in the project repository at https://github.com/KuznetsovKarazin/disability-job-matching-system . Authentic employment center data cannot be shared due to privacy regulations and institutional agreements. The complete source code, experimental configurations, and trained models are publicly available under academic license.

Code availability
All source code, experimental scripts, and analysis tools are available at https://github.com/KuznetsovKarazin/disability-job-matching-system. The repository includes complete installation instructions, configuration examples, and documentation for reproducing all reported results.

Declarations

Ethics approval and consent to participate
This research was conducted in accordance with European data protection regulations and ethical guidelines for AI systems in social applications. All synthetic data generation procedures were designed to prevent potential harm to individuals with disabilities. The federated learning framework preserves data locality and individual privacy while enabling collaborative improvement of employment services.

Competing interests
The authors declare no competing financial or non-financial interests related to this research. The author Alessandro Galdelli is a member of the Editorial board of this journal.

Received: 9 October 2025 / Accepted: 9 February 2026
Published online: 25 February 2026

References

1. Micangeli A, Puglisi A, Vignola R. Report on the Employment of Disabled People in European Countries: Italy. Brussels: Academic Network of European Disability Experts (ANED); 2008.
2. Buchanan J, Hammersley H, Uldry M, Couceiro Á, Moledo A. The Right to Work: The employment situation of persons with disabilities in Europe. Brussels: European Disability Forum; 2023.
3. Abid M, Ben-Salha O, Gasmi K, Alnor NHA. Modelling for disability: how does artificial intelligence affect unemployment among people with disability? An empirical analysis of linear and nonlinear effects. Res Dev Disabil. 2024;149:104732. https://doi.org/10.1016/j.ridd.2024.104732.
4. Rocha TdeO, Alonso CMdoC, Silva TNRda. Frontline inclusion workers' perceptions of employment barriers for people with disabilities in Brazil: advancing inclusion to leave no one behind. Appl Ergon. 2025;129:104584. https://doi.org/10.1016/j.apergo.2025.104584.
5. Agovino M, Marchesano K, Garofalo A. Policies based on mandatory employment quotas for disabled workers: the case of Italy. Mod Italy. 2019;24:295–315. https://doi.org/10.1017/mit.2019.14.
6. Khan H, Sohail SS, Madsen DØ, Nafis MT. Deep learning in disability research: a bibliometric analysis. Digit Eng. 2025;6:100046. https://doi.org/10.1016/j.dte.2025.100046.
7. Zhuang KV, Goggin G. New possibilities or problems for disability and inclusion? The case of AI and ADMs across work. Telemat Informatics. 2024;92:102156. https://doi.org/10.1016/j.tele.2024.102156.
8. Lukács A, Váradi S. GDPR-compliant AI-based automated decision-making in the world of work. Comput Law Secur Rev. 2023;50:105848. https://doi.org/10.1016/j.clsr.2023.105848.
9. Rosenberger J, Wolfrum L, Weinzierl S, Kraus M, Zschech P. CareerBERT: matching resumes to ESCO jobs in a shared embedding space for generic job recommendations. Expert Syst Appl. 2025;275:127043. https://doi.org/10.1016/j.eswa.2025.127043.
10. Wang Y. Design and implementation of student job matching system based on personalized recommendation algorithm. Syst Soft Comput. 2025;7:200302. https://doi.org/10.1016/j.sasc.2025.200302.
11. Sainju B, Hartwell C, Edwards J. Job satisfaction and employee turnover determinants in Fortune 50 companies: insights from employee reviews from Indeed.com. Decis Support Syst. 2021;148:113582. https://doi.org/10.1016/j.dss.2021.113582.
12. Sallach T, Mönke FW, Schäpers P. Cybervetting of organizational citizenship behavior expectations: profile summary as a key in LinkedIn-based assessments. Comput Human Behav. 2024;154:108113. https://doi.org/10.1016/j.chb.2023.108113.
13. Broda MD, Bogenschutz M, Dinora P, Prohn S, Lineberry S, West A. Understanding COVID-19 infection among people with intellectual and developmental disabilities using machine learning. Disabil Health J. 2024;17:101607. https://doi.org/10.1016/j.dhjo.2024.101607.
14. Chen C, Zhang W, Pan Y, Li Z. An interpretable hybrid machine learning approach for predicting three-month unfavorable outcomes in patients with acute ischemic stroke. Int J Med Inform. 2025;196:105807. https://doi.org/10.1016/j.ijmedinf.2025.105807.
15. Hoffman H, Wood J, Cote JR, Jalal MS, Otite FO, Masoud HE, et al. Development and internal validation of machine learning models to predict mortality and disability after mechanical thrombectomy for acute anterior circulation large vessel occlusion. World Neurosurg. 2024;182:e137–54. https://doi.org/10.1016/j.wneu.2023.11.060.
16. Michalakopoulos V, Sarantinopoulos E, Sarmas E, Marinakis V. Empowering federated learning techniques for privacy-preserving PV forecasting. Energy Rep. 2024;12:2244–56. https://doi.org/10.1016/j.egyr.2024.08.033.
17. Smajić A, Grandits M, Ecker GF. Privacy-preserving techniques for decentralized and secure machine learning in drug discovery. Drug Discov Today. 2023;28:103820. https://doi.org/10.1016/j.drudis.2023.103820.
18. Wahida F, Chamikara MAP, Khalil I, Atiquzzaman M. An adversarial machine learning based approach for privacy preserving face recognition in distributed smart city surveillance. Comput Netw. 2024;254:110798. https://doi.org/10.1016/j.comnet.2024.110798.

19. Kakandwar S, Bhushan B, Kumar A. Chapter 3—Integrated machine learning techniques for preserving privacy in Internet of Things (IoT) systems. In: Bhushan B, Sharma SK, Saračević M, Boulmakoul A. (eds.) Blockchain technology solutions for the security of IoT-based healthcare systems. pp. 45–75. Academic Press (2023). https://doi.org/10.1016/B978-0-323-99199-5.00012-4.
20. Chen Z, Wang Y. Privacy-preserving distributed optimization and learning. In: Ding Z. (ed.) Encyclopedia of systems and control engineering (first edition). pp. 308–324. Elsevier, Oxford (2026). https://doi.org/10.1016/B978-0-443-14081-5.00125-2.
21. Zhao X, Li G, Yao Y, Cui B. Balancing privacy and fairness: client selection in differential privacy-based federated learning. J Syst Archit. 2025;168:103576. https://doi.org/10.1016/j.sysarc.2025.103576.
22. Liu Z, Ma J, Gong X, Liu X, Liu B, An L. Differential privacy integrated federated learning for power systems: an explainability-driven approach. Comput Mater Continua. 2025;85:983–99. https://doi.org/10.32604/cmc.2025.065978.
23. Yang W, Xu X, Yu K, Li G. Byzantine-resilient federated learning with dynamic scoring matrix and variant PBFT consensus under differential privacy. Inf Sci. 122682 (2025). https://doi.org/10.1016/j.ins.2025.122682.
24. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Singh A, Zhu X (Jerry) (eds.) Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20–22 April 2017, Fort Lauderdale, FL, USA. pp. 1273–1282. PMLR (2017).
25. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, Zhang L. Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. pp. 308–318. Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2976749.2978318.
26. Mironov I. Rényi Differential Privacy. Presented at the 2017 IEEE 30th Computer Security Foundations Symposium (CSF) August 1 (2017). https://doi.org/10.1109/CSF.2017.11.
27. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, Ramage D, Segal A, Seth K. Practical secure aggregation for privacy-preserving machine learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. pp. 1175–1191. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3133956.3133982.
28. Darzi E, Sijtsema NM, van Ooijen P. Weight-space noise for privacy-robustness trade-offs in federated learning. Neural Comput Appl. 2025;37:19687–705. https://doi.org/10.1007/s00521-025-11420-1.
29. Darzi E, Sijtsema NM, van Ooijen PMA. A comparative study of federated learning methods for COVID-19 detection. Sci Rep. 2024;14:3944. https://doi.org/10.1038/s41598-024-54323-2.
30. Darzi E, Dubost F, Sijtsema NM, van Ooijen PMA. Exploring adversarial attacks in federated learning for medical imaging. IEEE Trans Ind Inform. 2024;20:13591–9. https://doi.org/10.1109/TII.2024.3423457.
31. Darzi E, Shen Y, Ou Y, Sijtsema NM, van Ooijen PMA. Tackling heterogeneity in medical federated learning via aligning vision transformers. Artif Intell Med. 2024;155:102936. https://doi.org/10.1016/j.artmed.2024.102936.
32. Darzi E, Marx A. Structured robustness for distribution shifts. presented at the workshop on spurious correlation and shortcut learning: foundations and solutions March 6 (2025).
33. Darzidehkalani E, Ghasemi-rad M, van Ooijen PMA. Federated learning in medical imaging: Part II: methods, challenges, and considerations. J Am Coll Radiol. 2022;19:975–82. https://doi.org/10.1016/j.jacr.2022.03.016.
34. Tawfik M, Abu-Ein AA, Noaman HM, Abdelhaliem AH, Fathi IS. FedMedSecure: federated few-shot learning with cross-attention mechanisms and explainable AI for collaborative healthcare cybersecurity. Sci Rep. 2025;15:40050. https://doi.org/10.1038/s41598-025-25107-z.
35. Tawfik M, Abdelhaliem AH, Fathi I. Quantum-resistant privacy-preserving IoT authentication via zero-knowledge proofs and blockchain integration. Stat Optim Inf Comput. 2025;14:1374–402. https://doi.org/10.19139/soic-2310-5070-2399.
36. Al-madni AM, Ying X, Tawfik M, Ahmed ZAT. An optimized blockchain model for secure and efficient data management in internet of things. In: 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS). pp. 1–11 (2024). https://doi.org/10.1109/ICITEICS61368.2024.10624817.
37. Piran FJ, Chen Z, Imani M, Imani F. Privacy-preserving federated learning with differentially private hyperdimensional computing. Comput Electr Eng. 2025;123:110261. https://doi.org/10.1016/j.compeleceng.2025.110261.
38. Gardner SH, Hoang T-M, Na W, Dao N-N, Cho S. Metaverse meets distributed machine learning: a contemporary review on the development with privacy-preserving concerns. ICT Express. 2025;11:507–22. https://doi.org/10.1016/j.icte.2025.04.008.

Publisher's Note