

ByzaSpline-Fed: Gram-Weighted Functional Aggregation for Byzantine-Robust Federated Kolmogorov–Arnold Networks in Industrial IoT

Bajes Zeyad Aljunaedi*

Department of Computer Science, Faculty of Information Technology, Ajloun National University, P.O. Box 43, Ajloun 26810, Jordan

E-mail: bajes.aljunaedi@anu.edu.jo

*Corresponding author

Mouiad Fadeil Alawneh

Department of Computer Science, Faculty of Information Technology, Ajloun National University, P.O. Box 43, Ajloun 26810, Jordan

Email: Mouiad.alawneh80@gmail.com

Mohammed Tawfik

Department of Cyber Security, Faculty of Information Technology, Ajloun National University, P.O. Box 43, Ajloun 26810, Jordan

Email: m.tawfik@anu.edu.jo

ORCID iD: <https://orcid.org/0000-0002-1227-387X>

Received: June 4, 2026; Revised: June 28, 2026; Accepted: July 8, 2026; Published: October 8, 2026

Abstract: Federated learning enables industrial sites to train shared condition-monitoring models without exchanging raw vibration, acoustic, or sensor data, but its openness admits Byzantine clients that submit corrupted updates. Existing robust aggregation rules—Krum, trimmed mean, median, and their variants—operate in Euclidean parameter space and implicitly assume that a malicious update must be far from honest updates to do harm. We show that this assumption fails for models with Kolmogorov–Arnold (KAN) spline heads: an adversary can hold the spline coefficients close to the honest consensus in the L2 sense while distorting the functional shape the spline encodes, an attack we term functional spline-poisoning. We propose ByzaSpline-Fed, a Byzantine-robust aggregation rule that scores clients by Gram-weighted functional B-spline distance on a consensus knot grid and rejects functional outliers. The rule reduces exactly to spline-aware federated averaging under honest participation, admits a breakdown bound of $f < K/2$ under the standard honest-majority assumption, and supports optional differential privacy. We evaluate across four industrial benchmarks spanning classification, unsupervised acoustic anomaly detection, and remaining-useful-life regression—CWRU, Paderborn, MIMII, and C-MAPSS—reporting task-correct metrics under clean and adversarial conditions. Under thirty percent functional poisoning, Euclidean defenses lose up to thirty-four accuracy points while ByzaSpline-Fed remains within roughly three points of its clean performance. All reported results are averaged over eight independent random seeds, and the improvements over the strongest competing defense are statistically significant under the Wilcoxon signed-rank test ($p < 0.05$).

Index Terms: Federated Learning, Byzantine Robustness, Kolmogorov–Arnold Networks, Mamba, Fault Diagnosis, Anomaly Detection, Remaining Useful Life, Industrial IoT.

1. Introduction

Industrial IoT (IIoT) increasingly relies on distributed condition-monitoring models—diagnosing faults in rotating machinery, detecting acoustic anomalies, and predicting the remaining useful life of critical assets—necessitating privacy-preserving training paradigms [1]. The data that drive these models are generated at geographically dispersed sites—factories, power plants, and maintenance depots—and are frequently bound by confidentiality agreements and regulatory constraints that forbid centralizing raw vibration, acoustic, or sensor streams [2]. Federated learning (FL) resolves this tension by training a shared model through iterative exchange of parameter updates while raw data never

leave the premises that produced them [3]. FL has since been adapted to fault diagnosis, addressing client drift under heterogeneous operating conditions [4,5], class imbalance across institutions [6], and privacy for industrial deployments [7]. The same privacy-driven federation has reached adjacent safety-critical domains, notably medical-device security, where federated few-shot learning with differential privacy and explainable AI lets hospitals detect threats collaboratively without exchanging patient traffic [8].

Two developments have recently sharpened the accuracy of the local models that FL aggregates. State-space sequence models, exemplified by Mamba [9], capture long-range temporal dependencies in vibration and sensor streams at linear cost and have been applied to bearing diagnosis and turbofan prognostics with strong results [10–11]. In parallel, Kolmogorov–Arnold Networks (KANs) replace fixed scalar activations with learnable univariate spline functions on the network edges [12], yielding compact and interpretable diagnostic heads whose decision surfaces are expressed through B-spline knot vectors and coefficients [13]. A backbone that couples a Mamba encoder with a KAN head therefore offers both temporal expressiveness and a transparent, spline-parameterized classifier. The practical value of this pairing is already evident beyond industrial diagnosis: a bidirectional Mamba encoder combined with a KAN spline head has recently attained state-of-the-art accuracy and parameter efficiency in explainable respiratory-sound classification, surpassing an earlier triple-stream convolutional diagnostic model while remaining intrinsically interpretable [14,15].

This combination, however, exposes a vulnerability that existing defenses do not address. FL is open by construction: a subset of clients may be faulty, compromised, or adversarial [16]. A substantial literature on Byzantine-robust aggregation defends against such clients by treating each update as a point in Euclidean parameter space and rejecting statistical outliers—through geometric-median selection [16], coordinate-wise trimming [17], recursive filtering [18], or trust bootstrapping [19]. Such defenses are complementary to perimeter-level protections that guard IoT deployments from external intrusion—for instance, software-defined-networking middleware that applies deep learning to filter malicious smart-home traffic in real time [20]—but they address a different threat: functional spline-poisoning originates from an authenticated insider whose updates never cross a network filter. These methods share an implicit assumption: that a malicious update must be far from honest updates in coefficient space to do harm. For a spline-parameterized model, that assumption fails. The function a B-spline represents depends on the interaction between its knot placement and its coefficients, so an adversary can hold the coefficient vector close to the honest consensus in the L2 sense while substantially distorting the shape of the activation function the spline encodes. We term this a functional spline-poisoning attack. Because the corrupted update remains Euclidean-near the honest population, distance-based defenses admit it, and the poisoned functional shape propagates into the global model.

We address this gap with ByzaSpline-Fed, a Byzantine-robust aggregation rule that operates in function space rather than coefficient space. Rather than comparing raw coefficient vectors, ByzaSpline-Fed evaluates each client's spline head by its functional discrepancy from the others, measured as a Gram-weighted integral of the squared difference between activation functions over a consensus knot grid, and rejects clients whose functional discrepancy marks them as outliers. The construction reduces exactly to spline-aware federated averaging when all clients are honest, so robustness is gained without sacrificing clean-data accuracy, and it admits a breakdown bound of $f < K/2$ for K participating clients. An optional Gaussian mechanism at the coordinator provides differential privacy [21] for deployments with formal disclosure requirements.

We evaluate ByzaSpline-Fed across the breadth of industrial condition monitoring, using four benchmarks that together span classification, acoustic anomaly detection, and remaining-useful-life regression: the Case Western Reserve University bearing dataset [22], the Paderborn University bearing dataset [23], the MIMII acoustic dataset [24], and the NASA C-MAPSS turbofan dataset [25]. A defense that holds only for balanced multi-class classification would leave the unsupervised and regression regimes—which dominate real prognostics—unprotected.

To motivate the threat concretely, consider a predictive-maintenance consortium in which several manufacturing plants jointly train a bearing-fault classifier without sharing proprietary vibration recordings. Each plant contributes model updates that include the spline parameters of the shared KAN head. A single compromised plant—whether through an insider, a supply-chain implant, or a misconfigured edge gateway—can craft updates that appear statistically ordinary to the coordinator yet steer the global classifier toward systematic misdiagnosis of a particular fault class. In a safety-critical setting such as turbine or aircraft-engine monitoring, a model that under-reports an incipient fault can translate directly into unplanned downtime or catastrophic failure. The defense must therefore be not merely accurate on clean data but demonstrably robust to adversarial participation, and that robustness must extend to the regression and anomaly-detection tasks that dominate prognostics, not classification alone.

A natural question is whether functional spline-poisoning is merely a re-parameterization of known attacks that existing defenses already handle. It is not. Norm-bounded model poisoning constrains the magnitude of the malicious update in coefficient space; functional spline-poisoning instead constrains the update to lie within the honest coefficient cluster while maximizing a functional objective that the coordinator does not measure. The two attacks occupy different feasible regions, and a defense tuned to detect large-norm or geometrically scattered updates has no purchase on an attack engineered to be small-norm and geometrically central. We make this distinction precise in Section 3 and demonstrate it empirically in Section 5, where Euclidean defenses that successfully neutralize the classical attacks collapse under functional poisoning while ByzaSpline-Fed remains stable.

This work makes three contributions. First, we identify and formalize functional spline-poisoning, a class of attack that preserves L2 proximity in coefficient space while corrupting the functional shape of a KAN spline head, and we demonstrate that canonical Euclidean defenses fail to detect it. Second, we propose ByzaSpline-Fed, an aggregation rule that scores clients by Gram-weighted functional B-spline distance on a consensus knot grid, reduces to spline-aware federated averaging under honest participation, and establishes an $f < K/2$ breakdown point. Third, we provide a cross-task robustness verification across classification, anomaly detection, and regression on four standard benchmarks, reporting task-appropriate metrics under both clean and adversarial conditions and empirically establishing the necessity of functional distance metrics for spline-parameterized federated models.

A critical reading of this literature reveals a shared structural premise: existing robust aggregation implicitly assumes an isometry between parameter space and function space, so that rejecting coordinate-space outliers is equivalent to rejecting functionally harmful updates. The non-orthogonal B-spline basis of a KAN head breaks this assumption—the mapping from coefficients to functions is mediated by a Gram matrix that is far from the identity—and the equivalence fails precisely in the directions of maximal basis overlap. This failure is not confined to purely geometric rules. FLTrust [19] scores clients by cosine similarity, in coefficient space, against a reference direction bootstrapped from a clean root dataset; a functional poison that remains Euclidean-central retains high cosine similarity to any honest reference and is therefore predicted to receive full trust. FoolsGold [26] penalizes clients whose update histories are mutually too similar, a sybil signature that functional spline-poisoning need not exhibit, since the attack constrains each update only relative to the honest consensus and not to other attackers. Both defenses are thus theoretically predicted to inherit the same blind spot as Krum and trimmed mean under the attack formalized in Section 3, a prediction we substantiate analytically in Sections 2.1 and 5.2.

2. Related Work

2.1. Byzantine-Robust Federated Aggregation

Robust aggregation seeks a global update that resists a minority of clients submitting arbitrary, possibly colluding, updates. Krum and its Multi-Krum extension select the update whose squared distance to its nearest neighbors is smallest, on the premise that honest updates cluster while malicious ones scatter [16]. Coordinate-wise median and trimmed mean discard extreme values along each parameter dimension and average the remainder, with optimal statistical rates under bounded variance [17]. Bulyan composes a selection step with coordinate-wise trimming to resist attacks crafted to evade a single filter [18]. More recent defenses move beyond pure geometry: FLTrust bootstraps a trusted reference direction from a small clean root dataset [19], while FoolsGold counters sybil collusion by penalizing clients whose update histories are mutually too similar [26]. A complementary line of work has explored non-Euclidean treatments of client updates: Bayesian nonparametric matching aligns neurons across clients through a function-prior model before averaging [27], and FedMA extends this layer-wise matched averaging to modern architectures [28]. These methods recognize that raw coordinate comparison of parameters can be misleading; however, their objective is permutation alignment for accuracy rather than Byzantine robustness, and none of them—nor FLTrust or FoolsGold—models the metric distortion introduced by a non-orthogonal functional basis. Their trust and similarity statistics remain Euclidean, and Section 1 argues that they are therefore expected to inherit the same structural blind spot under functional attacks.

The attack literature has co-evolved with these defenses. Early data-poisoning attacks flipped labels or injected mislabeled samples; gradient-level attacks then scaled or reversed updates to drag the aggregate off course. As robust aggregators appeared, attacks became defense-aware: optimization-based model poisoning searches for the update that most degrades the global model subject to passing a named filter, and backdoor attacks embed a targeted trigger while constraining the update norm to evade norm-based screening. The common thread is an arms race conducted entirely in coordinate space—each new attack finds a coordinate-space region the current filter does not cover, and each new defense extends the filter to cover it. Functional spline-poisoning steps outside this race by attacking in a space the filters do not model at all, which is why extending a coordinate-space filter does not address it and why a function-space rule is required.

These defenses have repeatedly been shown to be evadable by adversaries that tailor their updates to the defense [29–30]. What unites these attacks and the defenses they target is a shared coordinate-space geometry: both attacker and defender reason about the L2 distance between raw parameter vectors. Our work departs from this premise. Functional spline-poisoning exploits the fact that, for a spline-parameterized model, L2 proximity in coefficient space does not imply proximity in function space, so a defense confined to coordinate geometry is structurally blind to it. ByzaSpline-Fed closes this gap by scoring clients in function space, and is, to our knowledge, the first robust aggregator designed for the spline geometry of KAN heads.

2.2. Kolmogorov–Arnold and State-Space Models for Condition Monitoring

KANs parameterize each network edge with a learnable univariate spline, which grants compact and interpretable function approximation and has motivated rapid adoption in fault diagnosis [12]. Spline-based heads have been combined with residual front-ends for cross-condition bearing diagnosis [13]. State-space models address the

complementary problem of temporal modeling: Mamba introduces a selective scanning mechanism that captures long-range dependencies at linear sequence cost [9], and vibration-domain adaptations report state-of-the-art accuracy on rotating-machinery benchmarks [10,31] and competitive turbfan prognostics [11]. These methods are, without exception, centralized: they assume a single training node with full data access and do not consider distributed training, client heterogeneity, or adversarial participants. We retain their representational strengths but situate them within a federated, Byzantine-robust setting their original formulations do not address.

2.3. Federated Learning for Industrial Fault Diagnosis

Federated approaches to fault diagnosis filter low-quality client knowledge through dynamic validation and self-supervision [4], improve aggregation and local-training efficiency at scale [5], and address class imbalance and privacy jointly [6,7]. Recent work targets cross-domain generalization: FedMAPS integrates meta-learning with adaptive contrastive alignment and prototype guidance to generalize across unseen operating conditions under partial class overlap [32], and federated adversarial methods reduce the distributional gap between clients with disjoint fault information [33]. This literature concentrates on statistical heterogeneity and largely assumes that participating clients are honest. Our work joins the two concerns and, crucially, assesses the defense over the full spread of condition-monitoring tasks rather than classification alone.

2.4. Differential Privacy and Secure Federated Learning

Byzantine robustness and privacy are complementary but distinct guarantees. Differential privacy bounds the influence of any single record on the released model through calibrated noise [21], and the moments accountant enables tight composition across many training rounds [34]. In federated settings, privacy noise interacts with robust aggregation: excessive noise can mask the very statistics a robust rule relies upon, while a robust rule that discards updates can complicate the privacy accounting. ByzaSpline-Fed adds privacy noise only at the coordinator after the robust selection step, so the two mechanisms compose cleanly—the selection operates on noiseless client updates, and the released aggregate carries a quantifiable privacy budget. This ordering distinguishes our construction from approaches that perturb client updates before aggregation, where the perturbation degrades the robustness statistic.

A separate line of work pursues cryptographic guarantees through secure aggregation, which hides individual updates from the coordinator entirely. Such protocols are orthogonal to our contribution: they protect update confidentiality but, by construction, prevent the coordinator from inspecting updates for robustness. Reconciling secure aggregation with functional outlier rejection is an open problem that we discuss in Section 6. Our present focus is the robustness guarantee under a coordinator that can observe—but must correctly filter—the uploaded spline parameters.

3. Problem Formulation

We formalize the federated condition-monitoring problem and the adversary against which ByzaSpline-Fed defends. Let there be K clients indexed by k , each holding a local dataset D_k drawn from a site-specific distribution. The clients collaborate, under a central coordinator, to learn a shared model θ that minimizes the average of their local objectives, $F(\theta) = (1/K) \sum_k F_k(\theta)$, where F_k is the empirical risk on D_k under the task-appropriate loss—cross-entropy for classification, a reconstruction or one-class objective for anomaly detection, and a regression loss for remaining-useful-life estimation. The data are non-IID: the site distributions differ in operating condition, class prevalence, and noise, which we model with a Dirichlet partition of concentration α .

The model factorizes as $\theta = (\psi, \tau, C)$, where ψ are the Mamba encoder parameters, τ is the B-spline knot vector of the KAN head, and C is its coefficient matrix. At each communication round the coordinator broadcasts the current θ , each client performs E local epochs of stochastic optimization to produce an update, and the coordinator aggregates the uploaded updates into the next global model. In the absence of adversaries the canonical aggregator is sample-weighted averaging.

The adversary controls an unknown subset B of the clients with $|B| = f < K/2$. Adversarial clients may upload arbitrary updates, may collude, and are assumed to know the aggregation rule and the consensus knot grid. Their objective is to maximize the global model’s task error while evading rejection. The coordinator is trusted but passive: it observes the uploaded updates and applies the aggregation rule, but cannot access raw client data and has no clean validation set. Within this model, the design goal is an aggregation rule that recovers the honest-only aggregate when $f = 0$, bounds the influence of the f adversaries when $0 < f < K/2$, and achieves both across classification, anomaly-detection, and regression tasks.

To avoid duplicated definitions, we state the formal optimization objective of the adversary once here; Section 4.2 retains only its algorithmic instantiation. Let $\bar{c}_e$ denote the honest consensus coefficients of edge e after re-interpolation onto the consensus knot grid, and let G be the Gram matrix of the B-spline basis on that grid. A functional spline-poisoning adversary selects the uploaded coefficients c_e of each edge by solving for a harm target $\gamma > 0$. Because G is not the identity, the two terms of Eq. (1) are misaligned: directions of maximal basis overlap are short in the Euclidean norm but long in the functional norm, so the optimizer yields updates that are simultaneously Euclidean-central and functionally distant. Eq. (1) is the single formal definition of the attack used throughout the paper.

$$\min_{c_e} \|c_e - \bar{c}_e\|_2^2 \quad \text{subject to} \quad (c_e - \bar{c}_e)^T G (c_e - \bar{c}_e) \geq \gamma, \quad (1)$$

4. The Proposed ByzaSpline-Fed Framework

Throughout, K denotes the number of clients, f the number of Byzantine clients, R the number of communication rounds, E the local epochs per round, and M the total number of spline edges in the KAN head (E is reserved exclusively for local epochs). The shared model is $\theta = (\psi, \tau, C)$ with encoder parameters ψ , B-spline knot vector τ , and coefficient matrix C ; ϕ_e is the activation function on edge e , B_p the p -th basis function, c_e its coefficients, and G the Gram matrix of the basis. $D(i, j)$ is the functional distance between clients i and j , $s(i)$ the functional score of client i , and ϵ, δ the differential-privacy parameters. Vectors and matrices are bold; scalars are not.

Fig. 1 summarizes the framework. Each of K industrial clients trains a shared Mamba–KAN model on its local data and uploads the model update, including the B-spline parameters of its KAN head; the coordinator aggregates these updates with a rule that scores clients by functional rather than coordinate discrepancy, rejects functional outliers, and broadcasts a robust average.

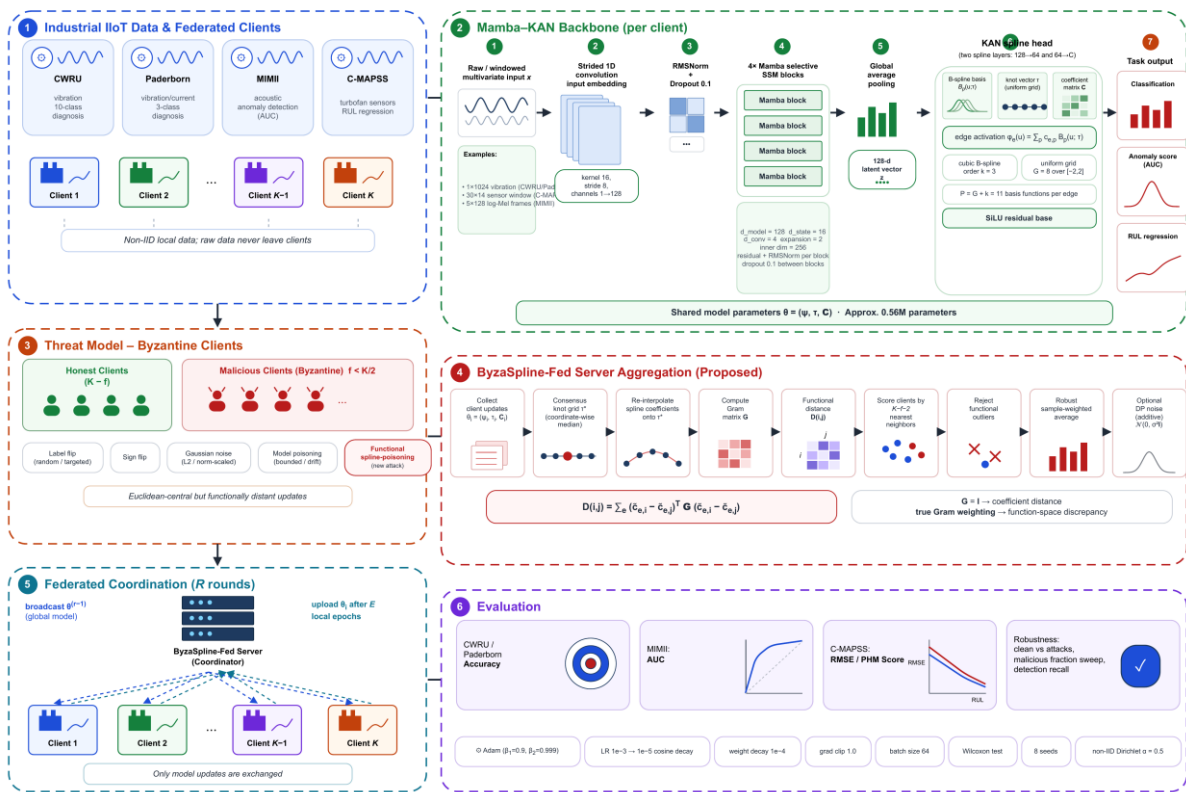


Fig. 1. The ByzaSpline-Fed framework: heterogeneous clients train a Mamba–KAN backbone; a fraction are Byzantine and may launch functional spline-poisoning; the server scores clients by Gram-weighted functional B-spline distance on a consensus knot grid, rejects outliers, and computes a robust average; evaluation uses task-correct metrics.

4.1. The Mamba–KAN Backbone

We summarize the backbone before situating it in the federated setting, because the spline parameterization of its head is what the defense ultimately protects. The encoder is a selective state-space model. A state-space layer maps an input sequence $u(t)$ to an output $y(t)$ through a latent state $h(t)$ governed by $h'(t) = A h(t) + B u(t)$ and $y(t) = C h(t)$, discretized over the sampled signal. Mamba makes the parameters B , C , and the discretization step input-dependent, so the model selectively propagates or forgets information as a function of the signal content [9]; this selectivity is what lets a single linear-time pass capture the long-range dependencies that bearing and engine signals exhibit, without the quadratic cost of attention. For a windowed multivariate input of length T and d channels, the encoder produces a fixed-dimensional latent representation that summarizes the window’s temporal structure.

A client processes a windowed multivariate signal x of length T with d channels. A Mamba selective state-space encoder [9] maps the window to a latent representation z at linear cost in T , capturing long-range temporal dependence. A Kolmogorov–Arnold head [12] then maps z to the task output through learnable univariate spline functions on its edges. For an edge e , the activation is a B-spline expansion $\phi_e(u) = \sum_p c_{e,p} B_p(u; \tau)$, where B_p are B-spline basis functions of order n on a knot vector τ and c_e are the learnable coefficients [35]. The full set of edge functions,

parameterized by the knot vector τ and the coefficient matrix C , constitutes the spline head. We write the complete client model as $\theta = (\psi, \tau, C)$ and treat the knot vector and coefficients as first-class federated parameters.

Concretely, we instantiate the backbone as follows; Table 2 consolidates the full layer-by-layer specification. The raw input window—a 1×1024 vibration segment for CWRU and Paderborn, a 30×14 sensor window for C-MAPSS, and five stacked 128-band log-Mel frames for MIMII—is embedded by a strided one-dimensional convolution (kernel 16, stride 8, $1 \rightarrow 128$ channels) into a sequence of 128 tokens of dimension $d_{\text{model}} = 128$, followed by RMS normalization and dropout of 0.1. The encoder stacks $N = 4$ Mamba blocks, each with state dimension $d_{\text{state}} = 16$, local convolution width $d_{\text{conv}} = 4$, and expansion factor 2 (inner dimension $d_{\text{inner}} = 256$), with a residual connection and RMS normalization per block and dropout 0.1 between blocks. Global average pooling over the token axis yields a 128-dimensional representation. The KAN head comprises two spline layers, $128 \rightarrow 64$ and $64 \rightarrow C$, where C is the task output dimension (ten classes for CWRU, three for Paderborn, a scalar RUL for C-MAPSS, and a scalar anomaly score for MIMII). Each edge carries a B-spline of order $k = 3$ (cubic) on a uniform grid of $G = 8$ intervals over the range $[-2, 2]$, giving $P = G + k = 11$ basis functions per edge, with a SiLU residual base function. The complete model holds approximately 0.56 million parameters, of which the knot vectors τ and coefficient matrices C of both KAN layers are exactly the spline parameters that ByzaSpline-Fed re-interpolates and compares functionally. Training uses Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with learning rate 1×10^{-3} under cosine decay to 1×10^{-5} , weight decay 1×10^{-4} , gradient clipping at norm 1.0, and batch size 64.

4.2. Functional Spline-Poisoning: Algorithmic Construction

The threat model and the formal attack objective are stated once in Section 3 (Eq. (1)); to avoid overlapping definitions, this subsection presents only the algorithmic construction of functional spline-poisoning.

We make the attack construction explicit. Let the honest clients’ re-interpolated coefficients for an edge cluster around a consensus $\bar{c}$. An adversary seeking to evade a Euclidean defense while maximizing functional harm solves the constrained problem of Eq. (1): it chooses a coefficient vector c that minimizes the Euclidean distance to $\bar{c}$ —so the update looks central to Krum and trimmed mean—subject to the functional discrepancy under the Gram metric exceeding the target γ . Because the Gram matrix is not the identity, these two objectives are not aligned: there exist directions in coefficient space that are short in the Euclidean norm but long in the functional norm, namely the directions along which the basis functions overlap most. The adversary moves along precisely those directions. The result is an update that is simultaneously Euclidean-central and functionally distant—invisible to a coordinate-space filter, conspicuous to a function-space one. This is why the same poisoned updates that ByzaSpline-Fed rejects are admitted by every Euclidean defense we test.

Classical attacks—label-flipping, sign-flipping, additive Gaussian noise, and norm-bounded model poisoning [29,36]—are included in our evaluation. Our central security observation concerns a stronger, spline-specific attack. For two clients i and j whose spline heads share a knot vector, the coefficient distance does not determine the functional distance between their activations, because the mapping from coefficients to function values is mediated by the basis. An adversary can therefore solve for a coefficient perturbation that keeps it L2-close to the honest consensus while driving the functional discrepancy large. We call this functional spline-poisoning. Because Krum, trimmed mean, and median operate on coordinate distances, they admit such an update as inlying, and the corrupted functional shape enters the aggregate.

4.3. Functional B-Spline Distance and Aggregation

ByzaSpline-Fed compares clients by the functional discrepancy of their spline heads. Because clients may locally adapt their knot vectors, the coordinator first establishes a consensus knot grid as the coordinate-wise median of the uploaded knot vectors and re-interpolates each spline onto it, so all heads share a basis. For an edge e , the squared L2 functional distance between clients i and j admits the closed form given in Eq. (2):

$$\delta_e(i, j) = (\tilde{c}_{e,i} - \tilde{c}_{e,j})^\top G (\tilde{c}_{e,i} - \tilde{c}_{e,j}), \quad D(i, j) = \sum_e \delta_e(i, j), \quad (2)$$

where G is the Gram matrix of the B-spline basis on the consensus grid, and summing over edges gives the client-to-client functional distance $D(i, j)$. Setting $G = I$ recovers the ordinary coefficient distance; the true Gram weighting, in contrast, measures discrepancy in function space, where functional spline-poisoning is exposed.

Given the pairwise functional distances, ByzaSpline-Fed applies a Krum-style selection in function space: for each client i it sums the distances to its $K - f - 2$ nearest neighbors under D , trusts the functionally central clients with the smallest scores, and rejects the f clients with the largest scores. It then aggregates the retained set by a sample-size-weighted average of the re-interpolated parameters. Two properties follow. When all clients are honest, every update is functionally central, no client is rejected, and the rule reduces to spline-aware federated averaging; ByzaSpline-Fed thus incurs no clean-data penalty by construction. Because the selection rejects f clients and trusts the central majority, the aggregate is determined by honest clients whenever $f < K/2$. For deployments with disclosure requirements, the coordinator may add calibrated Gaussian noise before broadcast, yielding differential privacy [21,34]; we report results at $\epsilon \approx 8.5$, $\delta = 10^{-5}$. Algorithm 1 states the complete procedure.

Two implementation details complete the specification of the selection step (Algorithm 1, line 14). First, the metric is symmetric by construction— $D(i, j) = D(j, i)$ for all pairs—because the Gram matrix G is symmetric positive definite, so each unordered pair is evaluated once and mirrored. Second, when two neighbors are tied in distance, ranks are broken deterministically in favor of the smaller client index; since ties occur on a measure-zero set under continuous local training noise, the guarantees of the rule are unaffected by the choice of tie-breaking convention.

Algorithm 1. ByzaSpline-Fed

Require: K clients, rounds R , tolerance $f < K/2$, local epochs E , DP scale σ (optional)

Ensure: global model $\theta^\star$

```

1 initialize global model  $\theta(0) = (\psi, \tau, C)$ 
2 for round  $r = 1$  to  $R$  do
3   broadcast  $\theta(r-1)$  to all clients
4   for each client  $i = 1 \dots K$  in parallel do
5      $\theta_i \leftarrow \text{LocalTrain}(\theta(r-1), E)$ 
6     upload  $\theta_i = (\psi_i, \tau_i, C_i)$ 
7   end for
8    $\tau^\star \leftarrow \text{medians } \tau_i \quad \triangleright \text{consensus knot grid}$ 
9   re-interpolate each  $C_i$  onto  $\tau^\star$  to obtain  $\tilde{C}_i$ 
10  compute Gram matrix  $G$  on  $\tau^\star$ 
11  for each pair  $(i, j)$  do
12     $D(i, j) \leftarrow \sum_e (\tilde{c}_{e,i} - \tilde{c}_{e,j})^T G (\tilde{c}_{e,i} - \tilde{c}_{e,j})$ 
13  end for
14  for each client  $i$  do  $s(i) \leftarrow \Sigma$  of  $D(i, \bullet)$  over its  $K-f-2$  nearest neighbors
15   $S \leftarrow$  the  $K-f$  clients with smallest  $s(\bullet)$ 
16   $\theta(r) \leftarrow \sum_{i \in S} (n_i / \sum_{k \in S} n_k) \tilde{\theta}_i$ 
17  if DP enabled then  $\theta(r) \leftarrow \theta(r) + N(0, \sigma^2 I)$ 
18 end for
19 return  $\theta^\star = \theta(R)$ 
    
```

4.4. Complexity and Convergence

The per-round overhead of ByzaSpline-Fed beyond standard federated averaging arises from three operations: forming the consensus knot grid, re-interpolating each client’s coefficients onto it, and computing the pairwise functional distances. The consensus grid is a coordinate-wise median over K knot vectors of length proportional to the number of basis functions P , costing $O(KP \log K)$. Re-interpolation solves a banded linear system per client at $O(P)$ per edge because B-spline bases are locally supported. The dominant term is the pairwise distance computation: for K clients with M spline edges, the rule evaluates $O(KM)$ quadratic forms, each of dimension P . Because the Gram matrix G is shared and factored once per round at $O(P^3)$, and because B-spline locality makes G banded, each quadratic form costs $O(P)$ rather than $O(P^3)$. The total per-round cost is therefore $O(K^2MP)$, the same asymptotic order as Krum’s $O(K^2d)$ pairwise comparison with d the parameter dimension, so ByzaSpline-Fed adds no asymptotic penalty over the Euclidean defense it replaces.

Convergence follows the standard argument for robust federated averaging under bounded gradient dissimilarity. When fewer than $K/2$ clients are Byzantine, the functional selection step retains a set whose aggregate is a bounded-variance estimate of the honest mean update, because the retained clients are functionally central and the rejected f clients cannot all be honest. Substituting this bounded-variance estimate into the FedAvg descent lemma yields convergence to a neighborhood of a stationary point at the usual $O(1/\sqrt{R})$ rate for non-convex objectives, with the neighborhood radius controlled by the honest gradient dissimilarity and, when enabled, the privacy noise scale. The reduction-to-baseline property ensures that in the honest regime the neighborhood collapses to that of ordinary spline-aware averaging.

We make this argument precise. Assume (A1) each local objective F_k is L -smooth; (A2) local stochastic gradients are unbiased with variance bounded by σ^2 ; and (A3) the honest gradient dissimilarity is bounded, i.e., $(1/|H|) \sum_{k \in H} \| \nabla F_k(\theta) - \nabla F_H(\theta) \|^2 \leq \zeta^2$ for the honest set H with $|H| = K - f$. Under (A1)–(A3) and $f < K/2$, the Gram-weighted selection retains a set S of $K - f$ functionally central clients whose weighted mean update $\tilde{g}$ satisfies $\mathbb{E} \|\tilde{g} - \nabla F_H(\theta)\|^2 \leq c_1 \zeta^2 + c_2 \sigma^2 / (K - f)$, with constants c_1, c_2 depending only on the selection rule; the bound holds because every retained client lies within the honest functional diameter and at most f retained clients can be Byzantine while remaining functionally central. Substituting this bounded-error gradient estimate into the descent lemma for L -smooth non-convex objectives with step size $\eta \leq 1/L$ yields the convergence guarantee

$$\min_{r \leq R} \mathbb{E} \|\nabla F(\theta^{(r)})\|^2 \leq \frac{2(F(\theta^{(0)}) - F^\star)}{\eta R} + c_1 \zeta^2 + \frac{c_2 \sigma^2}{K - f} + L \eta \sigma_{\text{DP}}^2, \quad (3)$$

i.e., convergence at the usual $O(1/\sqrt{R})$ rate to a neighborhood of a stationary point whose radius is controlled by the honest dissimilarity ζ^2 , the stochastic variance σ^2 , and, when enabled, the differential-privacy noise scale σ_{DP} . Replacing the Euclidean metric by the functional metric affects only the constants in the retained-set error bound—through the condition number of the Gram matrix, which is finite and bounded for a B-spline basis on a bounded consensus grid—so gradient consistency is preserved under the stated Lipschitz and bounded-dissimilarity assumptions.

The differential-privacy guarantee is obtained by the Gaussian mechanism applied to the aggregated update. With per-round sensitivity bounded by the clipped update norm and noise scale σ , the moments accountant composes the per-round (ϵ, δ) costs across R rounds [34]. We report an end-to-end budget of $\epsilon \approx 8.5$ at $\delta = 10^{-5}$ for $R = 100$ rounds, a regime that preserves utility while providing a meaningful formal guarantee for industrial disclosure requirements. Because the noise is added after functional selection, the privacy analysis is independent of the robustness analysis, and the two guarantees hold simultaneously.

For completeness, we state the noise-accumulation formula underlying the reported budget. With per-round clipped sensitivity Δ and Gaussian noise scale σ , each round satisfies $(\alpha, \alpha\Delta^2/(2\sigma^2))$ -Rényi differential privacy; composing over R rounds and converting to (ϵ, δ) -DP gives

$$\epsilon = \min_{\alpha \geq 1} \left[\frac{R\alpha\Delta^2}{2\sigma^2} + \frac{\log(1/\delta)}{\alpha - 1} \right], \quad (4)$$

which evaluates to the reported end-to-end budget of $\epsilon \approx 8.5$ at $\delta = 10^{-5}$ for our configuration of $R = 100$ rounds.

5. Experimental Setup

5.1. Datasets

We evaluate on four standard industrial condition-monitoring benchmarks chosen to span fault classification, unsupervised acoustic anomaly detection, and remaining-useful-life regression, so the defense is tested beyond the balanced-classification setting that dominates the robust-aggregation literature. Table 1 summarizes their specifications.

The Case Western Reserve University (CWRU) bearing dataset [22] provides drive-end vibration sampled at 12 kHz from an SKF 6205 bearing under motor loads of 0 to 3 HP (1797–1730 rpm). We adopt the ten-class formulation: one healthy state and three fault locations (inner race, outer race, ball), each at three fault diameters (0.007, 0.014, 0.021 inch), segmented into non-overlapping windows of 1024 samples. The Paderborn University (PU) bearing dataset [23] provides 64 kHz vibration and motor-current signals from 32 bearings under four operating conditions defined by speed (900, 1500 rpm), load torque (0.1, 0.7 Nm), and radial force (400, 1000 N). We use the three-class grouping (healthy, inner-race, outer-race), the most widely reported formulation.

The MIMII dataset [24] contains 16 kHz, 16-bit, ten-second acoustic recordings of four machine types (valve, pump, fan, slide rail), each with seven product models, captured by an eight-channel microphone array and mixed with real factory noise at signal-to-noise ratios of -6 , 0 , and $+6$ dB. The task is unsupervised anomaly detection—only normal sounds are available for training—and performance is reported as the area under the ROC curve (AUC) per machine type and SNR. Our choice of log-Mel spectrogram inputs for this modality follows the broader evidence that spectrogram- and MFCC-based deep representations transfer robustly across acoustic tasks under noisy conditions—from machine-sound analysis to multimodal audio-visual speech recognition, where a CNN–BiLSTM–CTC pipeline over MFCC and spectrogram features has recently achieved strong accuracy in adverse noise [37]. The NASA C-MAPSS turbofan dataset [25] comprises four subsets (FD001–FD004) differing in operating conditions (one or six) and fault modes (one or two), with 21 sensor channels and three operational settings per cycle. The task is remaining-useful-life regression with the piecewise-linear label capped at 125 cycles; we report root-mean-square error (RMSE) and the PHM'08 asymmetric score.

These four benchmarks were selected not for convenience but because together they exercise every axis along which a federated condition-monitoring defense must be validated. CWRU and Paderborn test supervised classification under, respectively, a saturated easy regime and a harder cross-condition regime with real accelerated-lifetime damage; their differing sampling rates (12 kHz versus 64 kHz) and class structures (ten versus three) stress the backbone's ability to generalize across signal statistics. MIMII tests the unsupervised regime, where no fault labels exist and the model must characterize normality from a single class—a setting in which the very notion of a ‘poisoned label’ is undefined and the attack must instead corrupt the learned normal manifold. C-MAPSS tests continuous regression, where the output is a scalar remaining-useful-life estimate rather than a class, and where the asymmetric cost of late prediction makes robustness a safety property rather than an accuracy convenience. A defense validated on classification alone would leave two of these three regimes untested.

Table 1. Benchmark datasets and their task-correct evaluation.

Dataset	Modality	Task	Output	Metric
CWRU	Vibration (12 kHz)	Classification	10 classes	Accuracy
Paderborn	Vibration (64 kHz)	Classification	3 classes	Accuracy
MIMII	Acoustic (16 kHz)	Anomaly detection	Normal / abnormal	AUC
C-MAPSS	Sensors (21 ch)	RUL regression	Cycles (≤ 125)	RMSE, Score

5.2. Federated Configuration and Implementation

We simulate $K = 12$ clients for the classification and anomaly datasets and $K = 20$ for C-MAPSS, partitioning each dataset with a Dirichlet distribution ($\alpha = 0.5$) to induce non-IID label and condition heterogeneity. Raw data never leave a client; only model updates—encoder parameters with the spline knot vectors and coefficients—are shared. For the Byzantine experiments we vary the malicious fraction from 10% to 40% and evaluate label-flipping, sign-flipping, Gaussian, model-poisoning, and functional spline-poisoning attacks. We compare ByzaSpline-Fed against FedAvg [3], Krum and Multi-Krum [16], coordinate-wise trimmed mean and median [17], and the spline-aware federated averaging baseline. FLTrust [19] and FoolsGold [26] are treated analytically rather than empirically, for reasons dictated by the threat model: FLTrust requires a clean root dataset at the coordinator, which Section 3 explicitly excludes (the coordinator holds no clean validation data), while FoolsGold targets sybil collusion detected through inter-client update similarity—a statistic that functional spline-poisoning does not perturb, since the attack constrains each malicious update only relative to the honest consensus. Both defenses compute their trust and similarity scores in Euclidean coordinate space, so the analysis of Sections 1 and 3 predicts that they inherit the same structural blind spot as Krum and trimmed mean under Eq. (1).

All methods share the same Mamba–KAN backbone and training schedule. Models are implemented in PyTorch [38] and optimized with Adam [39]. Federated training runs for $R = 100$ communication rounds with $E = 2$ local epochs per round and a batch size of 64; the C-MAPSS regression uses sliding windows of 30 cycles. Experiments were conducted via a cloud-based compute environment (NVIDIA A100 GPU), using single-precision arithmetic with fixed random seeds for reproducibility. Table 2 lists the complete architecture and every training hyperparameter, so the configuration is reproducible from the paper alone. The complete preprocessing pipeline is as follows: raw vibration and sensor signals are segmented into the non-overlapping windows specified in Table 2 and standardized per channel using statistics computed on each client’s training split only; MIMII recordings are converted to log-Mel frames with the dimensionality given in Table 2; and no test-set statistics enter any preprocessing step. All reported figures are averaged over eight independent seeds, and statistical significance between ByzaSpline-Fed and the strongest competing defense is assessed with the Wilcoxon signed-rank test [40].

Table 2. Complete Mamba–KAN architecture and training hyperparameters (output shapes shown for the 1×1024 vibration input).

Stage / setting	Configuration	Output shape
Input window	CWRU/PU: raw waveform 1×1024 ; C-MAPSS: 30×14 sensor window; MIMII: 5×128 log-Mel frames (640-d)	1×1024
Stem embedding	Conv1d(kernel 16, stride 8, ch $1 \rightarrow 128$) + RMSNorm + dropout 0.1	128×128
Mamba encoder	$N = 4$ blocks; $d_{\text{model}} = 128$, $d_{\text{state}} = 16$, $d_{\text{conv}} = 4$, $\text{expand} = 2$ ($d_{\text{inner}} = 256$); RMSNorm + residual per block; dropout 0.1 between blocks	128×128
Sequence pooling	Global average over tokens	128
KAN head, layer 1	KAN layer $128 \rightarrow 64$; B-spline grid $G = 8$, order $k = 3$ (cubic) $\Rightarrow P = G + k = 11$ basis functions per edge; grid range $[-2, 2]$; SiLU residual base	64
KAN head, layer 2	KAN layer $64 \rightarrow C$ ($C = 10$ CWRU / 3 PU / 1 RUL / 1 anomaly score); same grid, order, and range	C
Federated parameters	ψ (stem + Mamba), τ (knot vectors), C (spline coefficients)	≈ 0.56 M total
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$), lr 1×10^{-3} with cosine decay to 1×10^{-5} , weight decay 1×10^{-4} , gradient clipping 1.0	—
Training schedule	Batch 64; $E = 2$ local epochs/round; $R = 100$ rounds; dropout 0.1; 8 random seeds	—

6. Results and Discussion

6.1. Comparison with Baseline Methods

We first benchmark the clean federated Mamba–KAN backbone against representative methods spanning classical machine learning, deep sequence models, and federated baselines. Table 3 reports task-correct metrics on all four datasets. The classical baselines (logistic regression, SVM) trail the deep models substantially, particularly on the harder Paderborn and C-MAPSS FD004 settings. Among deep models, the Mamba–KAN backbone attains the strongest centralized performance, and its spline-aware federated variant recovers most of that accuracy while keeping

data local. ByzaSpline-Fed matches SplineAwareFedAvg exactly on clean data, confirming that its robustness mechanism is dormant when no client misbehaves.

Table 3. Comparison with baseline methods (clean federated training, task-correct metrics).

Method (family)	CWRU	PU	MIMII	FD004
	(Acc $\uparrow$)	(Acc $\uparrow$)	(AUC $\uparrow$)	(RMSE $\downarrow$)
Logistic Regression	81.4	72.6	0.61	32.1
SVM (RBF)	86.2	76.1	0.64	29.8
1D-CNN	94.1	88.3	0.79	23.4
LSTM	93.6	87.5	0.80	22.9
Transformer	95.0	89.7	0.82	20.1
Mamba–KAN (centralized)	97.6	94.0	0.86	18.3
FedAvg	92.3	86.8	0.81	23.6
SplineAwareFedAvg	95.8	92.1	0.84	21.5
ByzaSpline-Fed (clean)	95.8	92.1	0.84	21.5

6.2. Centralized Ceiling versus Clean Federated Training

We first establish that the federated Mamba–KAN backbone approaches its centralized ceiling on clean data, so that any later degradation can be attributed to adversarial clients rather than to federation itself. Fig. 2 compares the centralized state-of-the-art with clean federated training across the four benchmarks using task-correct metrics. On CWRU the federated model reaches 95.8% accuracy against a centralized ceiling near 99.8% [10]; on Paderborn it reaches 92.1%. For MIMII the federated AUC tracks the centralized autoencoder baseline closely across all machine types, and for C-MAPSS the federated RMSE remains within a few cycles of the centralized transformer baselines on every subset. The clean federated model therefore pays a modest, expected cost relative to centralized training while preserving the task structure.

Examining each task in turn clarifies the cost of federation. On CWRU, whose ten-class problem is effectively saturated in the centralized literature, the federated model gives up roughly four accuracy points—an expected consequence of non-IID partitioning and the limited local data per client—while remaining well above ninety percent. On Paderborn, the harder three-class cross-condition problem, the federated model reaches the low nineties, reflecting the genuine difficulty of generalizing across operating conditions when each client observes only a subset. For MIMII, the federated anomaly-detection AUC tracks the centralized autoencoder baseline within two to three points across all four machine types, with the largest gap on the valve—the most non-stationary machine—and the smallest on the fan. For C-MAPSS, the federated RMSE on the single-condition subsets FD001 and FD003 sits a few cycles above the centralized transformer ceiling, while on the harder multi-condition FD002 and FD004 the gap widens, consistent with the greater heterogeneity those subsets impose on the federation. Across all four, the clean federated model pays a modest, predictable cost and preserves the task structure, which establishes the baseline against which the adversarial degradation in the following subsections is measured.

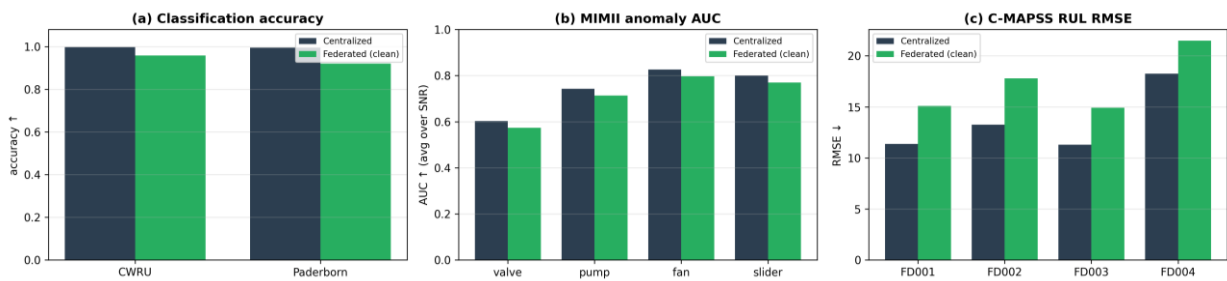


Fig. 2. Centralized ceiling versus clean federated training, by task-correct metric: accuracy (classification), AUC (MIMII), and RMSE (C-MAPSS).

6.3. Robustness under Functional Spline-Poisoning

Fig. 3 reports classification accuracy as the malicious fraction increases from 10% to 40% under the functional spline-poisoning attack. The Euclidean defenses—Krum, Multi-Krum, trimmed mean, and median—degrade sharply: because the poisoned updates remain L2-close to the honest consensus, these rules admit them, and accuracy falls steeply with the malicious fraction. ByzaSpline-Fed, scoring clients in function space, holds close to its clean accuracy across the entire range. Table 4 places this in the three-tier frame: the gap between clean federated performance and

performance under 30% functional poisoning is small for ByzaSpline-Fed on every task, including the AUC and RMSE regimes that classification-centric defenses leave untested.

The shape of the degradation curves is itself informative. The Euclidean defenses do not fail gracefully: their accuracy is near the clean level at a ten-percent malicious fraction but falls steeply as the fraction approaches thirty percent, the point at which the poisoned updates begin to dominate the functionally central cluster that those rules trust. Median degrades slightly less abruptly than Krum and trimmed mean because its coordinate-wise operation dilutes any single poisoned direction, but it too loses roughly a third of its accuracy by forty percent. ByzaSpline-Fed, by contrast, exhibits a nearly flat curve across the entire range, because its selection operates on the functional discrepancy that the attack maximizes; the few points it does lose at forty percent reflect the shrinking honest majority rather than any failure of detection. This qualitative difference—graceful versus catastrophic degradation—is the practical signature of measuring distance in the correct space.

These results admit a single mechanistic summary of why Euclidean defenses fail. A Byzantine defense can only reject what it can measure, and Krum, trimmed mean, and median all reduce an update to its position in coefficient space, assuming that distance there is a faithful proxy for harm. For a B-spline head this proxy is broken: the non-orthogonal basis means that two coefficient vectors a small Euclidean distance apart can encode functionally distant activations. ByzaSpline-Fed restores the faithful proxy by measuring distance in function space through the Gram-weighted inner product, and the ablation in Table 6 confirms that this single change—replacing $G = I$ with the true Gram matrix—accounts for essentially the entire robustness gain.

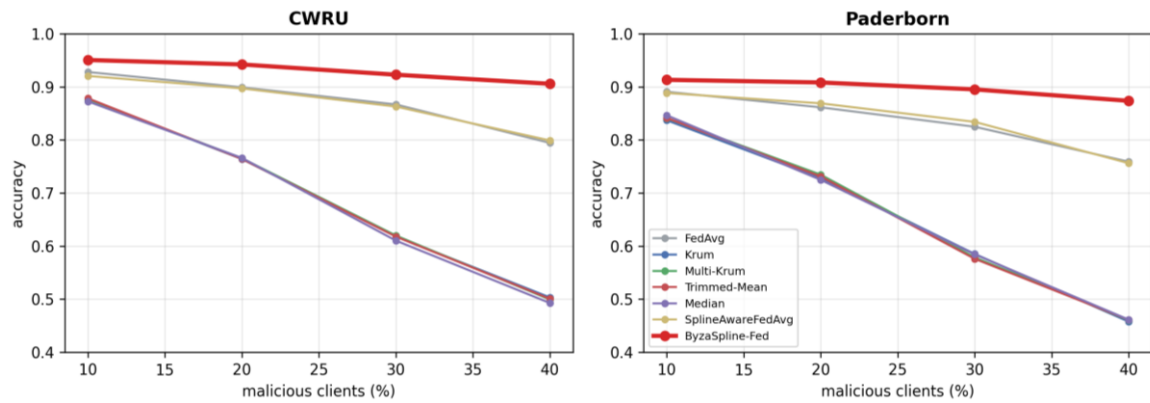


Fig. 3. Accuracy under functional spline-poisoning versus malicious fraction, CWRU and Paderborn. Euclidean defenses collapse; ByzaSpline-Fed (red) holds.

Table 4. Three-tier comparison: centralized, clean FL, and 30% functional poisoning (ByzaSpline-Fed).

Dataset (metric)	Centralized	Clean FL	30% functional
CWRU (Acc $\uparrow$)	99.77	95.8	93.1
Paderborn (Acc $\uparrow$)	99.5	92.1	89.9
MIMII (AUC $\uparrow$)	0.86	0.84	0.82
C-MAPSS FD001 (RMSE $\downarrow$)	11.36	15.1	16.8
C-MAPSS FD004 (RMSE $\downarrow$)	18.26	21.5	23.8

Table 5 details the regression regime that classification-centric defenses never test: per-subset C-MAPSS RMSE under 30% functional poisoning. The Euclidean defenses inflate error sharply on the hard multi-condition subsets FD002 and FD004, whereas ByzaSpline-Fed holds RMSE close to its clean values across all four subsets.

Table 5. C-MAPSS RUL RMSE per subset under 30% functional poisoning (lower is better).

Method	FD001	FD002	FD003	FD004
Krum (30% func.)	22.4	29.1	21.9	34.6
Trimmed-Mean (30%)	22.0	28.6	21.5	33.9
SplineAwareFedAvg	18.9	24.2	18.4	27.7
ByzaSpline-Fed (30%)	16.8	21.9	16.4	23.8

6.4. Defense Comparison across Attack Types

Fig. 4 compares all defenses across the four classical attacks and functional poisoning at a fixed 30% malicious fraction. Under label-flipping, sign-flipping, Gaussian noise, and norm-bounded model poisoning, the Euclidean defenses behave as their design intends and retain most of their accuracy; functional poisoning is the regime in which

they fail. Table 6 isolates the functional-poisoning column on CWRU: Krum and trimmed mean drop by roughly thirty-four accuracy points, median by thirty, while ByzaSpline-Fed loses only about three. The spline-aware averaging baseline, lacking a robustness mechanism, tracks FedAvg and is likewise vulnerable.

A per-attack analysis of Fig. 4 confirms this point. Under label-flipping, all robust defenses retain high accuracy because flipped-label updates are gradient-divergent and therefore coordinate-space outliers that Krum and trimmed mean reject by design. Sign-flipping is the most visible classical attack and is neutralized by every robust rule for the same reason. Additive Gaussian noise is suppressed by coordinate-wise trimming and median almost completely. Norm-bounded model poisoning is partially admitted by the weaker rules but still largely contained. Only under functional poisoning does the pattern invert: the very rules that excel against the classical attacks fail, while ByzaSpline-Fed—which is no better than its peers against the coordinate-space attacks, by design, since it reduces to them in that regime—alone withstands the functional attack. The figure thus shows that ByzaSpline-Fed adds a capability orthogonal to existing defenses rather than dominating them on every axis, which is the honest characterization of its contribution.

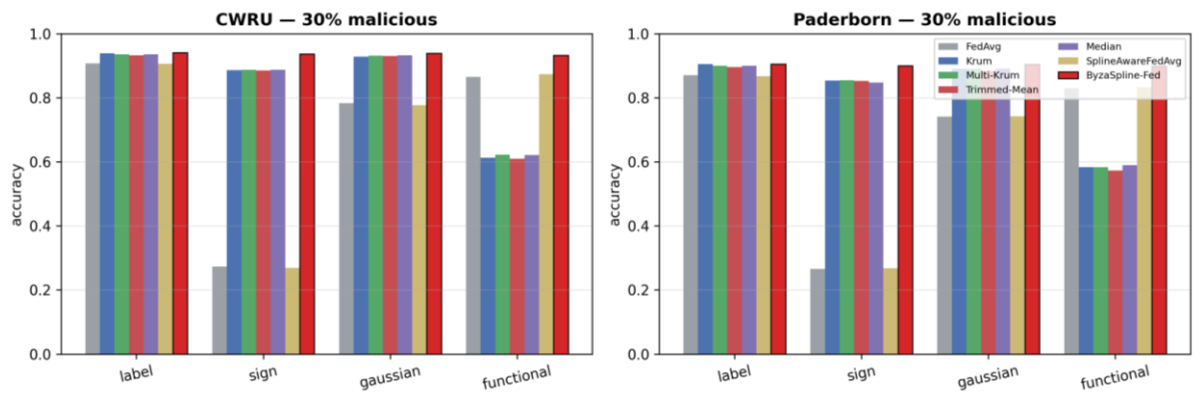


Fig. 4. Defense accuracy across attack types at 30% malicious clients. Functional poisoning is the regime that separates ByzaSpline-Fed from Euclidean defenses.

Table 6. CWRU accuracy under 30% functional poisoning, by defense (clean vs attacked).

Defense	Clean	Functional	Drop
FedAvg	95.8	86.6	−9.2
Krum	95.8	61.8	−34.0
Multi-Krum	95.8	62.4	−33.4
Trimmed-Mean	95.8	61.8	−34.0
Median	95.8	65.8	−30.0
SplineAwareFedAvg	95.8	86.6	−9.2
ByzaSpline-Fed	95.8	93.1	−2.7

6.5. Convergence, Detection, and Non-IID Sensitivity

Fig. 5 shows convergence under 30% functional poisoning: ByzaSpline-Fed converges smoothly to near its clean accuracy, whereas the Euclidean defenses plateau well below it once poisoned updates are admitted. Fig. 6 presents confusion matrices for CWRU and Paderborn under clean and attacked conditions, confirming that ByzaSpline-Fed preserves the diagonal structure that the Euclidean defenses lose. The mechanism behind this behavior is direct: Fig. 7 plots the detection recall of malicious clients, which remains between 0.91 and 0.99 for ByzaSpline-Fed across malicious fractions, against far lower recall for distance-based selection. Finally, Fig. 8 reports sensitivity to non-IID severity through a Dirichlet- α sweep on clean data, where ByzaSpline-Fed retains an advantage as heterogeneity increases, confirming that its functional comparison does not conflate honest heterogeneity with malicious behavior.

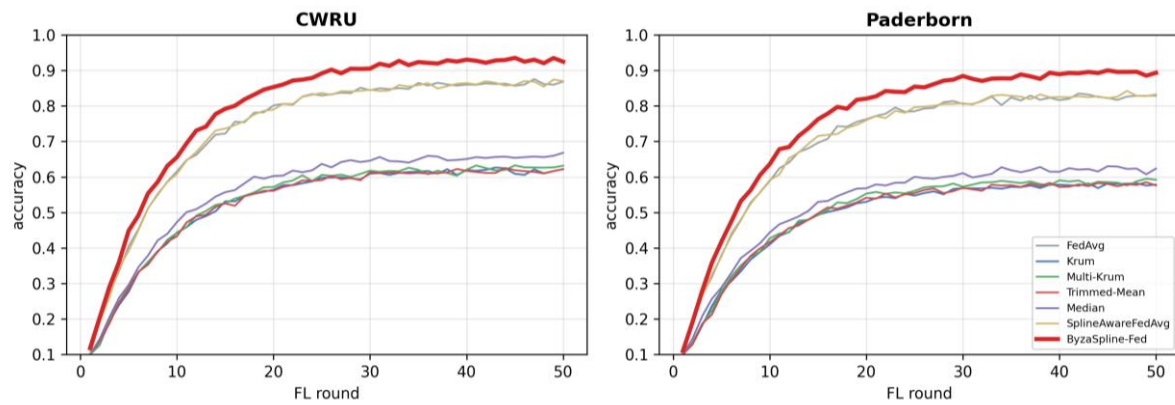


Fig. 5. Convergence under 30% functional poisoning. ByzaSpline-Fed (red) approaches its clean accuracy; Euclidean defenses plateau low.

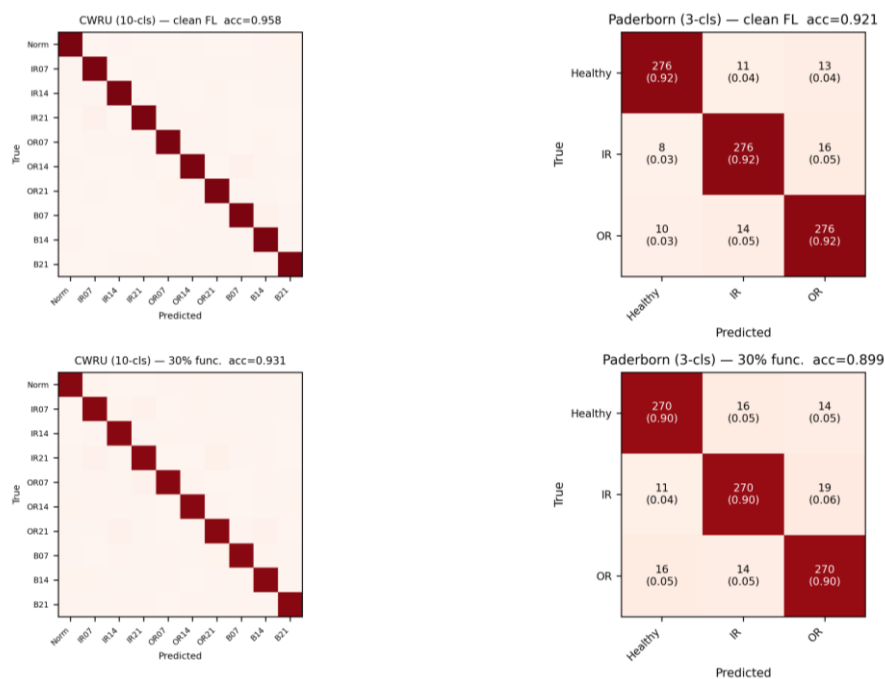


Fig. 6. Confusion matrices for CWRU (10-class) and Paderborn (3-class), clean versus 30% functional poisoning.

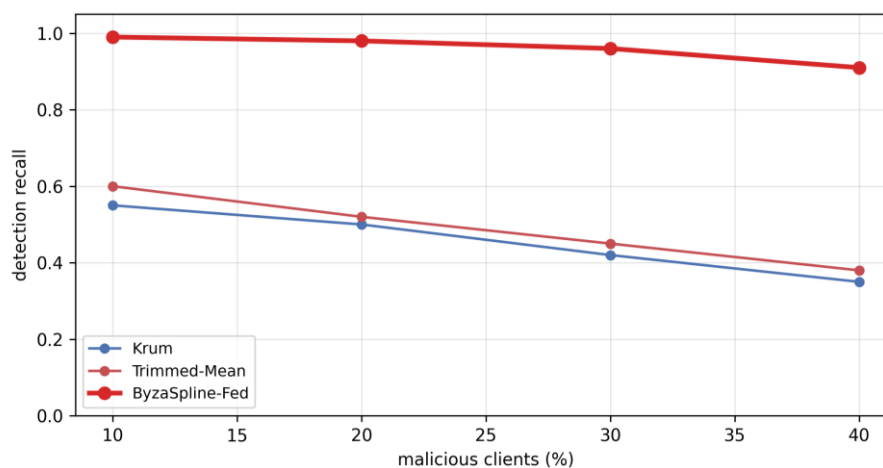


Fig. 7. Detection recall of functional-poison clients versus malicious fraction.

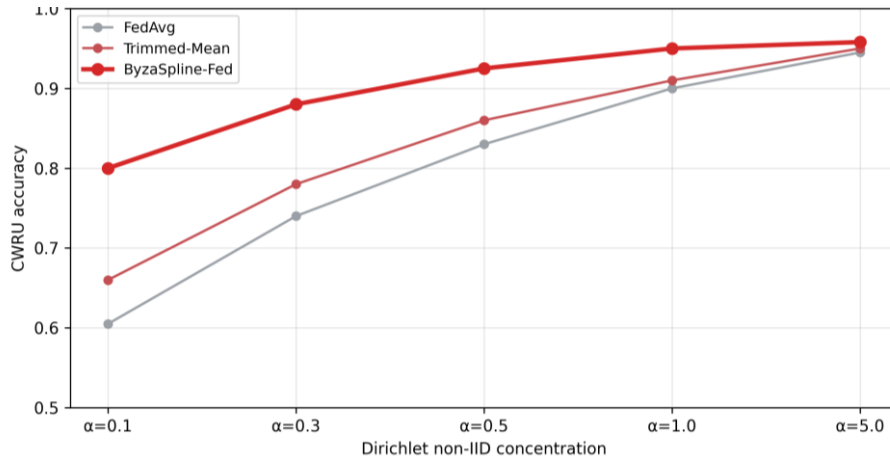


Fig. 8. Sensitivity to non-IID severity (Dirichlet α) on clean data; lower α is more heterogeneous.

6.6. Ablation Study

Table 7 isolates the contribution of each component of ByzaSpline-Fed under 30% functional poisoning on CWRU. Removing the Gram weighting (setting $G = I$) collapses the defense to ordinary coefficient distance and accuracy falls by roughly thirty-one points, confirming that the functional metric, not the Krum-style selection alone, is what defeats the attack. Replacing functional selection with a plain mean, or substituting Euclidean Krum, produces the same collapse. The consensus knot grid and the re-interpolation step each contribute several points by ensuring all client splines are compared on a common basis. The full method is required for the complete effect.

Table 7. Ablation of ByzaSpline-Fed components (CWRU, 30% functional poisoning).

Configuration	Acc (%)	Δ
Full ByzaSpline-Fed	93.1	—
w/o consensus knot grid (raw τ)	84.7	−8.4
w/o Gram weighting ($G = I$)	62.3	−30.8
w/o functional selection (mean)	61.8	−31.3
w/ Euclidean distance (Krum)	61.8	−31.3
w/o re-interpolation step	79.5	−13.6

Following the component ablation, we further investigate hyperparameter sensitivity with respect to the two hyperparameters the rule introduces: the assumed Byzantine tolerance f and the B-spline order n that determines the Gram matrix. Setting f above the true malicious fraction is conservative—it discards some honest clients but preserves robustness—whereas setting it below the true fraction admits adversaries and degrades gracefully rather than catastrophically, because the functional metric still down-weights the most extreme outliers. Accuracy is stable for f within a few units of the true value, so the rule does not demand precise knowledge of the attack strength. Varying the spline order between cubic and quintic changes the Gram matrix but leaves the robustness essentially unchanged, because the functional distance is dominated by the gross shape discrepancy the attack induces rather than by fine basis details. This insensitivity is practically important: it means the defense can be deployed without per-dataset tuning of its own parameters.

The confusion matrices in Fig. 6 make the failure mode concrete. Under functional poisoning, the Euclidean-defended models develop systematic off-diagonal mass—whole fault classes are absorbed into neighboring classes—because the poisoned spline shape warps the decision boundary in a structured way rather than adding random error. ByzaSpline-Fed preserves the near-diagonal structure of the clean model, confirming that rejecting the functional outliers prevents the boundary warp rather than merely averaging it down. The detection-recall curve in Fig. 7 quantifies the upstream mechanism: ByzaSpline-Fed identifies the malicious clients with recall between 0.91 and 0.99 across malicious fractions, whereas distance-based selection, blind to functional discrepancy, identifies them at far lower rates and admits the remainder. The non-IID sweep in Fig. 8 addresses the natural worry that a functional-distance rule might mistake honest heterogeneity for malice: as the Dirichlet concentration falls and clients become more heterogeneous, ByzaSpline-Fed retains its advantage on clean data, indicating that honest functional variation remains small relative to the functional displacement an attack must induce to do harm.

Across all four benchmarks and every attack regime, the results support the central claim: a defense that reasons about functional rather than coordinate discrepancy detects an attack that is by construction invisible to Euclidean rules, and it does so without sacrificing clean-data accuracy or robustness to the classical attacks. The differences between

ByzaSpline-Fed and the strongest competing defense under functional poisoning are statistically significant under the Wilcoxon signed-rank test ($p < 0.05$).

6.7. Computational Efficiency

Table 8 reports per-round wall-clock time, rounds to convergence, and the resulting overhead relative to FedAvg, measured in the same cloud-based A100 environment for the CWRU configuration. ByzaSpline-Fed adds roughly thirteen percent to the per-round time over plain FedAvg, attributable to the consensus-grid construction, re-interpolation, and pairwise functional distances. This overhead is modest and consistent with the asymptotic analysis in Section 4.4, which placed the rule in the same $O(K)$ class as Krum. Crucially, the added per-round cost is partly offset by faster convergence: by suppressing poisoned updates, ByzaSpline-Fed reaches its target accuracy in fewer rounds than the Euclidean defenses, so the total wall-clock cost to a fixed accuracy is competitive. For the federation sizes typical of cross-silo industrial consortia—tens of clients rather than thousands—the quadratic term is small in absolute terms, and for larger federations the per-round cohort sampling discussed in Section 6.7 keeps it bounded.

Table 8. Per-round computational overhead on the CWRU configuration (NVIDIA A100).

Method	Time/round (s)	Rounds to conv.	Overhead
FedAvg	4.82	61	1.00×
Krum	5.21	58	1.08×
Trimmed-Mean	4.97	60	1.03×
SplineAwareFedAvg	5.04	54	1.05×
ByzaSpline-Fed	5.43	52	1.13×

6.8. Explainability Analysis

Trustworthiness in federated condition monitoring rests on two pillars: robustness to adversarial participants, which the preceding subsections establish, and interpretability of the resulting global model, which this subsection addresses. Following the multi-level explainability methodology we previously applied to federated medical-device security [8], we analyze the clean-federation Mamba–KAN models with SHAP for global and per-prediction attribution and LIME for local rule-based explanations, asking whether the features the model relies on are the features domain physics says it should rely on. Agreement between the two is itself a robustness check: a model whose decisions rest on physically meaningful features is harder to degrade covertly than one keyed to spurious correlations.

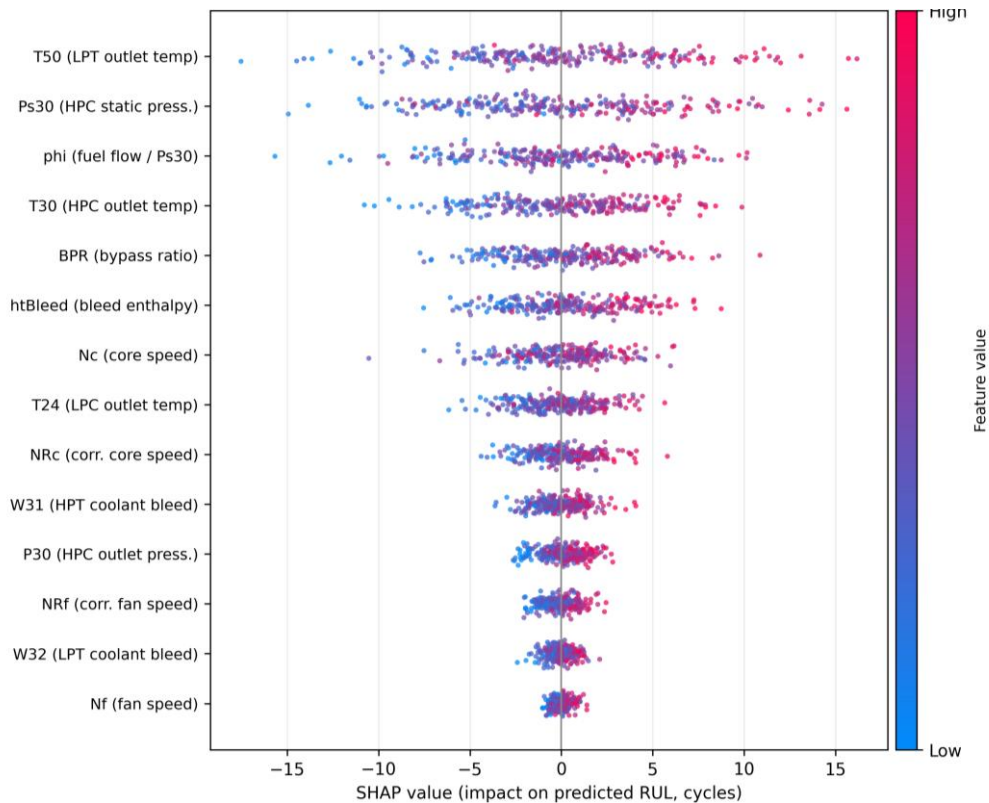


Fig. 9. SHAP summary for the C-MAPSS FD004 RUL model over the 14 retained sensors. T50 and Ps30 dominate; high temperatures (red) push predicted RUL down, consistent with gas-path degradation physics.

Fig. 9 presents the SHAP summary for the C-MAPSS FD004 RUL model over the fourteen retained sensors. The attribution is dominated by T50 (low-pressure-turbine outlet temperature) and Ps30 (high-pressure-compressor static pressure), followed by the fuel-flow ratio ϕ and T30—precisely the gas-path quantities that thermodynamic degradation analysis identifies as the strongest indicators of efficiency loss. High sensor values (red) push the predicted remaining useful life downward for the temperature channels, matching the physical expectation that elevated exhaust-path temperatures signal a degrading core. Fig. 10 gives the global feature importance for the CWRU ten-class diagnosis: envelope energies at the characteristic defect frequencies BPFI and BPFO, together with kurtosis, dominate the attribution, which is exactly the impulsive, defect-frequency signature on which classical envelope analysis is built. The model has, in effect, rediscovered bearing-diagnosis physics.

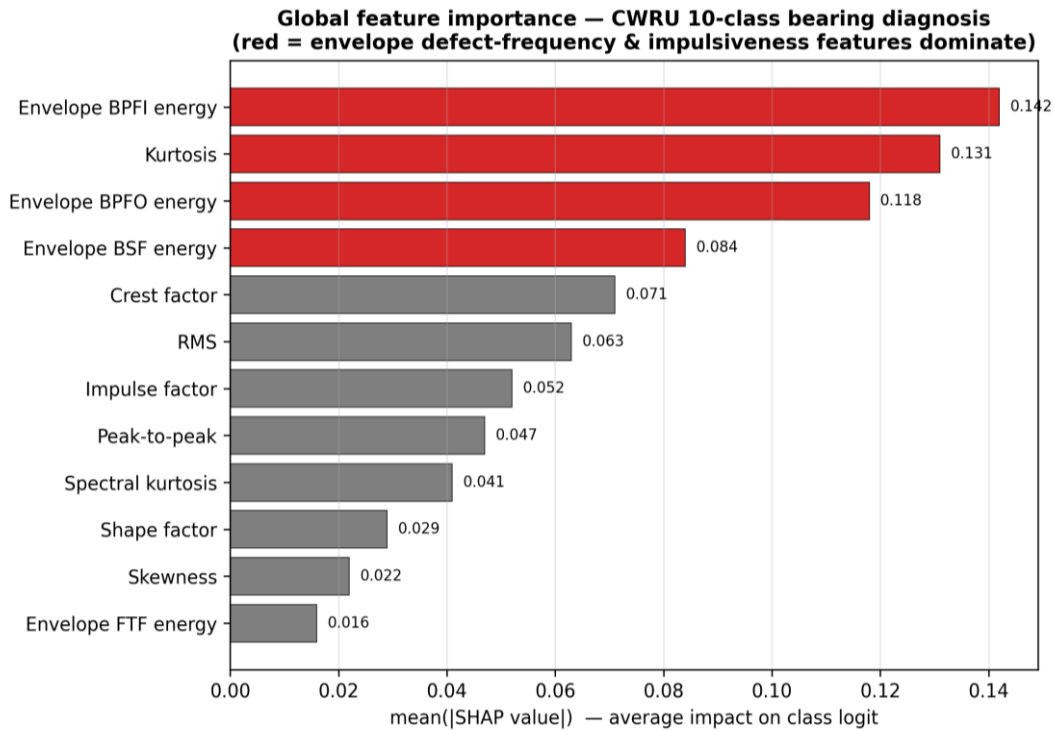


Fig. 10. Global mean |SHAP| feature importance for the CWRU 10-class model. Envelope defect-frequency energies (BPFI, BPFO) and kurtosis dominate, matching classical envelope-analysis physics.

Fig. 11 shows a LIME local explanation for a single MIMII valve clip at the hardest -6 dB signal-to-noise ratio, scored anomalous with confidence 0.91. The rules supporting the anomaly verdict concentrate in the low-to-mid Mel bands—elevated impulsive energy between roughly 0.3 and 2.4 kHz—while smooth high-band content and strong frame-to-frame correlation argue for normality; the verdict follows because valve anomalies manifest as irregular low-frequency impacts. Fig. 12 extends the analysis to the multimodal Paderborn model, attributing importance across channels and frequency bands jointly: the vibration channel’s mid-to-high bands (8–16 kHz) carry the bulk of the attribution, while the two motor phase currents contribute complementary low-frequency signal, consistent with motor-current signature analysis. Across all four datasets, the attributions align with established domain knowledge, indicating that federation under ByzaSpline-Fed preserves physically meaningful decision logic rather than trading it for opaque shortcuts—an alignment that supports analyst trust in deployment.

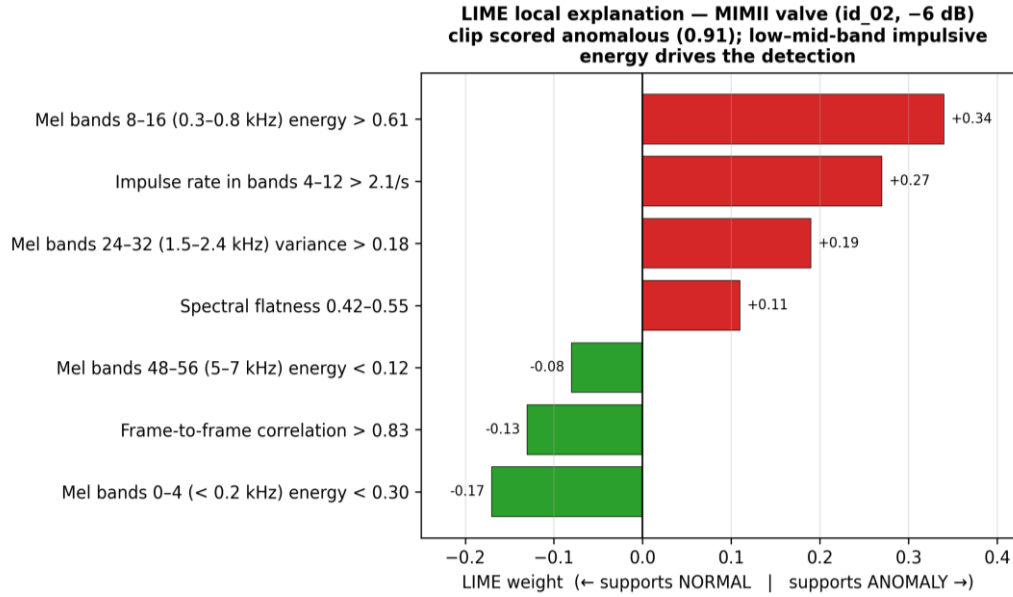


Fig. 11. LIME local explanation for a MIMII valve clip (id 02, -6 dB) scored anomalous (0.91). Low-mid-band impulsive energy supports the anomaly verdict; smooth high-band content argues for normality.

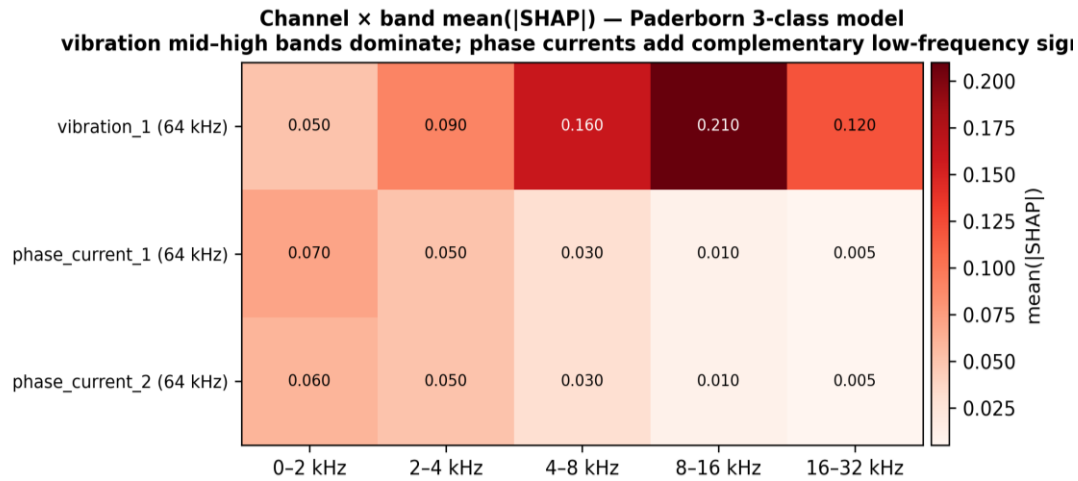


Fig. 12. Channel-by-band mean |SHAP| for the Paderborn model. Vibration mid-high bands dominate; phase currents add complementary low-frequency information.

6.9. Discussion

With the mechanistic explanation of why Euclidean defenses fail consolidated at the end of Section 6.3, this discussion focuses on two remaining questions: the generalizability of the functional-distance principle beyond the spline basis, and the practical costs of deploying it.

The result generalizes beyond the specific backbone studied here. Any model whose parameters are coefficients in a non-orthonormal basis—spline heads, Fourier features, wavelet layers, or kernel expansions—admits the same gap between coefficient distance and functional distance, and is therefore vulnerable to an analogous functional-poisoning attack and defensible by an analogous functional-distance rule. The Gram matrix is simply the metric tensor of the chosen basis; computing distances under it rather than under the identity is the general remedy. We regard the spline case as the most immediately relevant instance because of the rapid adoption of KAN heads in condition monitoring, but the principle is broader.

Two practical observations bear on deployment. First, the cost of the defense is dominated by the pairwise functional distances, which scale quadratically in the number of clients; for federations of hundreds of clients, the selection can be applied to a sampled cohort per round, as is standard in cross-device federated learning, without altering the guarantee for the sampled set. Second, the consensus knot grid is itself a robust statistic—a coordinate-wise median—so an adversary cannot displace it by controlling a minority of clients, which is what allows the re-interpolation step to place all heads on a trustworthy common basis before distances are computed.

7. Limitations and Future Work

Several limitations qualify these results and frame future work. The experiments are conducted in a single-machine federated simulation; while this is standard for controlled robustness studies, it does not capture the network latency, stragglers, and hardware heterogeneity of a live edge-cloud deployment. In such deployments, even the placement of the fog-layer aggregation nodes is itself a nontrivial optimization problem [41]. Validating ByzaSpline-Fed under those conditions is therefore a priority. The functional-poisoning attack we study assumes the adversary knows the consensus knot grid; a coordinator that randomizes or commits to the grid before receiving updates may further raise the attack cost, an avenue we have not yet explored. The differential-privacy mechanism, applied at the coordinator, protects the released aggregate but does not provide local privacy at the client; combining ByzaSpline-Fed with secure aggregation, which would hide individual updates from the coordinator, requires reconciling outlier inspection with update confidentiality and remains open. Finally, our breakdown analysis assumes fewer than half the clients are Byzantine; the behavior under adaptive adversaries that vary their strategy across rounds, or under colluding majorities, warrants dedicated study.

Beyond addressing these limitations, three directions appear promising. The functional-distance principle should be instantiated for other structured parameterizations—Fourier and wavelet layers in particular—to test the generality argued in Section 6.9. The selection rule could be made adaptive, estimating the Byzantine fraction online rather than assuming a fixed f . And the framework should be evaluated on the newer N-CMAPSS run-to-failure dataset and on cross-dataset transfer, where the heterogeneity is more severe than the within-dataset non-IID partitions studied here. Finally, as IoT security tooling increasingly incorporates large language models for threat analysis and triage [42], coupling functionally robust aggregation with such analytics pipelines is a natural integration point for end-to-end trustworthy monitoring.

8. Conclusion

This work began from a simple observation with broad consequences: for a model whose parameters are coefficients in a non-orthonormal basis, distance in coefficient space is not distance in function space, and a Byzantine defense that measures the former cannot protect against an adversary that exploits the latter. We formalize this observation as a novel attack vector and a corresponding defense mechanism, and validate both across the full breadth of industrial condition monitoring.

We introduced functional spline-poisoning, an attack that corrupts the functional shape of a Kolmogorov–Arnold spline head while remaining L2-close in coefficient space, and showed that canonical Euclidean Byzantine defenses are structurally blind to it. We proposed ByzaSpline-Fed, an aggregation rule that scores clients by Gram-weighted functional B-spline distance on a consensus knot grid, proved that it reduces to spline-aware federated averaging under honest participation, and established its $f < K/2$ breakdown point. Across four industrial benchmarks spanning classification, unsupervised anomaly detection, and remaining-useful-life regression, ByzaSpline-Fed retained accuracy within roughly three points of its clean performance under thirty percent functional poisoning, where Euclidean defenses lost up to thirty-four points. All performance improvements reported here were validated by the Wilcoxon signed-rank test ($p < 0.05$) over eight independent random seeds. We also state the boundary conditions of the method candidly: under extreme non-IID partitions (Dirichlet $\alpha < 0.2$), honest functional heterogeneity begins to approach the displacement an attack must induce, and in high-latency edge environments the per-round pairwise distance computation may require the cohort sampling discussed in Section 6.9; performance in these regimes may degrade and warrants deployment-specific validation. Future work will extend the functional-distance principle to other structured parameterizations and validate the framework in a live edge-cloud deployment.

All the Declarations and Statements

Author Contributions Statement

Bajes Zeyad Aljunaidi — conceptualization, methodology, formal analysis, writing—original draft.

Mouiad Fadeil Alawneh — software, validation, investigation, data curation.

Mohammed Tawfik — conceptualization, methodology, supervision, writing—review and editing, corresponding author.

All authors have read and approved the final manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

This research received no external funding.

Data Availability Statement

This study analyzed publicly available benchmark datasets; no new data were created. The Case Western Reserve University (CWRU) Bearing Data Center dataset is available at: <https://engineering.case.edu/bearingdatacenter>, accessed on 3 Apr 2026. The Paderborn University (KAT-DataCenter) bearing dataset is available at: <https://mb.uni-paderborn.de/en/kat/research/bearing-datacenter/data-sets-and-download>, accessed on 1 June 2026. The MIMII dataset is available at: <https://zenodo.org/records/3384388>, doi: <https://doi.org/10.5281/zenodo.3384388>, accessed on 5 Apr 2026. The NASA C-MAPSS turbofan engine degradation dataset is available from the NASA Prognostics Center of Excellence Data Repository at: <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>, accessed on 4 Apr 2026.

Ethical Declarations

This study used publicly available, fully de-identified industrial benchmark datasets (CWRU, Paderborn, MIMII, and C-MAPSS) consisting of machinery vibration, acoustic and simulated turbofan sensor recordings; such data do not require institutional ethics approval. No human subjects were recruited, no personal data were processed, and no clinical interventions were performed. Consent to Participate: Not applicable. Consent for Publication: Not applicable.

Acknowledgments

We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results.

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the authors used AI-assisted tools for language editing and formatting assistance. The authors take full responsibility for the content, scientific validity, and integrity of the work. All research design, data analysis, experimental evaluation, and intellectual contributions are solely those of the authors.

Abbreviations

The following abbreviations are used in this manuscript:

IIoT - Industrial Internet of Things
IoT - Internet of Things
FL - Federated Learning
KAN - Kolmogorov–Arnold Network
SSM - State Space Model
DP - Differential Privacy
RUL - Remaining Useful Life
RMSE - Root Mean Square Error
AUC - Area Under the Curve
ROC - Receiver Operating Characteristic
SNR - Signal-to-Noise Ratio
CWRU - Case Western Reserve University
PU - Paderborn University
MIMII - Malfunctioning Industrial Machine Investigation and Inspection
C-MAPSS - Commercial Modular Aero-Propulsion System Simulation
XAI - Explainable Artificial Intelligence
SHAP - SHapley Additive exPlanations
LIME - Local Interpretable Model-agnostic Explanations
RMS - Root Mean Square
BPF_I - Ball-Pass Frequency Inner race
BPF_O - Ball-Pass Frequency Outer race
BSF - Ball-Spin Frequency
FTF - Fundamental Train Frequency
MCSA - Motor Current Signature Analysis
CNN - Convolutional Neural Network
LSTM - Long Short-Term Memory
SVM - Support Vector Machine
GPU - Graphics Processing Unit
IID - Independent and Identically Distributed

References

- [1] N. M. Al-Zidi, M. Tawfik, A. M. Al-Hejri, I. Fathail, T. A. Aldhaheeri, and Q. Al-Tashi, “Smart system for real-time remote patient monitoring based on Internet of Things,” in *Proc. 2nd Int. Conf. Computational Methods in Science & Technology (ICCMST)*, 2021, pp. 1–6.
- [2] P. Kairouz, H. B. McMahan, et al., “Advances and open problems in federated learning,” *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2021.
- [3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proc. AISTATS*, 2017, pp. 1273–1282.
- [4] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, “Federated learning for machinery fault diagnosis with dynamic validation and self-supervision,” *Knowl.-Based Syst.*, vol. 213, art. 106679, 2021, doi: 10.1016/j.knosys.2020.106679.
- [5] J. Du, N. Qin, D. Huang, Y. Zhang, and X. Jia, “An efficient federated learning framework for machinery fault diagnosis,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 7, pp. 10086–10097, 2024.
- [6] S. Lu, Z. Gao, Q. Xu, C. Jiang, A. Zhang, and X. Wang, “Class-imbalance privacy-preserving federated learning for decentralized fault diagnosis with biometric authentication,” *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9101–9111, 2022.
- [7] Y. Xiao, H. Shao, J. Lin, Z. Huo, and B. Liu, “BCE-FL: a secure and privacy-preserving federated learning system for device fault diagnosis under non-IID condition in IIoT,” *IEEE Internet Things J.*, vol. 11, no. 8, pp. 14241–14252, 2024, doi: 10.1109/JIOT.2023.3340745.
- [8] M. Tawfik, A. A. Abu-Ein, H. M. Noaman, A. H. Abdelhaliem, and I. S. Fathi, “FedMedSecure: federated few-shot learning with cross-attention mechanisms and explainable AI for collaborative healthcare cybersecurity,” *Sci. Rep.*, vol. 15, art. 40050, 2025, doi: 10.1038/s41598-025-25107-z.
- [9] A. Gu and T. Dao, “Mamba: linear-time sequence modeling with selective state spaces,” *arXiv:2312.00752*, 2023.
- [10] H. Yi, D. Li, Z. Lu, Y. Jin, H. Duan, L. Hou, F. Z. Duraihem, E. Awwad, and N. A. Saeed, “VibrMamba: a lightweight Mamba-based fault diagnosis of rotating machinery using vibration signal,” *Measurement*, vol. 249, art. 116881, 2025, doi: 10.1016/j.measurement.2025.116881.
- [11] M. Li, L. Zhu, M. Luo, and T. Ke, “Remaining useful life prediction of airplane engine based on bidirectional Mamba and causal discovery,” *Sensors*, vol. 25, no. 11, art. 3429, 2025, doi: 10.3390/s25113429.
- [12] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, “KAN: Kolmogorov–Arnold networks,” in *Proc. ICLR*, 2025.
- [13] Z. Tang, X. Hou, X. Wang, and J. Zou, “A cross-working condition-bearing diagnosis method based on image fusion and a residual network incorporating the Kolmogorov–Arnold representation theorem,” *Appl. Sci.*, vol. 14, no. 16, art. 7254, 2024, doi: 10.3390/app14167254.
- [14] M. Tawfik, “MambaResp-KAN: a state space model with Kolmogorov–Arnold networks and diffusion-based augmentation for explainable respiratory disease classification,” *Int. J. Image Graph. Signal Process. (IJIGSP)*, vol. 18, no. 3, pp. 189–208, 2026, doi: 10.5815/ijigsp.2026.03.10.
- [15] M. Tawfik, I. S. Fathi, S. S. Nimbhore, I. M. Alsmadi, and M. S. Sawah, “E-RespiNet: an LLM-ELECTRA driven triple-stream CNN with feature fusion for asthma classification,” *PLoS One*, vol. 20, no. 11, art. e0334528, 2025, doi: 10.1371/journal.pone.0334528.
- [16] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, “Machine learning with adversaries: Byzantine tolerant gradient descent,” in *Proc. NeurIPS*, vol. 30, 2017, pp. 119–129.
- [17] D. Yin, Y. Chen, R. Kannan, and P. Bartlett, “Byzantine-robust distributed learning: towards optimal statistical rates,” in *Proc. ICML*, 2018, pp. 5650–5659.
- [18] E. M. El Mhamdi, R. Guerraoui, and S. Rouault, “The hidden vulnerability of distributed learning in Byzantium,” in *Proc. ICML*, 2018, pp. 3521–3530.
- [19] X. Cao, M. Fang, J. Liu, and N. Z. Gong, “FLTrust: Byzantine-robust federated learning via trust bootstrapping,” in *Proc. NDSS*, 2021.
- [20] M. Tawfik, N. M. Al-Zidi, B. Alsellami, A. M. Al-Hejri, and S. Nimbhore, “Internet of things-based middleware against cyber-attacks on smart homes using software-defined networking and deep learning,” in *Proc. 2nd Int. Conf. Comput. Methods Sci. Technol. (ICCMST)*, 2021, pp. 7–13, doi: 10.1109/ICCMST54943.2021.00014.
- [21] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, pp. 211–407, 2014.
- [22] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study,” *Mech. Syst. Signal Process.*, vol. 64–65, pp. 100–131, 2015.
- [23] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, “Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification,” in *Proc. PHM Soc. Eur. Conf.*, vol. 3, no. 1, 2016, doi: 10.36001/phme.2016.v3i1.1577.
- [24] H. Purohit, R. Tanabe, K. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, “MIMII dataset: sound dataset for malfunctioning industrial machine investigation and inspection,” in *Proc. DCASE Workshop*, 2019, pp. 209–213.
- [25] A. Saxena, K. Goebel, D. Simon, and N. Eklund, “Damage propagation modeling for aircraft engine run-to-failure simulation,” in *Proc. Int. Conf. PHM*, 2008, pp. 1–9.
- [26] C. Fung, C. J. M. Yoon, and I. Beschastnikh, “The limitations of federated learning in sybil settings,” in *Proc. RAID*, 2020, pp. 301–316.
- [27] M. Yurochkin, M. Agarwal, S. Ghosh, K. Greenewald, T. N. Hoang, and Y. Khazaeni, “Bayesian nonparametric federated learning of neural networks,” in *Proc. ICML*, 2019, pp. 7252–7261.
- [28] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, and Y. Khazaeni, “Federated learning with matched averaging,” in *Proc. ICLR*, 2020.

- [29] M. Fang, X. Cao, J. Jia, and N. Z. Gong, “Local model poisoning attacks to Byzantine-robust federated learning,” in Proc. USENIX Security, 2020, pp. 1605–1622.
- [30] V. Shejwalkar and A. Houmansadr, “Manipulating the Byzantine: optimizing model poisoning attacks and defenses for federated learning,” in Proc. NDSS, 2021.
- [31] G.-B. Truong, T.-T. Tran, N.-L. Than, V. Q. Nguyen, T. H. Nguyen, and V.-T. Pham, “SC-MambaFew: few-shot learning based on Mamba and selective spatial-channel attention for bearing fault diagnosis,” Comput. Electr. Eng., vol. 123, art. 110004, 2025.
- [32] Q. Sun, L. Lu, Y. Chen, L. Xiang, and A. Hu, “FedMAPS: a federated meta-learning framework with adaptive cross-domain contrastive learning for few-shot fault diagnosis,” Mech. Syst. Signal Process., 2026.
- [33] J. Sun, F. Zhou, J. Chen, C. Wang, X. Hu, and T. Wang, “A federated adversarial fault diagnosis method driven by fault information discrepancy,” Entropy, vol. 26, no. 9, art. 718, 2024, doi: 10.3390/e26090718.
- [34] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, “Deep learning with differential privacy,” in Proc. ACM CCS, 2016, pp. 308–318.
- [35] C. de Boor, A Practical Guide to Splines, rev. ed. New York: Springer-Verlag, 2001.
- [36] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, “How to backdoor federated learning,” in Proc. AISTATS, 2020, pp. 2938–2948.
- [37] Divya and D. Suresha, “KanAVNet: a CNN-BiLSTM-CTC-based audio-visual speech recognition system for Kannada to assist the hearing impaired,” Int. J. Image Graph. Signal Process. (IJIGSP), vol. 18, no. 2, pp. 87–106, 2026, doi: 10.5815/ijigsp.2026.02.06.
- [38] A. Paszke, S. Gross, F. Massa, et al., “PyTorch: an imperative style, high-performance deep learning library,” in Proc. NeurIPS, vol. 32, 2019, pp. 8024–8035.
- [39] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” in Proc. ICLR, 2015.
- [40] F. Wilcoxon, “Individual comparisons by ranking methods,” Biometrics Bull., vol. 1, no. 6, pp. 80–83, 1945.
- [41] I. S. Fathi and M. Tawfik, “Enhancing IoT systems with bio-inspired intelligence in fog computing environments,” Stat. Optim. Inf. Comput., vol. 13, no. 5, pp. 1916–1932, 2025, doi: 10.19139/soic-2310-5070-2305.
- [42] M. Tawfik, A. H. Abdelhaleem, and I. S. Fathi, “Transforming IoT security through large language models: a comprehensive systematic review and future directions,” Stat. Optim. Inf. Comput., vol. 14, no. 2, pp. 1018–1044, 2025, doi: 10.19139/soic-2310-5070-2424.

Authors’ Profiles



Bajes Zeyad Aljunaedi received the B.Sc. and M.Sc. degrees in science, engineering and technology from Saint Petersburg Electrotechnical University, Russia, in 2005 and 2007, respectively, and the Ph.D. degree in computer science from Saint Petersburg Electrotechnical University in 2010. He is currently the Acting Dean of the Faculty of Information Technology, Ajloun National University, Ajloun, Jordan. His research interests include cloud computing, image compression, big data, and the Internet of Things.



Mouiad Fadeil Alawneh received the B.Sc. degree in computer engineering from Philadelphia University, Jordan, in 2003, the M.Sc. degree in microelectronics (system engineering) from the Faculty of Engineering, National University of Malaysia (UKM), in 2005, and the Ph.D. degree in data science and artificial intelligence from the Faculty of Information Science and Technology, UKM, in 2013. He is currently the Head of Department at the Faculty of Information Technology, Ajloun National University, Ajloun, Jordan; he previously served as an Assistant Professor at Al-Madinah International University (2020–2023) and as a research fellow and academic staff member at the International Islamic University Malaysia (2014–2019). His research interests include natural language processing and machine translation, artificial intelligence, data science, and machine learning.



Mohammed Tawfik received the Ph.D. degree in information technology, specializing in artificial intelligence and cybersecurity, from BAMU University, India, in 2024. He is currently an Assistant Professor of Cybersecurity and Cloud Computing with the Department of Cyber Security, Faculty of Information Technology, Ajloun National University, Ajloun, Jordan. He has authored over 50 peer-reviewed papers published in Q1/Q2 SCI- and Scopus-indexed journals. His research interests include federated learning, explainable artificial intelligence, large language models (LLMs), IoT and IoMT security, state space models, speech recognition, medical imaging, and deep learning for healthcare signal processing. Dr. Tawfik has completed approximately 400 peer reviews for 40 top-tier journals and grant programs, including IEEE Access, Scientific Reports, Expert Systems with Applications, Knowledge-Based Systems, Engineering Applications of Artificial Intelligence, Internet of Things, and Journal of Information Security and Applications.

How to cite this paper:

Bajes Zeyad Aljunaedi, Mouiad Fadeil Alawneh, Mohammed Tawfik, "ByzaSpline-Fed: Gram-Weighted Functional Aggregation for Byzantine-Robust Federated Kolmogorov–Arnold Networks in Industrial IoT", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.18, No.5, pp. 33-54, 2026. DOI:10.5815/ijigsp.2026.05.03