Tracking Evolving Anomalies with Encrypted Federated Continual Learning

Zarka Bashir[1,2]
[1]*Center for AI and ML*, *IDRBT*, Hyderabad, India
[2]*Department of CSE*, *IIT*, Hyderabad, India
cs21resch11010@iith.ac.in

Mridula Verma[1]
[1]*Center for AI and ML*, *IDRBT*, Hyderabad, India
vmridula@idrbt.ac.in

C Krishna Mohan[2]
[2]*Department of CSE*, *IIT*, Hyderabad, India
ckm@cse.iith.ac.in

***Abstract*—Industrial Anomaly Detection (IAD) identifies deviations from normal operations in industrial systems, commonly using visual inspection data captured by camera sensors to monitor product quality. In practice, inspection data is generated across multiple industrial sites with evolving anomaly streams posing challenges in data centralization due to privacy constraints, cross-site generalization due to siloed deployments, and continual adaptation due to catastrophic forgetting. Federated continual learning (FCL) enables IAD systems to learn from distributed industrial data silos and adapt to evolving anomaly patterns without sharing raw inspection data. However, plaintext update exchanges remain vulnerable to gradient inversion and model extraction attacks. We propose Federated Approximate Gradient Matching (FedAGM), a fully homomorphic encryption (FHE)-native FCL framework for unsupervised IAD. FedAGM performs server-side aggregation of encrypted client gradients and aligns them via an FHE-compatible mechanism that approximates non-polynomial federated gradient matching operations with low-degree polynomials. We provide a provable error bound for this approximation, ensuring directional gradient alignment while preserving confidentiality across heterogeneous sites. Experiments on the benchmark MVTec-AD demonstrate that FedAGM achieves performance comparable to SOTA FCL methods, with enhanced privacy.**

***Index Terms*—Federated Continual Learning, Unsupervised Anomaly Detection, Fully Homomorphic Encryption.**

I. Introduction

Industrial Anomaly Detection (IAD) identifies industrial anomalies, such as equipment faults or surface defects, to ensure better product quality and operational safety [1], [2]. Since anomalies are rare and labeled data is costly, Unsupervised AD (UAD) models normal behavior from unlabeled data to flag deviations [1]–[3]. While multi-site collaborative learning enhances generalization, privacy regulations prohibit sharing raw data [4]. Furthermore, evolving anomaly patterns across categories necessitate continual adaptation without catastrophic forgetting (CF). These constraints thwart centralized training and limit scalability. Federated Continual Learning (FCL) [5] mitigates these challenges by integrating Federated Learning (FL) [4] and Continual Learning (CL) [6], [7], enabling privacy-preserving, collaborative updates that adapt to evolving industrial environments.

Our FCL-enabled IAD framework (FCL-IAD) utilizes a central server to aggregate updates from distributed sites training on non-stationary streams, facilitating cross-site knowledge transfer, adaptation to emerging faults and retain previously learned behaviors [8], [9]. To mitigate CF without the storage overhead or privacy risks of traditional replay-based methods [10], [11], we employ orthogonal gradient projection [6], [7], constraining updates to preserve prior knowledge. To handle task disjointness and cross-client distributional heterogeneity, we propose a server-side Gradient Matching (GM) strategy that aligns client gradients to mitigate negative transfer, ensuring stable global updates and improved knowledge consolidation across heterogeneous industrial tasks [12], [13].

Despite these advantages, existing FCL methods are vulnerable to privacy attacks such as gradient inversion, gradient leakage, and inference attacks due to sharing of raw gradients or model updates [14], [15]. Although Differential Privacy (DP), Secure Multi-Party Computation (SMPC), and Trusted Execution Environments (TEEs) mitigate these risks, they struggle in industrial IAD: DP may obscure critical rare anomaly signals [16], SMPC adds significant overhead, and TEEs face hardware-reliance and side-channel attacks [17]. In contrast, Fully Homomorphic Encryption (FHE) provides stronger privacy guarantees while preserving exact gradients during aggregation, ideally addressing deployment constraints of FCL-IAD, which remains underexplored in the literature.

To fill this research gap, we propose Federated Approximate Gradient Matching (FedAGM), an Fully Homomorphic Encryption (FHE) [18], [19] enhanced FCL-IAD framework for privacy-preserving continual IAD across heterogeneous sites. FedAGM enables the server to aggregate encrypted model gradients from sites directly, without decryption, providing strong end-to-end privacy guarantees throughout federated training [17], [20]. The aggregation in FedAGM assigns higher weights to encrypted client updates that align with the global gradient, reinforcing consistent updates, while down-weighting conflicting updates to maintain alignment across heterogeneous clients.

Existing FCL methods address cross-client heterogeneity using non-polynomial mechanisms [5], [8], [9], [12], which are incompatible with FHE [18]. FedAGM employs an FHE-compatible mechanism for server-side aggregation utilized in

FCL-IAD that aligns encrypted model gradients shared by clients by approximating standard non-linear GM operations, such as Euclidean norm scaling, and coefficient normalization, with low-degree polynomials, thereby preserving the geometric properties of conventional GM [21]. We further provide a provable error bound that quantifies the effect of polynomial and coefficient approximations by measuring the deviation of the encrypted aggregated gradient from the ideal global gradient. This bound ensures that the two remain close, thereby preserving directional alignment and ensuring stable model updates under FHE. Our main contributions are:

- We propose FedAGM, an FHE-enabled FCL framework that enables privacy-preserving continual adaptation to evolving industrial anomalies across distributed sites. It employs a server-side encrypted gradient aggregation mechanism that aligns updates across heterogeneous clients while keeping data confidential.
- We use polynomial approximations for non-linear GM operations and provide a theoretical bound on the induced approximation error.
- We evaluate FedAGM on the MVTec-AD dataset[1], demonstrating comparable performance to FCL baselines while ensuring strong privacy via FHE aggregation.

II. Related Work

a) Unsupervised Continual Industrial Anomaly Detection: UCAD equips UAD with CL via contrastively-learned prompts and memory banks to preserve task-invariant normal features [2]. Diffusion-based methods such as DiAD and CDAD address multi-class and continual anomaly detection via latent diffusion and gradient projection respectively [1], [3], while ControlNet requires gradient-based stabilization to prevent CF [22]. However, distributed and evolving data across multiple silos remain challenging for robust continual IAD.

b) Federated Continual Learning (FCL): FCL enables clients to collaboratively adapt to non-stationary streams while mitigating catastrophic forgetting across heterogeneous tasks [5], [8], [23]. Existing methods use replay [10], synaptic intelligence [8], parameter decomposition [5], and other mechanisms [9], [11], [24], but remain vulnerable to gradient leakage. Server-side gradient matching [12], [13] improves alignment across heterogeneous clients, but operates on plaintext gradients and relies on non-polynomial operations.

c) Privacy-Preserving FL: In FHE-based FL [17], [20], [25], clients send encrypted updates aggregated on the server without exposing plaintext, matching standard FL performance while ensuring strong privacy. The CKKS scheme [19] efficiently supports approximate real-number computation, making it well-suited for encrypting gradients [17]. Recent work [26] explores FHE-compatible CL via gradient projection memory, but FHE has not yet been applied to FCL, as cross-client heterogeneity typically requires non-polynomial operations incompatible with current FHE schemes.

[1]https://www.mvtec.com/company/research/datasets/mvtec-ad

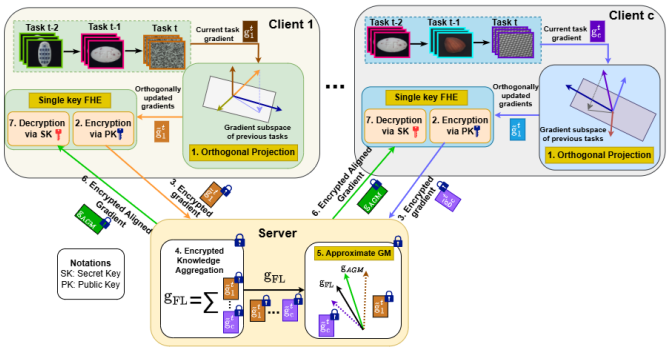

Fig. 1. Proposed framework for FedAGM.

III. Preliminaries

We consider a privacy-sensitive FCL for unsupervised IAD, where multiple clients sequentially observe a private stream of unlabeled tasks without task identifiers or access to past data. Let $\mathcal{C}$ denote the set of c clients, where each client receives a private sequence of unlabeled datasets $T_c^1, T_c^2, \ldots, T_c^t$, with each task corresponding to learning anomalies in a distinct object category. To simulate real-world heterogeneity, task sequences are sampled such that no two clients share the same task. An honest-but-curious server or potentially malicious clients may attempt to infer sensitive data from shared gradients, given knowledge of the model architecture and training procedure and possible auxiliary information about the data distribution.

IV. Proposed Method

Our proposed method, FedAGM, combines client-side orthogonal projection with FHE-enabled, gradient-invariant server-side aggregation to ensure data privacy. Clients compute, project, and encrypt local gradients, which the server aggregates to produce consistent global updates without accessing plaintext information. The overall framework is illustrated in Figure 1.

a) Client-Side Orthogonal Gradient Projection: Following [3], we project training gradients onto the orthogonal (null) space spanned by prior task inputs, ensuring that updates do not disrupt previously acquired knowledge while preserving model stability. Specifically, at task T_c^t in round r, client c computes the gradient of the unsupervised reconstruction loss $\mathcal{L}_{c,\text{rec}}^{(t,r)}$ on its local dataset $g_c^{(t,r)} = \nabla_\theta \mathcal{L}_{c,\text{rec}}^{(t,r)}(\theta_c^{(t,r)})$, where $\theta_c^{(t,r)}$ are the local autoencoder parameters prior to training on task T_c^t.

To preserve prior anomalies, each client maintains a compact low-rank subspace ($U_c^{(t-1,r)} \in \mathbb{R}^{p\times k}$) that captures k gradient directions important for preserving previously learned normal patterns. After task $t-1$, representative gradients are collected to form an orthonormal basis $\mathbf{U}_c^{(t-1,r)} = \text{span}\{\mathbf{u}_{c,1}^{(t-1,r)}, \ldots, \mathbf{u}_{c,k}^{(t-1,r)}\}$, where each column $\mathbf{u}_{c,i}^{(t-1,r)}$ represents a gradient direction in the subspace, p is the number of model parameters, k is the subspace rank, and $k \ll p$.

During training on task T_c^t in round r, the local gradient $g_c^{(t,r)}$ is projected onto the orthogonal complement of this subspace:

$$g'^{(t,r)}_c = \left(\mathbf{I} - \mathbf{U}_c^{(t-1,r)}(\mathbf{U}_c^{(t-1,r)})^\top\right) g_c^{(t,r)}. \quad (1)$$

This projection removes gradient components aligned with directions critical for prior tasks, ensuring updates do not interfere with previously learned representations while preserving the ability to learn new anomalies, i.e., $(\mathbf{U}_c^{(t-1,r)})^\top g_c'^{(t,r)} = \mathbf{0}$ so the update lies in the null space of important previous directions. The local model is updated using the projected gradient, $\theta_c^{(t,r)} = \theta_c^{(t-1,r)} - \eta_c\, g_c'^{(t,r)}$, where η_c denotes the client learning rate. After completing training on task $\mathcal{T}_c^{(t,r)}$, the client updates its memory subspace to incorporate information from the new task:

$$\mathbf{U}_c^{(t,r)} = \mathrm{SVD}_k\left(\left[\mathbf{U}_c^{(t-1,r)},\ g_c'^{(t,r)}\right]\right), \quad (2)$$

where, $\mathrm{SVD}_k(\cdot)$ retains only the top-k singular vectors.

While orthogonal projection mitigates temporal forgetting, it does not address the geometric bias caused by non-IID task distributions across clients. To alleviate this, we enforce hyperspherical uniformity to evenly distribute latent embeddings and apply variance regularization on the encoder's latent representations to prevent feature collapse [27]. The local objective for client c at task T_c^t becomes:

$$\mathcal{L}_c^t = \mathcal{L}_{\text{rec}}^{(t,r)} + \lambda_u\, \mathcal{L}_{\text{uni}}^{(t,r)} + \lambda_v\, \mathcal{L}_{\text{var}}^{(t,r)}, \quad (3)$$

where $\mathcal{L}_{\text{uni}}^{(t,r)}$ promotes hyperspherical uniformity, and $\mathcal{L}_{\text{var}}^{(t,r)}$ enforces variance across latent dimensions. Hyperparameters λ_u and λ_v balance these regularization terms, ensuring stable and informative embeddings (Summarized in Algorithm 1).

b) Approximate Gradient Matching under FHE: To protect local IAD models from leaking sensitive information, each client encrypts its projected gradients using FHE before sharing with the server [18], [19]. Clients jointly generate a CKKS key pair (pk, sk), sharing the public key pk with the server while keeping the secret key sk private [19]. After computing the projected gradient $g_c'^t$, each client encrypts it as $\tilde{g}_c^{(t,r)} = \text{Enc}_{\text{pk}}(g_c'^{(t,r)})$, and sends it to the server. The server performs aggregation directly in the encrypted domain, $\tilde{g}_{\text{FL}}^{(t,r)} = \sum_{c\in\mathcal{C}_S} \tilde{g}_c^{(t,r)}$, where $\mathcal{C}_S$ is the set of clients participating in round r. It then updates the encrypted global model as $\tilde{\theta}^{(G)} \leftarrow \tilde{\theta}^{(G)} - \eta_G \tilde{g}_{AGM}^{(t,r)}$, where η_G is the global learning rate and FHE-compatible invariant gradient $\tilde{g}_{AGM}^{(t,r)}$. The updated encrypted model is returned to clients for collective decryption, $\theta^{(G)} = \text{Dec}_{\text{sk}}(\tilde{\theta}^{(G)})$. Thus, the server never observes plaintext gradients or model updates, preserving privacy throughout FCL.

FedAGM aligns encrypted client updates toward a consistent direction, promoting shared, task-invariant representations while relying only on addition and multiplication operations. Let $\Gamma = \{\gamma_c^{(r)} \mid c \in \mathcal{C}_S \quad \sum_{c\in\mathcal{C}_S} \gamma_c^{(r)} = 1\}$ denote the set of learnable aggregation coefficients at round r and $\mathbf{g}^{(t,r)} = \{g_1^{(t,r)}, \ldots, g_{|\mathcal{C}_S|}^{(t,r)}\}$ the set of client gradients at task

Algorithm 1 FedAGM: Client-side Orthogonal Projection

Server: Initialize global model $\theta^{(G)}$
Clients: Initialize $\theta_c^{(t,0)} \leftarrow \theta^{(G)}$, subspace $\mathbf{U}_c^{(t,0)} \leftarrow \emptyset$; jointly generate (pk, sk) and share pk with server
Server encrypts and broadcasts model: $\tilde{\theta}^{(G)} = \text{Enc}_{\text{pk}}(\theta^{(G)})$ to $\mathcal{C}_S$
for all clients $c \in \mathcal{C}_S$ **in parallel do**
  Decrypt global model locally: $\theta_c^{(t,0)} = \text{Dec}_{\text{sk}}(\tilde{\theta}^{(G)})$
  for task $t = 1$ to T **do**
    for round $r = 1$ to R **do**
      Compute local gradient $g_c^{t,r}$ ▷ Eq. 3
      if $t > 1$ **then**
        Project gradient ▷ Eq. 1
      else
        $g_c'^{t,r} = g_c^{t,r}$
      end if
      Update local model: $\theta_c^{(t,r)} = \theta_c^{(t-1,r)} - \eta_c g_c'^{t,r}$
    end for
    Update subspace (Eq. 2), encrypt & send to server
    Server-side aggregation ▷ Algorithm 2
    Server updates $\tilde{\theta}^{(G)} = \tilde{\theta}^{(G)} - \eta_G g_{AGM}^{(t,r)}$
  end for
end for

$T_c^{(t)}$ and round r. In conventional GM, the Invariant Gradient Direction (IGD) is defined as [13]:

$$g_{IGD}^{(t,r)} = g_{FL}^{(t,r)} + \kappa \frac{\|g_{FL}^{(t,r)}\|}{\|\Gamma^\star \mathbf{g}^{(t,r)}\|} \Gamma^\star \mathbf{g}^{(t,r)},$$
$$\text{s.t.} \quad \Gamma^\star = \arg\min_{\Gamma} \left(\Gamma \mathbf{g}^{(t,r)} \cdot g_{FL}^{(t,r)} + \kappa \|g_{FL}^{(t,r)}\| \, \|\Gamma \mathbf{g}^{(t,r)}\|\right). \quad (4)$$

where $g_{FL}^{(t,r)}$ is the aggregated gradient, κ scales the contribution of aligned client gradients, and $\Gamma^\star$ denotes the optimal gradient matching coefficients. Intuitively, GM computes a convex combination of client gradients that aligns closely with the global direction while balancing magnitude contributions, promoting task-invariant updates across heterogeneous clients. To enable this under FHE, we quantify the directional alignment between each client's encrypted gradient $\tilde{g}_c^{(t,r)}$ and the aggregated global gradient $\tilde{g}_{FL}^{(t,r)}$ using the dot product

$$\tilde{s}_c^{(t,r)} = \tilde{g}_c^{(t,r)} \cdot \tilde{g}_{FL}^{(t,r)}. \quad (5)$$

A low-degree polynomial $\text{Poly}(\cdot)$ maps these alignment scores $\gamma_c^{t,r}$ to surrogate aggregation weights:

$$\tilde{\gamma}_c^{(t,r)} = \text{Poly}(\tilde{s}_c^{(t,r)}) = \alpha_0 + \alpha_1 \tilde{s}_c^{(t,r)} + \alpha_2 (\tilde{s}_c^{(t,r)})^2 + \alpha_3 (\tilde{s}_c^{(t,r)})^3 + \cdots + \alpha_d (\tilde{s}_c^{(t,r)})^d, \quad (6)$$

where the degree d balances approximation accuracy and FHE efficiency [19]. Gradients more aligned with the global direction receive higher surrogate coefficients $\tilde{\gamma}_c^{(t,r)}$, whereas conflicting gradients are downweighted, ensuring encrypted gradients maintain directional alignment across heterogeneous

Algorithm 2 Server-side Approximate Gradient Matching

Require: Encrypted client gradients $\{\tilde{g}_c^{(t,r)}\}_{c\in\mathcal{C}_S}$, polynomial degree d, scaling factor κ

Ensure: FHE-compatible aggregated invariant gradient $g_{AGM}^{(t,r)}$

Compute the encrypted global reference gradient $\tilde{g}_{FL}^{(t,r)}$
for all $c \in \mathcal{C}_S$ **do**
    Compute directional alignment score $\tilde{s}_c^{(t,r)}$ ▷ Eq. (5)
    Compute surrogate aggregation weight $\tilde{\gamma}_c^{(t,r)}$ using polynomial approximation ▷ Eq. (6)
end for
Compute approximated weighted gradient $\tilde{g}_\Gamma^{(t,r)}$ ▷ Eq. (7)
Compute FHE-compatible invariant gradient $g_{AGM}^{(t,r)}$ using Chebyshev polynomials ▷ Eq. (8)
return $g_{AGM}^{(t,r)}$

clients. The aggregated, globally aligned encrypted gradient is then computed as

$$\tilde{g}_\Gamma^{(t,r)} = \sum_{c\in\mathcal{C}_S} \tilde{\gamma}_c^{(t,r)} \tilde{g}_c^{(t,r)}. \tag{7}$$

Conventional GM also requires a scaling factor that involves division and square-root operations, $\frac{\|g_{FL}^{(t,r)}\|}{\|\Gamma^\star \mathbf{g}^{(t,r)}\|}$, which are not compatible with FHE. In FedAGM, we decompose this factor into a known magnitude and an inverse norm, and approximate both using low-degree Chebyshev polynomials [21]: $P_K(x) \approx \sqrt{x}, \quad Q_L(x) \approx \frac{1}{\sqrt{x}}$. This yields the FHE-compatible invariant gradient (See Algorithm 2):

$$g_{AGM}^{(t,r)} = \tilde{g}_{FL}^{(t,r)} + \kappa\, P_K(\|g_{FL}^{(t,r)}\|^2)\, Q_L(\|\tilde{g}_\Gamma^{(t,r)}\|^2)\, \tilde{g}_\Gamma^{(t,r)} \tag{8}$$

We provide a provable error bound for FedAGM in Lemma 1, showing that the encrypted aggregated gradient remain close to the ideal global representation.

Lemma 1 (AGM Alignment Guarantee)**.** *Assume all client gradients are bounded, $\|g_c^{(t,r)}\| \le G$ for all $c \in \mathcal{C}_S$, let $g_{FL}^{(t,r)}$, and $g_{IGD}^{(t,r)}$ be defined as in Equation* (4)*. Let κ be the scaling factor, and $\Gamma^\star$ denote the optimal GM coefficients. Define FHE-compatible surrogate coefficients $\tilde{\gamma}_c^{(t,r)}$ and $\tilde{g}_\Gamma^{(t,r)}$ as in Equations* (5)-(7)*. Then the FHE-compatible invariant gradient $g_{AGM}^{(t,r)}$ defined in Equation* (8) *satisfies*

$$\begin{aligned} g_{AGM}^{(t,r)} &= g_{IGD}^{(t,r)} + \Delta^{(t,r)}, \ \textit{where} \\ \|\Delta^{(t,r)}\| &\le \kappa\, P_K\Big(\|g_{FL}^{(t,r)}\|^2\Big)\, Q_L\Big(\|\tilde{g}_\Gamma^{(t,r)}\|^2\Big) \\ &\quad \times \sum_{c\in\mathcal{C}_S} \big|\tilde{\gamma}_c^{(t,r)} - \gamma_c^{\star(t,r)}\big|\, \|g_c^{(t,r)}\| \\ &\quad + \kappa\, \varepsilon_K \Big\| \sum_{c\in\mathcal{C}_S} \gamma_c^{\star(t,r)} g_c^{(t,r)} \Big\| \end{aligned} \tag{9}$$

If $\mathrm{Poly}(\cdot)$ *is monotone increasing, the directional ordering of client contributions is preserved: gradients more aligned with $g_{FL}^{(t,r)}$ receive higher weights.*

Proof. FHE preserves addition and multiplication, so we work in the plaintext domain. Let $\Delta^{(t,r)} = g_{AGM}^{(t,r)} - g_{IGD}^{(t,r)}$. Subtracting (4) from (8) and adding and subtracting $\kappa P_K(\|g_{FL}\|^2) Q_L(\|\tilde{g}_\Gamma\|^2) \sum_c \gamma_c^\star g_c$ decomposes the deviation into two terms, one from surrogate coefficient error and one from norm-approximation error:

$$\begin{aligned} \Delta^{(t,r)} &= \kappa P_K Q_L \sum_c \big(\tilde{\gamma}_c - \gamma_c^\star\big) g_c \\ &\quad + \kappa \Big[P_K Q_L - \tfrac{\|g_{FL}\|}{\|\Gamma^\star \mathbf{g}\|}\Big] \sum_c \gamma_c^\star g_c, \end{aligned}$$

where superscripts (t,r) are suppressed for brevity. Applying the triangle inequality to the first term and bounding $\|\sum_c (\tilde{\gamma}_c - \gamma_c^\star) g_c\| \le \sum_c |\tilde{\gamma}_c - \gamma_c^\star| \|g_c\|$, and bounding the norm-approximation factor in the second term by the uniform Chebyshev approximation error ε_K [21] yields (9). Finally, monotonicity of $\mathrm{Poly}(\cdot)$ gives $\tilde{s}_{c_1} > \tilde{s}_{c_2} \Rightarrow \tilde{\gamma}_{c_1} > \tilde{\gamma}_{c_2}$, preserving the directional ordering of client contributions. □

V. Experimentation Setup and Result Analysis

A. Experimentation Setup

We utilize the *MVTec-AD* benchmark comprising 5,354 images across 15 industrial object and texture categories. Training sets contain only normal samples, while test sets include both normal and defective images with real-world anomalies (e.g., scratches, contamination). To simulate FCL, the 15 categories are partitioned into sequential, disjoint tasks following [12]. We consider four standard configurations: 14-1, 10-5, 3×5, and 10-1$\times$5, varying task size and incremental difficulty [3]. Centralized baselines use identical splits for fair comparison. We adopt a 5-client federated setup with non-overlapping task streams per client. Even within a configuration, different class groupings yield distinct task sequences (e.g., 3 classes per task result in 455 sequences). To model non-IID settings, client data heterogeneity is simulated using a Dirichlet distribution with $\alpha \in \{0.1, 1, 10, 100\}$, where smaller α indicates stronger imbalance [12], [27]. We use a ResNet-34-based convolutional autoencoder, where the encoder removes the classification head and a symmetric CNN decoder reconstructs the input. Encrypted gradient alignment is implemented via the CKKS scheme using TenSEAL. Performance is evaluated using image-level AUROC (Avg. Acc) and average forgetting (FM) [3], with results averaged across clients. All FCL baselines use the same backbone to ensure fair comparison. For both heterogeneous and non-IID FCL settings, we use a batch size of 32, 5 local epochs, and full client participation (fraction = 1.0). The learning rates are 0.001 (heterogeneous) and 0.005 (non-IID), with 20 and 30 communication rounds, respectively; the non-IID heterogeneity level is set to 10, and $\kappa, \lambda_u, \lambda_v = 0.5, 0.7, 0.5$.

B. Result Analysis

a) Comparison with FCL Baselines: Table I reports results across four task configurations: simpler single-step tasks (14-1, 10-5) and challenging multi-step tasks (3$\times$5, 10-1$\times$5). Classical federated baselines (FedAvg [4], [27],

TABLE I

PERFORMANCE COMPARISON WITH FL AND FCL BASELINES. * COMBINES UCAD [2] AT EACH CLIENT. FCL-IAD USES CLIENT-SIDE ORTHOGONAL PROJECTION WITH UV REGULARIZATION AND SERVER-SIDE GM. BEST RESULTS ARE SHOWN IN BOLD.

Model	14 - 1 with 1 Step		10 - 5 with 1 Step		3 × 5 with 5 Steps		10 - 1 × 5 with 5 Steps	
	Avg Acc (↑)	FM (↓)	Avg Acc (↑)	FM (↓)	Avg Acc (↑)	FM (↓)	Avg Acc (↑)	FM (↓)
FedAvg* [4]	83.38 ± 0.74	15.91 ± 0.49	80.12 ± 0.68	10.66 ± 0.37	73.63 ± 0.78	13.92 ± 0.41	75.27 ± 0.74	17.73 ± 0.58
FedUV* [27]	85.41 ± 0.69	16.12 ± 0.47	82.91 ± 0.64	9.81 ± 0.36	75.37 ± 0.81	13.24 ± 0.44	80.83 ± 0.67	15.16 ± 0.47
FedOMG* [13]	86.15 ± 0.66	14.82 ± 0.45	83.51 ± 0.59	9.93 ± 0.33	77.72 ± 0.68	14.55 ± 0.60	82.17 ± 0.77	14.42 ± 0.46
GLFC [9]	87.17 ± 0.58	13.76 ± 0.42	85.52 ± 0.61	9.66 ± 0.31	78.27 ± 0.85	11.35 ± 0.33	85.18 ± 0.62	14.78 ± 0.41
FedCIL [10]	88.58 ± 0.52	13.63 ± 0.39	87.23 ± 0.55	8.96 ± 0.28	80.18 ± 1.19	10.81 ± 0.36	86.63 ± 0.58	13.76 ± 0.69
FedKnow [11]	91.64 ± 0.41	10.82 ± 0.34	88.40 ± 0.47	8.21 ± 0.29	82.15 ± 0.51	8.74 ± 0.30	88.25 ± 0.54	13.21 ± 0.38
TARGET [23]	89.26 ± 0.56	12.94 ± 0.41	86.54 ± 0.59	9.13 ± 0.32	79.83 ± 1.03	10.26 ± 0.35	86.14 ± 0.61	13.52 ± 0.42
LANDER [24]	88.74 ± 0.49	12.07 ± 0.38	86.31 ± 0.52	7.99 ± 0.27	80.32 ± 0.56	9.76 ± 0.32	87.46 ± 0.57	12.14 ± 0.52
FedSSI [8]	92.48 ± 0.36	9.84 ± 0.29	89.14 ± 0.41	7.33 ± 0.24	83.24 ± 0.91	8.91 ± 0.38	89.82 ± 0.49	9.86 ± 0.47
STAMP [12]	91.62 ± 0.34	9.63 ± 0.26	90.27 ± 0.39	6.91 ± 0.23	**85.43** ± 0.89	8.22 ± 0.41	89.14 ± 0.57	10.31 ± 0.49
Ours (FCL-IAD)	**92.74** ± 0.69	**8.96** ± 0.72	**90.66** ± 0.91	**5.23** ± 0.54	84.51 ± 0.75	**7.39** ± 0.65	**91.87** ± 0.48	**9.17** ± 0.86
Ours (FedAGM)	87.77 ± 0.47	11.06 ± 0.39	86.49 ± 0.51	8.84 ± 0.28	80.31 ± 0.74	10.03 ± 0.37	86.16 ± 0.62	14.73 ± 0.54

FedOMG [13]) suffer from naive averaging over heterogeneous updates, while continual FL baselines (GLFC [9], FedCIL [10], FedKnow [11], TARGET [23], LANDER [24]) reduce but do not eliminate forgetting. FedAGM maintains stable performance under FHE, confirming that encrypted gradient alignment supports robust privacy-preserving FCL. Without FHE, our FCL-IAD framework consistently achieves the best performance across all configurations.

Table II shows that overlapping tasks across clients consistently improve performance, as shared tasks reduce cross-client gradient interference and allow updates to reinforce each other. Classical federated methods still lag behind centralized IAD, while advanced FCL baselines such as FedSSI [8] and STAMP [12] surpass it by effectively leveraging cross-client knowledge. IPoLoRA-C [26] underperforms centralized IAD but offers strict privacy guarantees. FedAGM maintains strong performance under FHE in both settings, demonstrating robustness to varying task distributions while preserving privacy.

b) Ablation Study: Table III evaluates each component of FedAGM. Removing AGM or UV degrades performance due to non-IID client distributions. FHE introduces a slight drop but remains robust, preserving privacy while maintaining competitive results. AGM provides the largest gain by reducing cross-client conflicts in heterogeneous tasks, while UV regularization further stabilizes representations. For reference, the upper bound, FCL-enabled IAD without FHE (Table I) achieves the best performance, indicating that FedAGM under FHE closely approaches ideal aggregation.

c) Impact of Data Heterogeneity: Figure 2 shows Avg. test accuracy under varying levels of data heterogeneity. Avg. Accuracy decreases for all methods as heterogeneity increases (smaller α). Despite the performance drop from encrypted gradient operations, FedAGM remains robust across all heterogeneity levels, demonstrating its effectiveness in mitigating client drift while preserving privacy.

d) Impact of Client Fraction: Table IV shows that accuracy improves and forgetting decreases as client fraction increases, as more representative updates reduce drift and stabilize learning. At the sparse setting (fraction $= 0.1$), FedAGM trails STAMP by $\sim 6\%$ in both settings due to limited gradient alignment under encrypted computation, but recovers substantially at higher fractions, confirming robustness as participation scales.

e) Cross-Data and Open-set Anomaly Detection: Figure 3 evaluates cross-dataset continual IAD: MVTec-VisA[2] (sequential training), Mix-order (random task mixing), and Isolated (single-dataset baselines). While FedAGM exhibits an accuracy gap compared to SOTA due to encrypted aggregation constraints, all methods see performance declines in the open-set setting. This underscores the difficulty of generalizing to novel categories under distribution shifts in FCL-IAD.

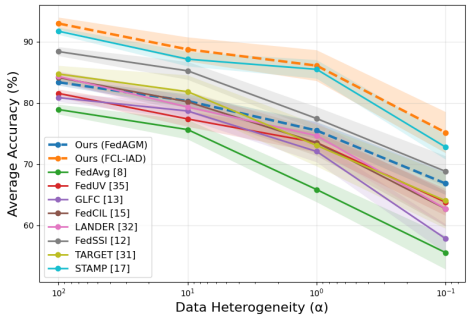

Fig. 2. Performance w.r.t. α for non-overlapping 3×5 task configuration.

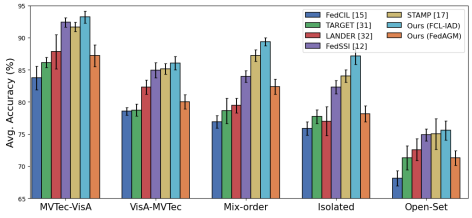

Fig. 3. Cross-dataset and Open-set continual anomaly detection results.

VI. CONCLUSION

We introduced FedAGM, an FHE-native FCL framework for unsupervised industrial anomaly detection. By combining client-side orthogonal projection with server-side FHE-

[2]https://www.kaggle.com/datasets/ess1004/visa-anomaly-detection

TABLE II
PERFORMANCE COMPARISON OF BASELINES WITH 3 CLASSES PER TASK OVER 3 SEQUENTIAL STEPS. * IS COMBINED WITH UCAD [2].

Model	Non-Overlapped		Overlapped	
	Avg Acc (↑)	FM (↓)	Avg Acc (↑)	FM (↓)
DiAD [1]	-	-	80.5 ± 0.64	11.8 ± 0.55
UCAD [2]	-	-	84.8 ± 0.46	10.3 ± 0.39
ControlNet [22]	-	-	79.2 ± 0.78	13.9 ± 0.52
CDAD [3]	-	-	89.1 ± 0.40	3.69 ± 0.68
IPoLoRA-C [26]	-	-	78.56 ± 0.88	10.74 ± 0.68
FedAvg* [4]	74.52 ± 0.78	13.61 ± 0.41	76.12 ± 0.72	12.63 ± 0.38
FedUV* [27]	75.08 ± 0.81	13.14 ± 0.44	77.83 ± 0.76	12.27 ± 0.40
FedOMG* [13]	76.62 ± 0.68	14.78 ± 0.60	80.48 ± 0.63	13.41 ± 0.52
GLFC [9]	77.83 ± 0.85	11.48 ± 0.33	83.92 ± 0.79	9.36 ± 0.29
FedCIL [10]	80.96 ± 1.19	9.89 ± 0.36	85.75 ± 1.08	8.42 ± 0.31
FedKnow [11]	82.81 ± 0.51	8.97 ± 0.30	89.42 ± 0.47	7.13 ± 0.27
TARGET [23]	80.65 ± 1.03	10.04 ± 0.35	85.01 ± 0.94	7.42 ± 0.31
LANDER [24]	81.16 ± 0.56	9.31 ± 0.32	88.58 ± 0.52	7.16 ± 0.28
FedSSI [8]	84.63 ± 0.91	**8.20** ± 0.38	90.37 ± 0.84	6.52 ± 0.34
STAMP [12]	**85.94** ± 0.89	8.36 ± 0.41	**92.11** ± 0.83	**6.41** ± 0.36
Ours (FedAGM)	79.16 ± 0.74	11.10 ± 0.37	87.07 ± 0.69	8.27 ± 0.33

TABLE III
ABLATION STUDY (3 × 5)

FHE	AGM	UV	Acc ↑	FM ↓
×	×	×	75.17±0.71	11.15±0.60
×	×	✓	78.73±0.32	10.67±0.48
✓	×	×	71.86±0.55	12.08±0.41
✓	✓	×	77.19±0.63	10.51±0.76
✓	×	✓	75.44±0.27	10.96±0.46
✓	✓	✓	**80.31±0.74**	**10.03±0.37**

TABLE IV
IMPACT OF CLIENT FRACTION IN FEDAGM (50 CLIENTS, 20 ROUNDS).

Client Fraction	Non-Overlapped		Overlapped	
	Avg. Acc (%) ↑	FM (↓)	Avg. Acc (%) ↑	FM (↓)
STAMP† (0.10)	73.67 ± 0.51	11.73 ± 0.64	84.19 ± 0.38	6.56 ± 0.52
FedAGM (0.10)	67.57 ± 1.33	13.47 ± 0.82	79.17 ± 0.74	8.33 ± 0.27
FedAGM (0.25)	70.80 ± 0.50	12.22 ± 0.60	81.25 ± 0.36	6.94 ± 0.48
FedAGM (0.40)	72.51 ± 0.49	10.65 ± 0.57	83.83 ± 0.35	6.76 ± 0.45
FedAGM (0.75)	77.25 ± 0.48	10.01 ± 0.54	87.21 ± 0.33	6.51 ± 0.40
FedAGM (1.00)	80.41 ± 0.47	9.24 ± 0.50	91.73 ± 0.32	6.44 ± 0.38

compatible approximate gradient matching, the method mitigates catastrophic forgetting, stabilizes representations under non-IID data, and enables secure coordination across clients. Experiments on MVTec-AD demonstrate performance comparable to SOTA FCL methods while ensuring strict privacy.

REFERENCES

[1] H. He, J. Zhang, H. Chen, X. Chen, Z. Li, X. Chen, Y. Wang, C. Wang, and L. Xie, "A diffusion-based framework for multi-class anomaly detection," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 38, no. 8, 2024, pp. 8472–8480.

[2] J. Liu, K. Wu, Q. Nie, Y. Chen, B.-B. Gao, Y. Liu, J. Wang, C. Wang, and F. Zheng, "Unsupervised continual anomaly detection with contrastively-learned prompt," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 38, no. 4, 2024, pp. 3639–3647.

[3] X. Li, X. Tan, Z. Chen, Z. Zhang, R. Zhang, R. Guo, G. Jiang, Y. Chen, Y. Qu, L. Ma *et al.*, "One-for-more: Continual diffusion model for anomaly detection," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 4766–4775.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[5] J. Yoon, W. Jeong, G. Lee, E. Yang, and S. J. Hwang, "Federated continual learning with weighted inter-client transfer," in *International conference on machine learning*. PMLR, 2021, pp. 12 073–12 086.

[6] A. Chaudhry, N. Khan, P. Dokania, and P. Torr, "Continual learning in low-rank orthogonal subspaces," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9900–9911, 2020.

[7] D. Cheng, Y. Hu, N. Wang, D. Zhang, and X. Gao, "Achieving plasticity-stability trade-off in continual learning through adaptive orthogonal projection," *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[8] Y. Li, Y. Wang, H. Wang, Y. Qi, T. Xiao, and R. Li, "Rehearsal-free continual federated learning with synergistic synaptic intelligence," *arXiv preprint arXiv:2412.13779*, 2024.

[9] J. Dong, L. Wang, Z. Fang, G. Sun, S. Xu, X. Wang, and Q. Zhu, "Federated class-incremental learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 164–10 173.

[10] D. Qi, H. Zhao, and S. Li, "Better generative replay for continual federated learning," *arXiv preprint arXiv:2302.13001*, 2023.

[11] Y. Luopan, R. Han, Q. Zhang, C. H. Liu, G. Wang, and L. Y. Chen, "Fedknow: Federated continual learning with signature task knowledge integration at edge," in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 2023, pp. 341–354.

[12] D. M. Nguyen, L.-T. Nguyen, and Q.-V. Pham, "Spatio-temporal gradient matching for federated continual learning," in *ICML 2025 Workshop on Collaborative and Federated Agentic Workflows*, 2025.

[13] T.-B. Nguyen, M.-D. Nguyen, J. Park, Q.-V. Pham, and W. J. Hwang, "Federated domain generalization with data-free on-server matching gradient," *arXiv preprint arXiv:2501.14653*, 2025.

[14] K. N. Kumar, C. K. Mohan, and L. R. Cenkeramaddi, "The impact of adversarial attacks on federated learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 5, pp. 2672–2691, 2023.

[15] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," *Advances in neural information processing systems*, vol. 32, 2019.

[16] L. Fan and L. Xiong, "Differentially private anomaly detection with a case study on epidemic outbreak detection," in *2013 IEEE 13th international conference on data mining workshops*. IEEE, 2013, pp. 833–840.

[17] W. Jin, Y. Yao, S. Han, C. Joe-Wong, S. Ravi, S. Avestimehr, and C. He, "Fedml-he: An efficient homomorphic-encryption-based privacy-preserving federated learning system," in *International Workshop on Federated Learning in the Age of Foundation Models in Conjunction with NeurIPS 2023*.

[18] C. Gentry, *A fully homomorphic encryption scheme*. Stanford university, 2009.

[19] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *International conference on the theory and application of cryptology and information security*. Springer, 2017, pp. 409–437.

[20] E. Gronberg, L. d'Aliberti, M. Saebo, and A. Hook, "Blindfl: Segmented federated learning with fully homomorphic encryption," *arXiv preprint arXiv:2501.11659*, 2025.

[21] A. K. Saibaba, "Approximating monomials using chebyshev polynomials," *arXiv preprint arXiv:2101.06818*, 2021.

[22] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 3836–3847.

[23] J. Zhang, C. Chen, W. Zhuang, and L. Lyu, "Target: Federated class-continual learning via exemplar-free distillation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4782–4793.

[24] M.-T. Tran, T. Le, X.-M. Le, M. Harandi, and D. Phung, "Text-enhanced data-free approach for federated class-incremental learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 23 870–23 880.

[25] Z. Zuo, N. Su, B. Li, and T. Zhang, "Pack: Towards communication-efficient homomorphic encryption in federated learning," in *Proceedings of the 2024 ACM Symposium on Cloud Computing*, 2024, pp. 470–486.

[26] H. B. Ta, D. Nguyen, Q. Tran, T. Tran, and T. Pham, "Low-rank adaptation in multilinear operator networks for security-preserving incremental learning," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 24 341–24 350.

[27] H. M. Son, M.-H. Kim, T.-M. Chung, C. Huang, and X. Liu, "Feduv: uniformity and variance for heterogeneous federated learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 5863–5872.