

Manifold-Aligned In-Context Learning: Hybrid Retrieval-Augmented Tabular Foundation Models for Ambient Intelligence

Baobing Zhang, Sehrish Rafique, Mohamad Reza Shahabian Alashti, Shadiya Alingal Meethal, Vignesh Velmurugan, Patrick Holthaus, Gabriella Lakatos, Angela Dickinson and Farshid Amirabdollahian

Abstract—Indoor human localisation is a critical measure in smart homes and service robotics, yet ambient-sensor streams are characterised by long-horizon non-stationary distribution shifts, asynchronous sampling, and inherent sparsity. Under such conditions, tabular foundation model inference becomes highly sensitive to context quality and prone to instability under strict deployment budgets. This paper investigates a real-world multi-sensor environment and proposes a hybrid retrieval-augmented tabular foundation inference framework, framed through the lens of Bayesian Meta-Learning. Under a fixed context budget, we perform structured context optimisation to fortify the model's context-driven inference mechanism. Specifically, we introduce a dual-stream temporal-anchor-semantic-retrieval approach: the temporal anchor retains recent sequences to capture local dynamics. At the same time, semantic retrieval performs stochastic neighbour search within a differentiable manifold learned by Neighborhood Components Analysis (NCA) to recall long-tail sparse events. Experiments demonstrate that the proposed framework consistently outperforms state-of-the-art baseline models. Ablation studies further reveal that naive Euclidean retrieval in the original feature space leads to catastrophic degradation, underscoring the necessity of task-specific metric learning for retrieval-augmented in-context learning. Overall, the framework enables practical indoor localisation with superior robustness under noisy and incomplete sensing conditions. Code and data are publicly available at <https://github.com/baobingzhang/Retrieval-Augmented-In-Context-Learning>.

Index Terms—Internet of Things, Indoor localisation, Soft sensing, Ambient sensors, Foundation models, In-context learning, Metric Learning, Smart Homes.

I. INTRODUCTION

Indoor human localisation is a fundamental technical cornerstone for Ambient Assisted Living (AAL) and smart healthcare, underpinning critical applications ranging from fall detection to long-term behavioural profiling [1]. In these scenarios, distributed ambient sensors transform smart homes from simple automation platforms into continuous and precise measurement systems [2]. Specifically for home-based rehabilitation and elderly care, such systems are required to provide interpretable and auditable monitoring of occupant spatial distribution, room transitions, and abnormal events to inform timely risk alerts and care interventions [2], [3].

However, large-scale real-world deployment must satisfy strict privacy acceptability and cost constraints, which often

limits the practical adoption of high-bandwidth modalities such as cameras, mmWave radar [4], and Wi-Fi CSI [5]. Consequently, achieving reliable indoor localisation using low-cost, privacy-preserving ambient sensor networks (e.g., PIR motion detectors and contact sensors) has emerged as a foundational challenge at the intersection of IoT and healthcare [3].

Unlike high-information sensing, ambient sensor streams exhibit severe *measurement-chain challenges*: observations are frequently asynchronous and sparse, while signals are compounded by noise and missing data points. Moreover, inherent factors such as device drift and behavioural diversity induce significant *long-term non-stationary distribution shifts*. These factors lead to high intra-state variability and inter-state observability overlap, causing traditional discriminative models (e.g., random forests and LSTMs) to struggle with complex contextual dependencies and cross-condition generalization [3]. From an instrumentation and measurement perspective, the objective is not merely high average accuracy; rather, the inference must remain stable and predictable under conditions of data drift and sensing incompleteness.

Recent advances in tabular foundation models [6], [7] offer a promising paradigm shift. Through *In-Context Learning (ICL)*, these models can guide inference using a context set that encodes task structures, which is particularly attractive under limited data and distribution shifts [8]. Nevertheless, bringing this paradigm to long-term IoT streams faces a critical obstacle: tabular foundation inference is constrained by a *fixed context-length limit*, and uncertainty remains in selecting representative context samples. Naive use of full history or simplistic sampling can inject noise and class-distribution bias, leading to unstable performance [9]. Furthermore, we observe that simple Euclidean retrieval in the original feature space leads to *catastrophic metric collapse*—where physically distinct states become indistinguishable in the high-dimensional sparse space, failing to recall critical long-tail events.

Furthermore, real-world sensor networks often operate at highly irregular time steps, creating a severe “representation gap” between the foundation model's pre-trained priors and the actual topological structure of real-time sensor streams. Since the weights of foundation models are fixed during inference, relying solely on temporal sliding windows often overlooks deep semantic correlations masked by temporal noise. This alignment failure, compounded by the discreteness of sensor arrays, further distorts the metric space, rendering retrieval mechanisms based on raw feature similarity fragile when facing “concept drift” caused by environmental variations.

Motivated by these challenges and inspired by the Robot House multi-sensor instrumentation environment [10], we

The authors are with the *Robotics Research Group, University of Hertfordshire, Hatfield, United Kingdom* (e-mail: {b.zhang6, s.rafiq, m.r.shahabian, s.alingal-meethal4, v.velmurugan, p.holthaus, g.lakatos, a.m.dickinson, f.amirabdollahian2}@herts.ac.uk).

Copyright (c) 2026 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

propose a *Manifold-Aligned In-Context Learning (MA-ICL)* framework. We prioritise in-context manifold alignment as the primary design objective and introduce a dual-stream retrieval strategy: a *Temporal Anchor* to capture local Markovian dynamics and mitigate short-term drift, and a *Semantic Spark* to retrieve globally relevant evidence. Crucially, the semantic retrieval is performed within a *differentiable manifold* learned via Neighborhood Components Analysis (NCA), which aligns temporally dispersed but semantically similar activities. This framework improves the foundation model's context-conditioned inference by injecting high-fidelity, distribution-aware exemplars into its in-context prompt. The main contributions of this paper are as follows:

- 1) We propose *Manifold-Aligned In-Context Learning*, a novel inference paradigm that treats context construction as a retrieval optimisation problem for a frozen Tabular Foundation Model, which itself approximates the Bayesian posterior predictive distribution, on non-stationary ambient sensor streams, without parameter fine-tuning.
- 2) We identify and resolve the “*Metric Collapse*” phenomenon in high-dimensional sparse sensor data by introducing a *Dual-Stream Context Retrieval* strategy that synergizes *Temporal Anchors* with *Semantic Sparks* driven by *Neighborhood Components Analysis*, ensuring the retrieval of semantically coherent evidence from a learned differentiable manifold.
- 3) We provide extensive empirical evidence on the real-world *Robot House* smart-home platform, demonstrating robustness against non-stationary distribution shifts and long-tail events while establishing a *Pareto-optimal* balance between accuracy and context budget compared to modern deep-learning baselines; generalization under continuous-stream distribution shift is further characterised in the supplementary material.
- 4) From an instrumentation perspective, we systematically quantify the trade-offs between context budget and inference accuracy. By revealing the performance response under varying computational constraints, we provide empirical guidelines for applying tabular foundation models in resource-constrained IoT scenarios without the need for costly online retraining.

The rest of this paper is organised as follows. Section II reviews related work, providing the foundations of ambient sensing and tabular foundation models, and compares our approach to earlier methods. Section III details the proposed method, including the Bayesian context-aware inference formulation, NCA-based manifold learning, and the dual-stream retrieval strategy. Section IV describes the experimental setup, implementation, data analysis. Section V provides the experimental results and discussion. Section VI concludes the paper and outlines future directions.

II. RELATED WORK

In recent years, device-free indoor presence sensing and localisation technologies have received increasing attention in smart home and healthcare scenarios. Studies have shown

that WiFi channel state information (CSI) and received signal strength (RSSI) can enable non-intrusive presence detection and activity recognition without requiring wearable devices [11], [12]. Passive infrared (PIR) sensors are widely used for occupancy detection due to their low cost and privacy advantages [13], while ultra-wideband (UWB) radar provides sub-meter localisation accuracy suitable for fine-grained tracking [14]. Wearable sensors such as IMUs remain useful for fall detection and behaviour analysis in clinical monitoring [15]. At the higher-fidelity end of the sensing spectrum, LiDAR-based geometric pipelines attain precise localisation and mapping in complex environments through active scanning control [16] and large-scale bundle-adjustment optimisation [17]; these approaches, however, rely on dense point-cloud geometry and higher-cost, power-intensive sensors, and are therefore complementary to, rather than a substitute for, the low-cost non-visual ambient sensors and learning-driven inference targeted in this work. Deep neural networks, including CNNs and LSTMs, have been extensively applied to improve sensing accuracy and robustness [18]. Despite these advancements, existing methodologies often struggle to maintain robust performance when faced with the long-horizon non-stationary distribution shifts and inherently sparse event patterns characteristic of real-world ambient sensor streams. This paper addresses these challenges by proposing a hybrid retrieval-augmented tabular foundation model framework that leverages manifold-aligned in-context learning to synergistically capture both local temporal dynamics and long-tail semantic evidence.

Beyond sensing capabilities, recent studies have focused on adapting presence detection, localisation, and fall recognition to complex real-world environments. For instance, Ngamakeur et al [19] proposed an indoor localisation framework based on PIR signals combined with a CNN-LSTM model. Alteneiji et al [11] used deep learning to extract presence states from WiFi reflections, effectively mitigating occlusion and multipath interference. For fall detection, Hu et al [12] introduced DeFall, a system that detects fall events through anomalies in WiFi CSI, enabling deployment without user instrumentation or prior environment modelling. Overall, multimodal fusion and deep models are emerging as key enablers of robust, efficient, and deployable indoor presence measurement systems [18], [20].

A. Modelling Sparse and Irregular Event Streams

In real-world multi-sensor environments, heterogeneous sensors exhibit asynchronous sampling rates, making unified modelling a challenge. To address this, continuous-time models such as Neural Ordinary Differential Equations (ODEs) and Controlled Differential Equations (CDEs) have been introduced to handle irregular timestamps, demonstrating robustness and expressiveness in healthcare and IoT contexts [21], [22]. Moreover, neural temporal point processes (Neural TPPs) have been widely applied to model the dynamics of sparse events [23]. For cross-modal alignment, attention-based approaches such as MAESTRO [24] align asynchronous modalities without relying on a primary modality.

On the other hand, real sensor streams often suffer from missing values and noise, impairing modelling accuracy

and deployment stability. The CSDI model [25] leverages diffusion-based generation to produce high-quality time-series imputations, while Wang et al. [26] conducted a systematic evaluation of deep imputation methods' impact on prediction reliability. Additionally, Transformer-based architectures have shown strong robustness under modality drop-out [27], with models like CARD [28] using channel alignment to stabilize sequence modelling, supporting reliable deployment over long asynchronous multi-sensor streams.

B. Tabular Foundation Models and In-Context Learning

Tabular foundation models have gradually evolved from task-specific deep architectures into more unified modelling frameworks with certain pretraining and generalization capabilities. Representative works include TabNet [29], which improves interpretability and small-sample performance via sequential feature-selection attention, TabTransformer [30], which models categorical feature dependencies using Transformer-based contextual embeddings, and SAINT [31], which introduces row-column dual attention and contrastive pretraining. By explicitly modelling inter-feature context, these methods significantly narrow the performance gap to GBDT-based approaches on multiple benchmarks. Furthermore, TabPFN [6], [7] brings large-scale synthetic task meta-learning and in-context learning into the tabular domain, enabling the model to solve new tasks via a single forward pass without any parameter updates, and to systematically outperform classical methods on small and medium-sized datasets, marking the emergence of a true “foundation model” paradigm for tabular learning.

Along the line of few-shot and in-context learning, one class of methods reformulates tabular tasks as language modelling problems to exploit the contextual reasoning ability of large language models, such as TabLLM [8], which serializes table rows into natural language prompts for zero-shot and few-shot classification. Another class of approaches improves small-sample generalization through task generation and meta-learning, for example STUNT [32], which constructs pseudo-tasks from unlabeled tables for training. Meanwhile, tabular foundation models have started to be deployed in real-world sensor and industrial scenarios, including TabNet for traffic incident duration prediction [33], TabTransformer for IoT intrusion detection [34], and TabPFN for medical radiomics tasks with demonstrated robustness under extremely small sample sizes [35]. However, existing studies also clearly indicate that real sensor data typically exhibit severe distribution shift, extreme data scarcity, and missing semantic information, leading to a substantial gap between small-data advantages on benchmark datasets and robust deployment in real-world environments. This has recently motivated the introduction of semantic augmentation and knowledge distillation strategies, such as FeatLLM [36] and Latte [37], to mitigate this gap. Crucially, these foundation models are designed for static, independently and identically distributed (i.i.d.) tables. TabPFN [6], [7] consumes the entire (small) training set as an order-agnostic context in a single forward pass, and its only prescribed remedy once a dataset exceeds its bounded

context is uniform random subsampling. Neither assumption holds for a continuous, non-stationary sensor stream, whose history is unbounded and where uniform subsampling discards precisely the recent drift and long-tail events that determine the current label; indeed, the same line of work subsequently introduced a separate Drift-Resilient TabPFN [38] to cope with temporal distribution shift, confirming that vanilla in-context inference is not drift-aware. A parallel effort in language modelling—in-context example (demonstration) selection, such as KATE [39] and EPR [40]—retrieves relevant exemplars to populate the prompt, but it operates on text, uses a single similarity stream, and models neither temporal non-stationarity nor a fixed streaming budget. Our work departs from all of these by keeping the foundation model entirely frozen and instead treating context construction itself as an inference-time, budget-constrained selection problem for non-stationary tabular streams.

C. Metric learning and representation-driven context retrieval

In recent years, research on context-aware retrieval can be summarized as learning a retrieval-friendly semantic embedding space via metric learning and representation learning, such that semantically similar items are pulled closer while dissimilar items are pushed further apart in the representation space. In supervised metric learning, representative studies focus on designing more effective similarity constraints and exploiting training signals more efficiently: for example, jointly weighting multiple similarity pairs to improve the utilization of informative pairs [41], replacing explicit triplet sampling with a classification-style objective using multiple centers while preserving intra-class diversity [42], and adopting a proxy-based objective to balance fast convergence with modelling data-to-data relations [43]. Collectively, these lines of work improve retrieval embeddings in terms of discriminability, stability, and trainability [41]–[43].

On the representation-learning side, self-supervised contrastive learning methods (e.g., SimCLR [44] and MoCo [45]) use instance-level consistency across different augmented views as supervision, providing transferable general-purpose encoders that substantially reduce annotation dependence and improve robustness in open-world retrieval settings [44], [45]. For multimodal context-aware retrieval, UNITER [46] and CLIP [47] learn a unified cross-modal semantic space through large-scale image-text alignment pretraining, while LightningDOT [48] further targets practical deployment by optimising real-time dot-product retrieval for low-latency systems. Meanwhile, “context” often manifests as structured relations (e.g., session sequences, global transitions, and multi-hop knowledge-graph relations), making graph modelling an important way to enhance retrieval: GCE-GNN [49] fuses short-term and long-term context by combining session graphs with global transition graphs, and GNN-RAG [50] performs graph neural retrieval to identify query-relevant subgraphs in knowledge graphs, providing highly relevant evidence context for downstream reasoning (including integration with large language models). Closest to our retrieval mechanism, TabR [51] augments a tabular network with a k-nearest-

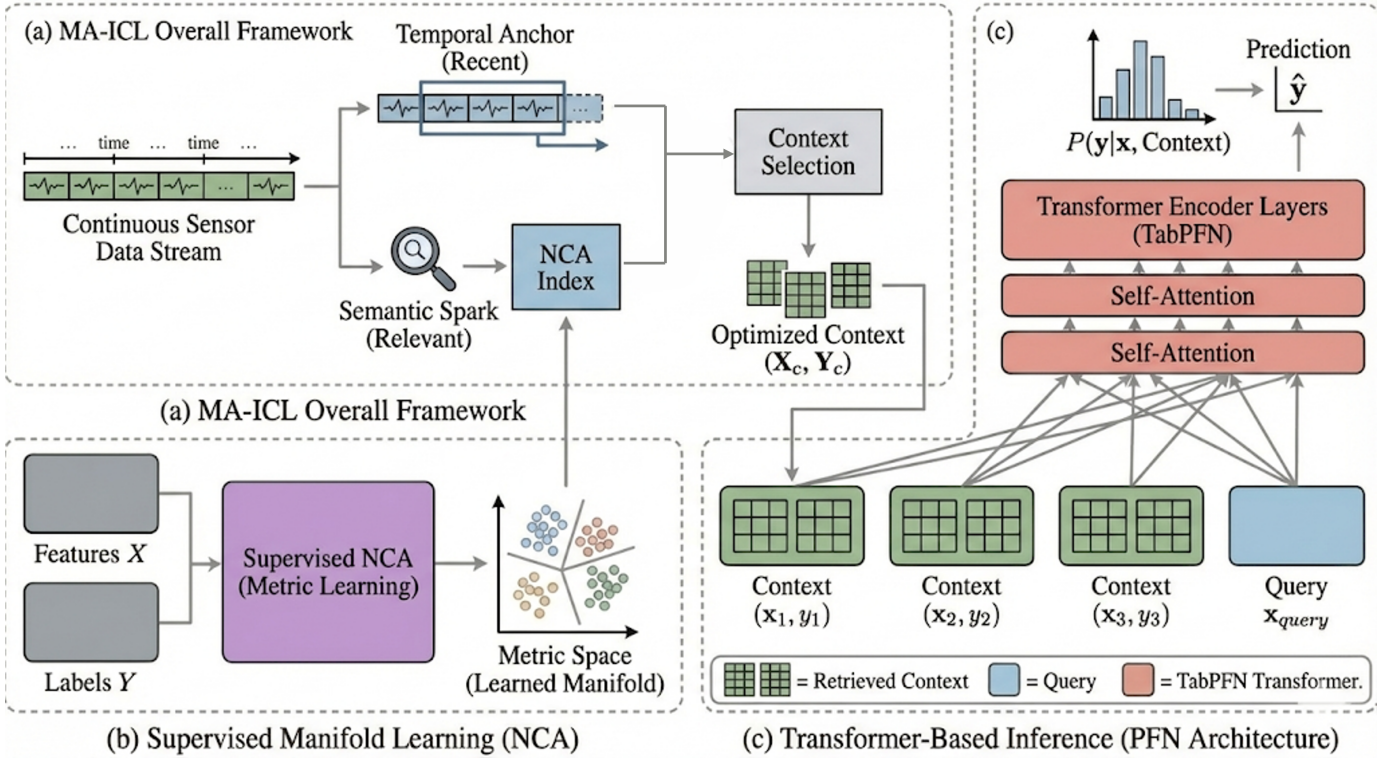


Fig. 1. **Overview of the proposed Manifold-Aligned In-Context Learning (MA-ICL) framework.** (a) **Dual-Stream Inference:** The system combines a **Temporal Anchor** (for stability) and **Semantic Spark** (via NCA retrieval) to generate the optimised context set $(\mathbf{X}_C, \mathbf{Y}_C)$. (b) **Supervised NCA:** Projects features into a learned metric space by maximising intra-class stochastic neighbor probabilities to build the retrieval index. (c) **Transformer-Based Inference:** The context sequence $(\mathbf{x}_i, \mathbf{y}_i)$ and query $\mathbf{x}_{query}$ are fed into the PFN. The architecture leverages **In-Context Learning (ICL)** to directly approximate the posterior predictive distribution $P(y|\mathbf{x}, \text{Context})$.

neighbour module; however, its similarity metric is a differentiable layer trained end-to-end with the predictor over a static candidate pool under the i.i.d. setting, and the model is trained from scratch per dataset rather than kept frozen. Similarly, Neighbourhood Components Analysis [52] is conventionally employed as a metric for k-NN classification, where the learned metric is itself the decision rule. In contrast, we decouple metric learning from prediction: NCA is used solely to shape a retrieval-index geometry over which exemplars are selected and then handed, unmodified, to a frozen foundation model — a training-free, model-agnostic curation policy rather than an internal differentiable component. This distinction is not cosmetic: performing the same dual-stream retrieval in the raw Euclidean feature space collapses accuracy to 2.86% (Section V), so a learned selection geometry is a necessary, load-bearing part of the design.

III. METHODOLOGY

This section introduces our novel approach to manifold-aligned in-context learning, first formally defining Bayesian context-aware inference, followed by outlining our solution to NCA-based manifold learning, and finally explaining our novel dual-stream retrieval strategy.

A. Problem Formulation: Bayesian Context-Aware Inference

We consider the task of continuous ambient intelligence as a sequential posterior estimation problem. Let $\mathcal{X} \subset \mathbb{R}^d$ denote

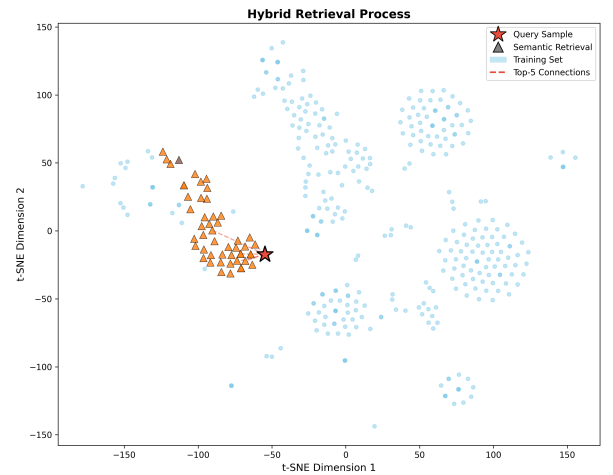


Fig. 2. **Visualisation of the Semantic Retrieval Process in the Learned Manifold.** The high-dimensional sensor features are projected into 2D space using t-SNE. The **Red Star** represents the current query sample $\mathbf{x}_t$ (from the test stream), while the **Triangles** denote the retrieved historical context samples $\mathcal{C}_{sem}$ via the NCA metric. The dashed lines highlight the top-5 nearest neighbors. This visualisation demonstrates that the learned manifold effectively clusters semantically similar activities, enabling the robust retrieval of relevant context even for sparse events.

the sensor feature space and $\mathcal{Y} = \{1, \dots, C\}$ be the discrete set of labels. At time t , the observation vector is $\mathbf{x}_t \in \mathcal{X}$. To capture temporal Markovian dependencies, we construct a

sliding window feature vector $\mathbf{X}_t = \Psi(\mathbf{x}_{t-w+1:t}) \in \mathbb{R}^{d \times w}$.

Unlike traditional discriminative models that learn a static mapping $f_\theta : \mathcal{X} \mapsto \mathcal{Y}$, we interpret the task through the lens of *Bayesian meta-learning*, casting inference as context selection for a meta-learned posterior approximator, to handle distribution shifts in non-stationary environments.

Definition 1 (Posterior Predictive Distribution). Given a query $\mathbf{X}_t$ and a historical context set $\mathcal{D}_{hist} = \{(\mathbf{X}_i, y_i)\}_{i=1}^N$, our goal is to compute the posterior predictive distribution (PPD):

$$P(y_t | \mathbf{X}_t, \mathcal{D}_{hist}) = \int P(y_t | \mathbf{X}_t, \phi) P(\phi | \mathcal{D}_{hist}) d\phi \quad (1)$$

where ϕ represents latent task-specific generative parameters.

The Tabular Foundation Model, specifically the Prior-Data Fitted Network (PFN) [7] approximates this intractable integral via a Transformer-based architecture $\text{PFN}(\cdot)$. However, due to the quadratic complexity of attention, the model is constrained to a limited context size $K \ll N$ (e.g., $K \approx 2000$). Thus, the optimisation objective transforms into a *Variational Context Selection* problem: finding the optimal subset $\mathcal{C}^*$ that minimizes the Kullback-Leibler (KL) divergence between the true posterior and the approximation:

$$\mathcal{C}^* = \arg \min_{\mathcal{C} \subset \mathcal{D}_{hist}, |\mathcal{C}| \leq K} D_{KL} \left(P(y | \mathbf{X}_t, \mathcal{D}_{hist}) \parallel P_{\text{PFN}}(y | \mathbf{X}_t, \mathcal{C}) \right) \quad (2)$$

We emphasise that Eq. (2) serves as a guiding objective; our dual-stream strategy (Section III-C) is a tractable heuristic that approximates this selection rather than an exact KL-minimisation solver. Different from conventional schemes requiring explicit posterior updates, our approach approximates Bayesian inference implicitly through meta-learned priors within the PFN framework.

B. Task-Specific Manifold Learning via NCA

To select the optimal $\mathcal{C}^*$ that minimizes the divergence in Eq. (2), a metric reflecting semantic similarity is essential. Euclidean distance often degenerates in high-dimensional sparse sensor spaces. We employ Neighborhood Components Analysis (NCA) to learn a low-dimensional manifold embedding.

Let $\mathbf{A} \in \mathbb{R}^{d' \times d}$ be the linear transformation matrix to be learned. It induces a Mahalanobis-like distance metric in the projected space:

$$d_{\mathbf{A}}(\mathbf{X}_i, \mathbf{X}_j) = \|\mathbf{A}\mathbf{X}_i - \mathbf{A}\mathbf{X}_j\|^2 \quad (3)$$

Proposition 1 (Stochastic Neighbor Probability). We define the probability that point i selects point j as its stochastic neighbor using a Softmax kernel:

$$p_{ij} = \frac{\exp(-d_{\mathbf{A}}(\mathbf{X}_i, \mathbf{X}_j))}{\sum_{k \neq i} \exp(-d_{\mathbf{A}}(\mathbf{X}_i, \mathbf{X}_k))}, \quad p_{ii} = 0 \quad (4)$$

This probability p_{ij} quantifies the semantic relevance of sample j to sample i .

optimisation Objective. We maximise the expected Leave-One-Out (LOO) accuracy over the training set, which is equivalent to maximising the log-likelihood of correctly classifying each point based on its neighbors:

$$\mathcal{L}(\mathbf{A}) = \sum_{i=1}^N \log \left(\sum_{j \in \Omega_i} p_{ij} \right), \quad \Omega_i = \{j \mid y_j = y_i\} \quad (5)$$

Gradient Derivation. To optimise $\mathbf{A}$ via gradient ascent, we derive the closed-form gradient of the objective function:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{A}} = 2\mathbf{A} \sum_i \left(p_i \sum_k p_{ik} \mathbf{v}_{ik} \mathbf{v}_{ik}^\top - \sum_{j \in \Omega_i} p_{ij} \mathbf{v}_{ij} \mathbf{v}_{ij}^\top \right) \quad (6)$$

where $p_i = \sum_{j \in \Omega_i} p_{ij}$ is the classification probability, and $\mathbf{v}_{ik} = \mathbf{X}_i - \mathbf{X}_k$. Maximising this objective forces $\mathbf{A}$ to collapse intra-class samples into compact clusters, ensuring the semantic purity of the retrieved context.

C. Hybrid Retrieval Strategy: A Bias-Variance Perspective

We propose a dual-stream retrieval strategy that balances the bias-variance trade-off in non-stationary streams.

1) Theoretical Justification:

Proposition 2 (Bias-Variance Decomposition). The generalisation error of the estimator is bounded by:

$$\mathcal{E}_{\text{error}} \approx \alpha \cdot \underbrace{\text{Bias}(\mathcal{C}_{anc})}_{\text{Distribution Shift}} + \beta \cdot \underbrace{\text{Var}(\mathcal{C}_{sem})}_{\text{Sample Sparsity}} \quad (7)$$

Relying solely on recent history ($\mathcal{C}_{anc}$) minimises bias from concept drift but suffers high variance for rare events. Relying solely on semantic retrieval ($\mathcal{C}_{sem}$) reduces variance via global support but introduces bias from outdated distributions. Our hybrid approach optimally combines both.

2) *Mechanism Design:* **1. Temporal Stream (Anchor):** $\mathcal{C}_{anc} = \{(\mathbf{X}_\tau, y_\tau)\}_{\tau=t-K_{anc}}^{t-1}$. This preserves Markovian continuity and prevents mode collapse.

2. Semantic Stream (Prototype Spark): Instead of point-wise retrieval, we compute the *Batch Centroid* $\mu_B = \frac{1}{B} \sum \mathbf{A}\mathbf{X}_b$ in the learned manifold. $\mathcal{C}_{sem}$ is formed by the k -NNs of μ_B . This prototype-based approach implicitly denoises the query. This choice rests on three grounds. **De-noising:** writing each query embedding as $\mathbf{z}_i = \mathbf{z}^* + \epsilon_i$ (a shared latent state plus independent noise of variance σ^2), the prototype has variance σ^2/B , a B -fold reduction that gives a more stable retrieval anchor. **Temporal coherence:** a batch is B consecutive stream windows that, under the Markovian continuity exploited by the temporal anchor, almost always share the same location state, so a common context is an appropriate approximation; if a batch does straddle distinct states, the temporal anchor still supplies each sample's recent evidence. **Computational cost:** since the batch shares one context, the dominant $O(K^2)$ self-attention that encodes the K -exemplar context is computed once per batch rather than once per sample, and the k -NN query is issued once instead of B times, giving a per-sample-to-batch cost ratio of $B(K+1)^2/(K+B)^2 \approx B$ for $K \gg B$ (about $48 \times$ at $K=2048, B=50$). The batch size B is a tunable parameter,

Algorithm 1 Task-Specific Metric Learning

Require: Training set $\mathcal{D}_{hist}$, Learning rate η , Epochs E

Ensure: Transformation matrix $\mathbf{A}^*$

- 1: Initialize $\mathbf{A}$ randomly
 - 2: **for** epoch = 1 to E **do**
 - 3: Compute pairwise distances matrix D_{ij} via Eq. 3
 - 4: Compute probability matrix P_{ij} via Eq. 4
 - 5: Compute gradient $\nabla_{\mathbf{A}} \mathcal{L}$ via Eq. 6
 - 6: Update $\mathbf{A} \leftarrow \mathbf{A} + \eta \nabla_{\mathbf{A}} \mathcal{L}$
 - 7: **end for**
 - 8: Index $\mathcal{Z}_{hist} \leftarrow \mathbf{A} \mathcal{D}_{hist}$
 - 9: **return** $\mathbf{A}^*, \mathcal{Z}_{hist}$
-

Algorithm 2 Hybrid Retrieval & Inference

Require: Stream Buffer $\mathbf{B}_t$, History $\mathcal{D}_{hist}$, Matrix $\mathbf{A}^*$, Budget K

Ensure: Batch Predictions $\hat{\mathbf{Y}}_t$

- 1: **Step 1: Temporal Anchoring**
 - 2: $K_{anc} \leftarrow \lfloor \lambda K \rfloor$
 - 3: $\mathcal{C}_{anc} \leftarrow \text{Last}(\mathcal{D}_{hist}, K_{anc})$
 - 4: **Step 2: Prototype Semantic Retrieval**
 - 5: $\mathbf{Z}_{batch} \leftarrow \mathbf{B}_t \cdot (\mathbf{A}^*)^\top$
 - 6: $\boldsymbol{\mu} \leftarrow \text{Mean}(\mathbf{Z}_{batch}, \text{axis} = 0)$ ▷ Compute Prototype
 - 7: $\mathcal{I}_{sem} \leftarrow \text{KNN}(\mathcal{Z}_{hist}, \boldsymbol{\mu}, k = K - K_{anc})$
 - 8: $\mathcal{C}_{sem} \leftarrow \{(\mathbf{X}_j, y_j) \mid j \in \mathcal{I}_{sem}\}$
 - 9: **Step 3: Fusion & Prediction**
 - 10: $\mathcal{C}_{hybrid} \leftarrow \text{Sort}_{\text{time}}(\mathcal{C}_{anc} \cup \mathcal{C}_{sem})$
 - 11: $P(\mathbf{Y}) \leftarrow \text{PFN}(\mathbf{B}_t; \mathcal{C}_{hybrid})$ ▷ Forward Pass
 - 12: $\hat{\mathbf{Y}}_t \leftarrow \arg \max P(\mathbf{Y})$
 - 13: **return** $\hat{\mathbf{Y}}_t$
-

and $B=1$ recovers exact per-sample retrieval. To intuitively validate the efficacy of the learned manifold structure, we visualise the retrieval process in Fig. 2. As observed, although the query sample (indicated by the red star) is temporally detached from the training history, the NCA-induced metric space successfully aligns it within a compact semantic cluster. This geometric alignment confirms that our method establishes cross-temporal semantic connectivity within the latent space, thereby effectively compensating for the information deficit caused by the scarcity of samples in local windows.

3. Sequential Fusion: To respect the causal structure required by the Transformer's positional embeddings, we define the fusion operator:

$$\mathcal{C}_{hybrid} = \text{Sort}_{\text{time}}(\mathcal{C}_{anc} \cup \mathcal{C}_{sem}) \quad (8)$$

The metric learning and inference procedures are detailed in Algorithms 1 and 2. The complete flowchart is shown in Fig. 1.

IV. EXPERIMENTAL SETUP

This section details the experimental configuration designed to validate the proposed Manifold-Aligned In-Context Learning (MA-ICL) framework. The purpose of this section is to provide a comprehensive overview of the evaluation methodology in a real-world setting. Specifically, we describe

the experimental environment and data collection procedures within the Robot House testbed, followed by the technical implementation details of the system. Note that to ensure a clear distinction between methodology and evaluation, the empirical results and performance analysis are presented separately in Section V.

A. Experimental Environment: The Robot House

The evaluation of the proposed framework is conducted within the Robot House at the University of Hertfordshire, a fully-sensorized residential testbed dedicated to Ambient Assisted Living (AAL) research. As illustrated in the floor plan Fig 4 and Fig 5, the environment replicates a standard domestic living space including a Kitchen, Living Area, Bedroom, Bathroom, Sofa Area and Corridor, providing an ideal setting for authentic activity recognition. For more detailed information about Robot House, please refer to [10].

1) *Heterogeneous Ambient Sensing Network:* To address sensing uncertainty without invading privacy. Robot House consist of a non-intrusive network of over 60 heterogeneous sensors integrated via a ROS (Robot Operating System)-based infrastructure. The sensing matrix consists of:

- **Binary Event-based Sensors:** PIR motion detectors and magnetic contact sensors for monitoring doors and cabinets.
- **Continuous Stream-based Sensors:** Appliance-level power meters and ambient monitors (light, temperature).
- **Contextual Anchors:** Pressure mats embedded in sofas and beds to stabilize occupancy detection during static phases.

2) Dataset Curation and High-Precision Ground Truth:

Our evaluation utilizes sequential data collected under real-world conditions. A defining characteristic of this dataset is its high-fidelity ground truth. Room-level labels were manually annotated using synchronized video recordings as the reference oracle (solely for annotation purposes). This rigorous labelling process ensures high temporal precision, particularly for cross-room transitions. The finalised dataset is formalised into 8 logical spatial categories, reflecting the comprehensive range of domestic activities and room-to-room movements.

3) *Dataset Characteristics and Feature Visualisation:* The collected dataset comprises 8,067 records spanning 21 heterogeneous sensing dimensions with 1.6 Hz average state-update frequency. As synthesised in Fig. 3, the data exhibits significant multi-dimensional complexity. Subplot (1) illustrates the class imbalance, with the Sofa Area and Bedroom accounting for 68% of total samples. Subplot (2) highlights the temporal occupancy variance; while the Dining Area exists as a low-frequency event, it maintains the highest average stay duration of 18.0 minutes. The spatial relationships and hardware trigger frequencies are quantified in Subplots (3) and (4), respectively. Finally, Subplot (5) presents the activity duration distribution. Note that the Dining Area is neglected in Subplot (5) to ensure statistical significance; as a single-segment activity in this session, it lacks the variance required for quartile-based boxplot analysis. By filtering the visualisation to the top 8 frequent categories, we provide a clearer contrast between rapid inter-room transitions and stable sedentary states.

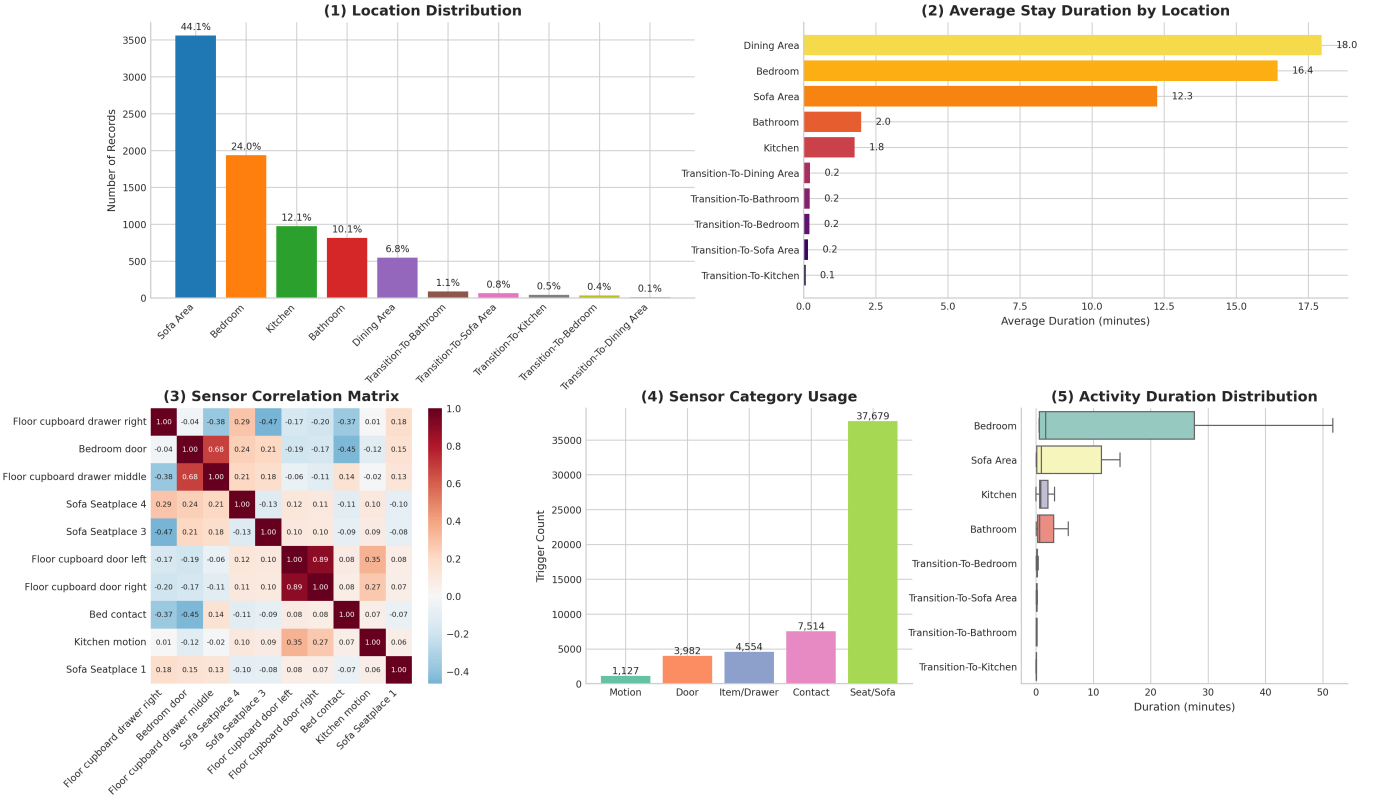


Fig. 3. Multi-dimensional feature visualisation of the Robot House sensor dataset, illustrating spatial distribution, occupancy duration, sensor correlations, and trigger frequency profiles.

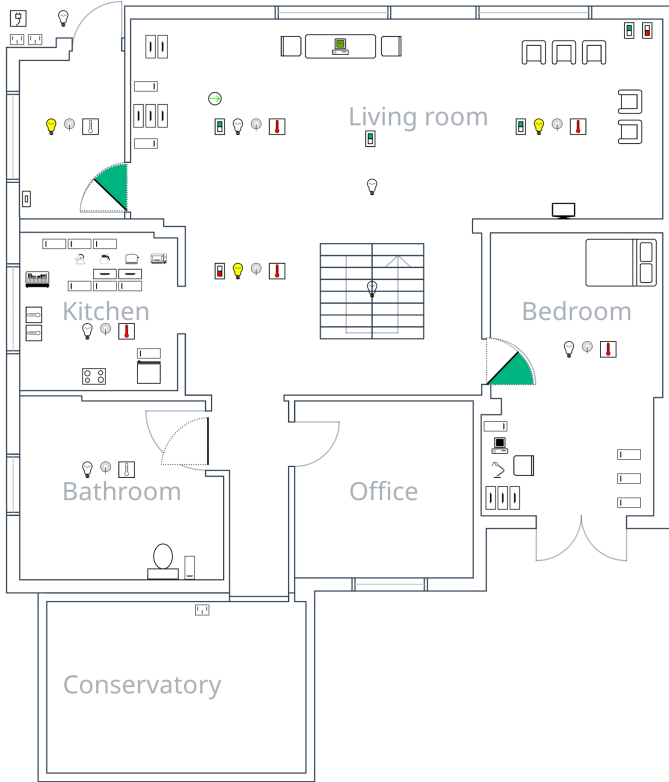


Fig. 4. Spatial layout of heterogeneous ambient sensors in the Robot House.

4) Preprocessing and Input formalisation: The sensory input is structured into a 76-dimensional feature vector. To capture local temporal dynamics and intermittent shifts, we implement a sliding window mechanism with a size of $W = 10$. All numerical features are normalised via Min-Max scaling to the range $[0, 1]$, mitigating the dominance of high-magnitude power readings and ensuring balanced influence across diverse sensing modalities.

B. Implementation Details

The proposed framework was implemented using PyTorch and Scikit-learn. In the metric learning phase, the NCA model was initialized with a target latent dimension of $d' = 16$, and the transformation matrix was optimised using the L-BFGS-B solver to maximise stochastic neighbor accuracy. This solver performs a Wolfe-condition line search rather than fixed-step gradient ascent, so there is no learning rate to tune or warm up and no risk of gradient explosion; the schematic update in Algorithm 1 denotes the objective gradient, not a fixed-step rule. The stochastic-neighbour softmax is evaluated over negative squared distances with the log-sum-exp stabilisation, so large distances underflow harmlessly to zero rather than overflowing, and the Min-Max normalisation of the inputs keeps all pairwise distances bounded. The training objective consequently converges smoothly and monotonically, as shown by the NCA convergence curve in the Supplementary Material.

For the PFN foundation model, the total context budget was fixed at $K = 2048$. In accordance with the hybrid strategy



Fig. 5. Main living spaces in the Robot House smart home environment: (top row, left to right) Living area, Kitchen, Sofa area; (bottom row, left to right) Bedroom, Bathroom, Corridor. Each scene is equipped with environmental sensors and a mobile robot for data collection and validation. This platform provides an ideal testbed for Ambient Intelligence research and supports understanding and research across multiple scenarios.

TABLE I

PERFORMANCE COMPARISON WITH MODERN TABULAR DEEP LEARNING METHODS ON THE INDOOR LOCALISATION TASK. BEST RESULTS ARE IN **BOLD**, SECOND BEST ARE UNDERLINED. Δ DENOTES THE IMPROVEMENT OVER THE BEST BASELINE.

Method	Venue	Type	Acc. (%)	F1-Mac. (%)	Prec.-W (%)	Bal. Acc. (%)	MCC (%)
TabTransformer [53]	arXiv'20	Transformer	80.88	33.69	82.58	50.01	44.80
SAINT [54]	NeurIPS'21	Attention	83.80	37.10	90.10	52.29	50.80
NODE [55]	ICLR'20	Tree-NN	83.67	33.33	81.09	49.44	44.40
DCN V2 [56]	WWW'21	Interaction	83.49	39.50	81.55	54.43	47.00
Trompt [57]	ICML'23	Prompt	86.34	42.03	80.63	59.52	56.70
TabR [51]	ICLR'24	Retrieval	83.30	36.44	79.06	51.35	43.40
ExcelFormer [58]	KDD'24	Transformer	<u>86.59</u>	<u>43.41</u>	<u>86.94</u>	<u>60.80</u>	61.30
Ours	—	Hybrid-Retrieval	87.09	48.23	88.03	61.67	<u>60.50</u>
Δ vs. Best Baseline			+0.50	+4.82	-2.07	+0.87	-0.80

hyperparameter $\lambda = 0.5$, this budget was allocated equally between the temporal anchor ($K_{anc} = 1024$) and the semantic stream ($K_{sem} = 1024$). This configuration balances the capture of local Markovian continuity with the capacity to retrieve global semantic prototypes. Unless otherwise specified, all experimental results presented in this work are derived from 10 independent runs to ensure statistical significance.

V. RESULTS AND DISCUSSION

This section provides a comprehensive evaluation of the proposed Manifold-Aligned In-Context Learning (MA-ICL) framework through a series of structured experiments. First, the comparative analyses against a wide range of state-of-the-art baselines demonstrate the framework's robustness and accuracy on the real-world Robot House instrumented residential environment. Following this, the ablation study dissects the synergistic effects of the dual-stream retrieval mechanism and manifold alignment, validating their critical role in resolving the metric collapse issue inherent in high-dimensional sparse sensor data. Finally, the stability and hyperparameter

sensitivity analysis systematically quantifies the trade-off between inference precision and computational budget, providing practical empirical guidelines for deployment in resource-constrained IoT scenarios.

A. Comparative Analysis with Baseline Models

Predicting spatial locations from high-dimensional sensor readings represents a quintessential challenge in tabular deep learning. Due to the inherent heterogeneity of tabular data and the absence of spatial inductive bias, traditional neural architectures often struggle to achieve robust generalization. As illustrated in Table I, standard deep learning models, such as *ResNet-1D*, exhibit significant performance degradation in capturing non-linear dependencies between heterogeneous features, leading to suboptimal F1-Macro scores. Within the traditional machine learning paradigm, we observe a distinct hierarchical performance distribution. High-capacity ensemble algorithms, such as *Random Forest* and *LightGBM*, maintain competitive baselines. However, several foundational algorithms demonstrate severe limitations. *AdaBoost* proves in-

TABLE II

PERFORMANCE EVALUATION AGAINST BASELINES. BEST RESULTS ARE IN **BOLD**, SECOND BEST ARE UNDERLINED. Δ DENOTES THE IMPROVEMENT OVER THE BEST BASELINE. GIVEN THE STRONG CLASS IMBALANCE OF THE TASK, THE ACCURACY MARGIN IS MODEST BY DESIGN, WHEREAS THE IMPROVEMENT IS MARKEDLY LARGER ON THE ROBUSTNESS-ORIENTED METRICS (F1-MACRO, MCC).

Category	Algorithm	Acc. (%)	F1-Macro (%)	MCC (%)
Deep Learning	BiLSTM	<u>85.79</u>	<u>46.46</u>	<u>55.46</u>
	GRU	84.36	39.60	51.11
	ResNet-1D	70.27	16.27	20.62
Traditional ML	Random Forest	84.17	37.47	47.75
	LightGBM	84.11	33.03	46.82
	XGBoost	83.80	33.87	44.76
	SVM	83.74	39.90	45.71
	AdaBoost	80.38	25.38	39.18
	GaussianNB	10.99	20.16	32.65
	KNN	10.86	25.44	18.10
Ours	Hybrid-Retrieval	87.09	48.23	60.50
Δ vs. Best Baseline		+1.30	+1.77	+5.04

sufficient in characterizing complex localisation boundaries due to its weak learners' preference for simplistic feature subsets. More strikingly, *GaussianNB* and *KNN* exhibit near-total failure. This is primarily because GaussianNB imposes an unrealistic assumption of feature independence, contradicting the highly correlated nature of RSSI (Received Signal Strength Indicator) values in sensor arrays. Meanwhile, *KNN* suffers from the "curse of dimensionality" and susceptibility to noise in the unweighted high-dimensional space. By contrast, our proposed Hybrid-Retrieval method attains the best overall accuracy of 87.09% together with a leading F1-Macro of 48.23%. Because the majority Bedroom class accounts for about 80% of the test set, overall accuracy is compressed against a high base-rate ceiling and is therefore relatively insensitive; the +1.30% accuracy gain over the strongest baseline (*BiLSTM*) should be read together with the substantially larger gains on the imbalance-aware metrics, namely +1.77% in F1-Macro and +5.04% in MCC, which are the decisive criteria for long-tail localisation. By synergizing local temporal dynamics with learned global semantic contexts, our model mitigates the data-inefficiency typical of tabular deep learning and improves robustness on the minority location classes.

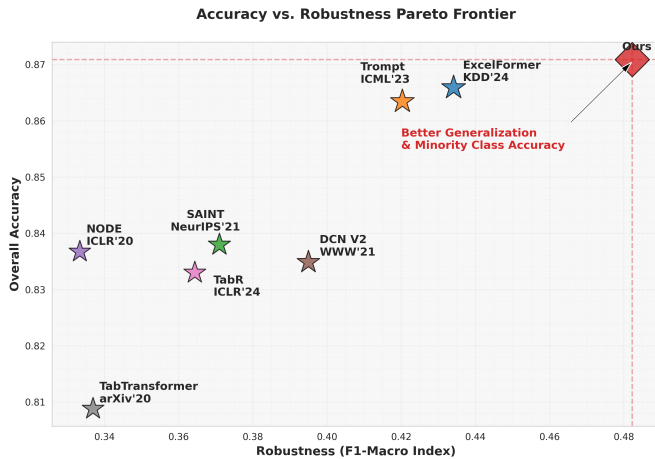


Fig. 6. Pareto Frontier analysis comparing our method against baselines.

TABLE III

COMPARISON OF ARCHITECTURAL PROPERTIES ACROSS TABULAR DEEP LEARNING METHODS. “✓” INDICATES THE METHOD SUPPORTS THE CORRESPONDING CAPABILITY, “×” INDICATES NO SUPPORT, AND “ Δ ” INDICATES PARTIAL SUPPORT.

Method	Feature Types	Temporal	Inter-sample	Retrieval
TabTransformer	Δ Categorical	×	×	×
SAINT	✓ Mixed	×	✓ Row attn.	×
NODE	✓ Mixed	×	×	×
DCN V2	✓ Mixed	×	×	×
Trompt	✓ Mixed	×	×	×
TabR	✓ Mixed	×	✓ k-NN	✓ Static
ExcelFormer	✓ Mixed	×	×	×
Ours	✓ Mixed	✓	✓ Semantic	✓ Hybrid

B. Comparative Analysis with Modern state-of-the-art baseline methods

a) *Comparison with Modern Tabular Deep Learning Methods:* We evaluate our method against seven state-of-the-art tabular deep learning approaches spanning diverse architectural paradigms: transformer-based methods (TabTransformer [53], ExcelFormer [58]), attention mechanisms (SAINT) [54], tree-neural hybrids (NODE) [55], feature interaction networks (DCN V2) [56], prompt-based learning (Trompt) [57], and retrieval-augmented models (TabR) [57]. As shown in Table I, our method attains the best overall accuracy (87.09%). As expected under the strong class imbalance, this accuracy margin over ExcelFormer is small (+0.50%); the improvement instead concentrates on the class-imbalance-aware metrics, most notably +4.82% in F1-Macro and +0.87% in Balanced Accuracy. We do not claim uniform dominance: relative to ExcelFormer the method is marginally lower on MCC (−0.80%) and weighted precision (−2.07%), so we position the result as a Pareto-optimal accuracy/robustness balance rather than an across-the-board gain. Furthermore, the Accuracy vs. Robustness Pareto Frontier depicted in Fig. 6 illustrates this trade-off. Among all evaluated models, our approach resides firmly in the top-right corner, establishing a performance optimum by achieving a Pareto-optimal balance between overall classification accuracy and minority-class robustness. This improvement is particularly significant given the inherent class imbalance in indoor localisation scenarios, where minority location classes are often underrepresented.

Notably, while SAINT achieves the highest weighted precision (90.10%), this comes at the cost of being overly conservative—predicting predominantly high-confidence samples while sacrificing recall on minority classes, as evidenced by its considerably lower F1-Macro (37.10% vs. our 48.23%). In contrast, our temporal-semantic hybrid retrieval mechanism effectively captures discriminative patterns across all location classes, achieving superior balanced accuracy (61.67%) without compromising prediction confidence. The marginal difference in MCC (−0.8%) compared to ExcelFormer further confirms that our performance gains are achieved through genuinely improved classification rather than prediction bias. The architectural properties of all methods are shown in Table III.

C. Ablation Study

TABLE IV

ABLATION STUDY OF DIFFERENT RETRIEVAL PARADIGMS ON THE TABPFN-HYBRID FRAMEWORK. ALL STRATEGIES SHARE THE IDENTICAL FROZEN TABPFN FOUNDATION MODEL ($n_{EST}=32$) AND DIFFER ONLY IN THE RETRIEVAL POLICY; IN PARTICULAR, NAIVE (PURE-TEMPORAL) USES THE SAME MODEL WITH ONLY THE TEMPORAL ANCHOR $\mathcal{C}_{anc}$ AND NO SEMANTIC RETRIEVAL.

Retrieval Strategy	Acc. (%)	F1-Macro (%)	Recall-Macro (%)
Naive (Pure-Temporal)	84.42	38.25	58.05
Semantic-Only (NCA)	80.32	37.95	40.09
Hybrid (Euclidean)	2.86	8.06	21.77
Proposed Hybrid	87.09	48.23	61.67

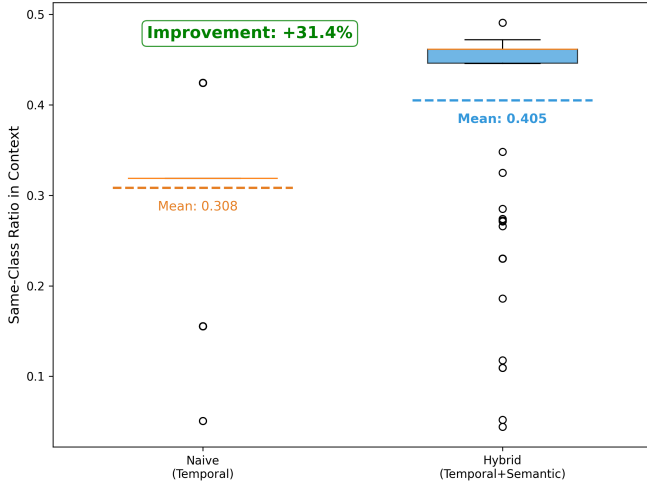


Fig. 7. Quantifies the significant gain in context retrieval quality. By integrating semantic augmentation, the hybrid strategy elevates the same-class sample ratio by **31.4%**, providing high-fidelity demonstrations for effective In-Context Learning.

To further investigate the synergistic effects of our hybrid retrieval mechanism, we conduct an ablation study comparing the proposed method against two degenerate baselines: *Temporal-Only* and *Semantic-Only* (NCA) context mining. Crucially, all variants are evaluated on the identical frozen TabPFN foundation model and differ only in how the in-context set is retrieved: the Naive (Pure-Temporal) baseline feeds the same model with only the temporal anchor $\mathcal{C}_{anc}$ (the most recent window) and performs no semantic retrieval, so the improvement from 84.42% to 87.09% isolates the contribution of the NCA-based semantic stream. As summarized in Table IV, the *Temporal-Only* strategy, while capturing local signal continuity, fails to retrieve informative patterns outside the immediate sliding window. Conversely, the *Semantic-Only* approach yields suboptimal results due to the neglect of temporal dependencies in sensor sequences. Notably, applying hybrid retrieval directly in the unlearned raw feature space (*Hybrid-Euclidean*) results in severe “metric collapse,” with accuracy plummeting to 2.86%. This may be because sparse high-dimensional sensor spaces lack meaningful semantic neighborhoods under Euclidean distance. This catastrophic failure serves as compelling evidence that high-dimensional sparse sensor data lacks effective semantic

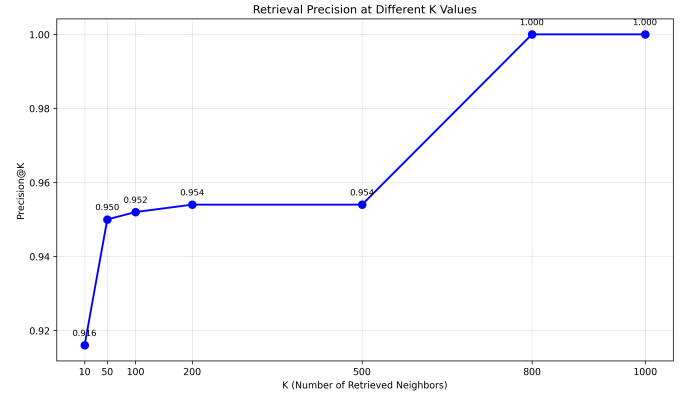


Fig. 8. **Sensitivity Analysis of Semantic Context Budget K .** The curve illustrates the retrieval hit rate (labeled Precision@K) across varying neighbor counts. The rapid convergence at $K = 50$ confirms the manifold’s compactness for typical classes, while the saturation point at $K = 800$ justifies our deployed budget ($K \approx 1024$) as a necessary condition to encompass sparse, long-tail events within the reliable retrieval horizon, ensuring robust noise resistance.

neighborhood structures in Euclidean space, necessitating the reconstruction of the metric space via NCA-induced manifold learning. Our proposed hybrid strategy bridges these two paradigms, achieving a substantial improvement in both F1-Macro (+9.98% absolute gain) and Recall-Macro (+3.62% absolute gain). Fig. 7 quantitatively evaluates the quality of context retrieved by different strategies using the *Same-Class Ratio* metric. Empirical results demonstrate that our proposed hybrid strategy achieves a significant **31.4%** improvement over the *Naive* temporal baseline, elevating the average context purity from 0.308 to 0.405. This validation suggests that the semantic embedding learned via NCA effectively transcends temporal boundaries to retrieve highly representative historical exemplars. Such high-fidelity context injection serves as the primary driver for the improved discriminative capability of foundation model in heterogeneous and imbalanced indoor environments. This evidence demonstrates that integrated spatio-temporal context retrieval is essential for enhancing model robustness in imbalanced indoor localisation tasks, improving robustness on the imbalance-aware metrics.

To validate the rationality of the context budget K , we evaluate the semantic coverage capability of the NCA manifold. The retrieval precision curve in Fig. 8 not only validates the parameter selection but also reveals a *three-tier robustness mechanism* inherent in the system’s design for non-stationary streams. First, the rapid ascent to 0.950 precision at a minimal retrieval depth ($K = 50$) confirms the extreme *topological compactness* of the NCA-induced manifold, demonstrating that 95% of canonical activities are densely clustered and easily retrievable. Second, the subsequent plateau requiring a broad search radius up to $K = 800$ (to capture the remaining 5% gain) highlights the *physical necessity* of a large context budget. This expanded horizon is strictly required not for general inference, but specifically to anchor **long-tail, sparse anomalies** that lie on the manifold’s periphery. Third, our configured operational set-point ($K_{sem} \approx 1024$) extends beyond this saturation threshold, establishing a critical **computational**

safety margin. This design ensures that the system maintains a sufficient semantic buffer to handle potential distribution shifts or label space expansions without overfitting, thereby improving robustness to distribution shift.

D. Hyperparameter Sensitivity and Stability Analysis

To gain deeper insights into the operational mechanics of the proposed hybrid framework, we investigated the influence of key hyperparameters on localisation accuracy across three dimensions. As illustrated in Table V, the experimental results reveal distinct performance patterns. First, regarding the **fusion ratio** (α), the system reaches its peak accuracy of 87.09% when temporal anchors and semantic retrieval are balanced at an equal ratio ($\alpha = 0.5$). Second, the **context scale** (K) involves an information-noise trade-off. While increasing K from 1024 to 2048 enhances the reference depth, further expansion to 4096 results in a 2.73% accuracy drop due to the inclusion of irrelevant historical noise. Finally, the **feature dimension** (Dim) demonstrates that a 16-dimensional embedding is optimal; higher dimensions introduce redundant variance that hinders retrieval precision. These findings confirm ($K = 2048, \alpha = 0.5, Dim = 16$) as the most reliable configuration for practical deployment. The complete test data are provided in the supplementary materials.

TABLE V

PERFORMANCE IMPACT OF KEY HYPERPARAMETERS ON ACCURACY. BOLD VALUES REPRESENT THE BEST CONFIGURATION ACHIEVED BY OUR HYBRID DESIGN.

Trajectory	Variant	Acc. (%)	F1-W (%)	Observation
Ratio α (Fixed K, Dim)	$\alpha = 0.2$	80.76	83.50	Single-mode collapse Optimal Synergy Conservative bias
	$\alpha = 0.5$	87.09	86.56	
	$\alpha = 0.8$	86.59	83.68	
Scale K (Fixed α, Dim)	$K = 1024$	86.84	89.38	Information sparsity Saturation point Context pollution
	$K = 2048$	87.09	86.56	
	$K = 4096$	84.36	83.50	
Metric Dim (Fixed K, α)	$Dim = 16$	87.09	86.56	Low-rank efficient Redundancy noise
	$Dim = 32$	85.79	84.01	

E. Diversity Analysis of Retrieval Augmentation

To further investigate the impact of context representation on ICL efficiency, we compare our proposed hybrid strategy with five representative single-mode retrieval strategies (S_1 – S_5). As summarised in Table VI, the experimental results present an insightful trend: in indoor localisation scenarios, single-mode retrieval based solely on feature similarity often fails to surpass the *Naive* temporal baseline. This phenomenon reveals that temporal consistency serves as an *indispensable prior constraint* in localisation tasks—implying that short-range evolutionary patterns between samples carry higher mutual information than global, static feature similarities.

However, a standalone temporal window exhibits clear limitations when facing signal fluctuations or long-tail distributions. Although S_1 (NCA) and S_3 (MMR) capture certain global semantic features through metric learning and diversity sampling, they still fall short in terms of comprehensive reliability. The most significant observation lies in the catastrophic failure of S_2 , S_4 , and S_5 . From a representation learning

TABLE VI
PERFORMANCE COMPARISON OF VARIOUS CONTEXT RETRIEVAL STRATEGIES. THE “BASE” CONFIGURATION RELIES SOLELY ON RECENT TEMPORAL HISTORY, WHILE “S1–S5” REPRESENT SINGLE-MODE SEMANTIC OR STRUCTURAL RETRIEVALS. OUR PROPOSED “HYBRID” APPROACH INTEGRATES THE TEMPORAL ANCHOR WITH NCA-BASED RETRIEVAL, ACHIEVING BEST RESULTS.

ID	Retrieval Strategy	Acc (%)	F1-Weighted (%)
Base	Naive (Recent History Only)	84.42	82.54
S1	Contrastive (NCA) [52]	80.32	80.61
S3	Diversity (MMR) [59]	73.81	81.84
S2	Spectral (FFT) [60]	9.19	13.32
S4	Textual (TF-IDF Proxy) [61]	5.15	7.28
S5	Tree-Kernel (RF Path) [62]	4.41	5.79
Ours	Hybrid (Temporal + NCA)	87.09	86.55

perspective, the collapse of S_4 (Symbolic Proxy) and S_5 (Tree-Kernel) stems from the “over-coarsened” signal gradients. These schemes process continuous features through rigid interval partitioning or discretization, which inadvertently flattens the subtle signal fluctuations essential for reflecting spatial displacements. This leads the retrieval process into a trap of “false correlation”—where physical locations are far apart yet feature representations appear highly similar—thereby introducing severe noise bias into the model. Ultimately, our hybrid strategy achieves a definitive performance leap by dynamically decoupling and fusing the stability of temporal anchors with the global semantic reach of Neighborhood Components Analysis (NCA). This approach successfully elevates accuracy from 84.42% to 87.09%. This improvement not only validates the necessity of the hybrid design but also establishes an optimal solution for context selection in complex environments.

In summary, the experimental results validate the effectiveness of the MA-ICL framework in processing complex indoor sensing data. By effectively integrating temporal anchors with semantic retrieval, our approach not only outperforms state-of-the-art models in overall accuracy and minority-class robustness but also successfully overcomes the metric collapse inherent in high-dimensional sparse sensor streams. These findings directly fulfill the research objectives of this paper, providing a stable, drift-aware sensing-inference solution built on a foundation model that approximates the Bayesian posterior for resource-constrained IoT scenarios without the need for costly online retraining.

Limitations

While the Hybrid TabPFN framework demonstrates strong robustness in handling non-stationary data streams, we acknowledge limitations regarding computational efficiency and context constraints. Compared to lightweight Decision Trees or Random Forests, our method introduces a higher per-sample inference cost. As quantified in the supplementary material, this cost is dominated by the frozen Transformer’s quadratic self-attention over the K -exemplar context (a single forward pass of 2.56 ms per sample), whereas the NCA retrieval adds only 0.025 ms and is constant in the history size, and the semantic memory is kept at a fixed capacity rather than

growing without bound. The method performs no gradient training, and a streaming deployment runs in real time (above 1.6 Hz) across a tunable latency–accuracy operating range. The principal practical constraint is memory: the frozen foundation model occupies about 1 GB at peak, so the current implementation targets edge devices with a modest GPU rather than ultra-low-power microcontrollers, for which model distillation would be required. Furthermore, although our hybrid retrieval strategy effectively maximises information density within the fixed context window, the theoretical upper bound on sequence length still restricts the model from accessing infinite historical horizons; we take a first step with a bounded, fixed-capacity online memory (supplementary material), and more advanced memory-compression mechanisms for lifelong learning remain an important direction. Finally, our evaluation is conducted on a real-world testbed (the Robot House); although we additionally characterise robustness under continuous-stream distribution shift in the supplementary material, generalization across different homes, residents and sensor layouts is not established here and is an important direction for future work.

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a novel context-aware framework built upon the *Tabular Foundation Model*, fulfilling the core objective of robust, drift-aware context selection for a foundation model that approximates Bayesian posterior prediction, on non-stationary ambient sensor streams, through direct in-context inference. By resolving the “Metric Collapse” phenomenon in high-dimensional sparse sensor data through differentiable *Neighborhood Components Analysis (NCA)*, we constructed a hybrid retrieval system that synergizes temporal stability with semantic relevance. Extensive experiments on the real-world Robot House dataset show consistent improvements on the robustness-oriented metrics, indicating improved robustness against long-term non-stationary distribution shifts and long-tail events. Furthermore, our findings provide systematic guidelines for quantifying the trade-offs between context budget and inference accuracy, directly addressing the requirements for practical deployment in resource-constrained IoT scenarios. This marks a successful application of the retrieval-augmented in-context learning paradigm for tabular foundation models in IoT sensing. Looking ahead, future research will focus on two strategic directions: first, exploring *Knowledge Distillation* to transfer the Bayesian priors of the Foundation Model into lightweight edge networks for deployment on ultra-low-power microcontrollers that lack a GPU; and second, investigating the transferability of the model across different buildings or floors. By leveraging meta-learning or domain adaptation techniques, we aim to enable the model to rapidly transfer learned spatial metric knowledge even with minimal target-site samples, addressing the prohibitive site-survey costs associated with deployment.

ACKNOWLEDGMENTS

This work has been supported by the Hospital@Home project funded by the Dinwoodie Charitable Company and further support received from the University of Hertfordshire development funds.

APPENDIX

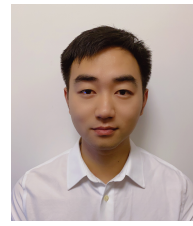
Please see supplementary materials for technical details, list of abbreviations, detailed algorithmic parameter configurations, extended experimental results and in-depth introduction to the Robot House platform.

REFERENCES

- [1] G. Forbes, S. Massie, and S. Craw, “Fall prediction using behavioural modelling from sensor data in smart homes,” *Artificial Intelligence Review*, vol. 53, no. 2, pp. 1071–1091, 2020.
- [2] Y. Hu, Y. Li, J. Wang, and Y. Yan, “Indoor human localization using electrostatic sensors and compressive sensing techniques,” *IEEE Transactions on Instrumentation and Measurement*, 2025.
- [3] S. Rafique, P. Holthaus, G. Fang, and F. Amirabdollahian, “Interpretable room-level human presence detection using ambient sensors in smart homes,” in *17th International Conference on Ubiquitous Computing and Ambient Intelligence (UCAI 2025)*, 2025.
- [4] S. M. Kahya, M. S. Yavuz, and E. Steinbach, “Hood: Real-time human presence and out-of-distribution detection using fmcw radar,” *IEEE Transactions on Radar Systems*, 2024.
- [5] G. Zhang, D. Zhang, Y. He, J. Chen, F. Zhou, and Y. Chen, “Multi-person passive wifi indoor localization with intelligent reflecting surface,” *IEEE Transactions on Wireless Communications*, vol. 22, no. 10, pp. 6534–6546, 2023.
- [6] N. Hollmann, S. Müller, K. Eggensperger, and F. Hutter, “TabPFN: A transformer that solves small tabular classification problems in a second,” in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=cp5PvcI6w8_
- [7] N. Hollmann, S. Müller, L. Purucker, A. Krishnakumar, M. Körfer, S. B. Hoo, R. T. Schirmeister, and F. Hutter, “Accurate predictions on small data with a tabular foundation model,” *Nature*, vol. 637, no. 8045, pp. 319–326, 2025.
- [8] S. Hegselmann, A. Buendia, H. Lang, M. Agrawal, X. Jiang, and D. Sontag, “Tabllm: Few-shot classification of tabular data with large language models,” in *International conference on artificial intelligence and statistics*. PMLR, 2023, pp. 5549–5581.
- [9] V. Thomas, J. Ma, R. Hosseinzadeh, K. Golestan, G. Yu, M. Volkovs, and A. Caterini, “Retrieval and fine-tuning for in-context tabular models,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 108 439–108 467, 2024.
- [10] University of Hertfordshire, “Robot house,” <https://robohouse.herts.ac.uk/>, accessed: 2025-06-12.
- [11] A. Alteneiji, A. Suliman, K. Poon, and U. Ahmad, “Wifi sensing for presence detection using deep learning,” in *2024 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*. IEEE, 2024, pp. 1–6.
- [12] Y. Hu, F. Zhang, C. Wu, B. Wang, and K. R. Liu, “Defall: Environment-independent passive fall detection using wifi,” *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8515–8530, 2021.
- [13] J. Yun and J. Woo, “A comparative analysis of deep learning and machine learning on detecting movement directions using pir sensors,” *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2855–2868, 2019.
- [14] B. Yang, J. Li, Z. Shao, and H. Zhang, “Robust uwb indoor localization for nlos scenes via learning spatial-temporal features,” *IEEE Sensors Journal*, vol. 22, no. 8, pp. 7990–8000, 2022.
- [15] M. Musci, D. De Martini, N. Blago, T. Facchinetti, and M. Piastra, “Online fall detection using recurrent neural networks on smart wearable devices,” *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1276–1289, 2020.
- [16] J. Li, X. Xu, Z. Liu, S. Yuan, M. Cao, and L. Xie, “Aeos: Active environment-aware optimal scanning control for uav lidar-inertial odometry in complex scenes,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 232, pp. 476–491, 2026.
- [17] J. Li, T.-M. Nguyen, M. Cao, S. Yuan, T.-Y. Hung, and L. Xie, “Graph optimality-aware stochastic lidar bundle adjustment with progressive spatial smoothing,” *IEEE Transactions on Intelligent Transportation Systems*, 2025.
- [18] I. Ahmad, A. Ullah, and W. Choi, “Wifi-based human sensing with deep learning: Recent advances, challenges, and opportunities,” *IEEE Open Journal of the Communications Society*, vol. 5, pp. 3595–3623, 2024.

- [19] K. Ngamakeur, S. Yongchareon, J. Yu, and Q. Z. Sheng, "Deep cnn-lstm network for indoor location estimation using analog signals of passive infrared sensors," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22 582–22 594, 2022.
- [20] J. Ding and Y. Wang, "A wifi-based smart home fall detection system using recurrent neural network," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 4, pp. 308–317, 2020.
- [21] Y. Rubanova, R. T. Chen, and D. K. Duvenaud, "Latent ordinary differential equations for irregularly-sampled time series," *Advances in neural information processing systems*, vol. 32, 2019.
- [22] P. Kidger, J. Morrill, J. Foster, and T. Lyons, "Neural controlled differential equations for irregular time series," *Advances in neural information processing systems*, vol. 33, pp. 6696–6707, 2020.
- [23] O. Shchur, A. C. Türkmen, T. Januschowski, and S. Günnemann, "Neural temporal point processes: A review," *arXiv preprint arXiv:2104.03528*, 2021.
- [24] P. Mohapatra, Y. Sui, A. Pandey, S. Xia, and Q. Zhu, "Maestro: Adaptive sparse attention and robust learning for multimodal dynamic time series," *arXiv preprint arXiv:2509.25278*, 2025.
- [25] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "Csdi: Conditional score-based diffusion models for probabilistic time series imputation," *Advances in neural information processing systems*, vol. 34, pp. 24 804–24 816, 2021.
- [26] J. Wang, W. Du, Y. Yang, L. Qian, W. Cao, K. Zhang, W. Wang, Y. Liang, and Q. Wen, "Deep learning for multivariate time series imputation: A survey," *arXiv preprint arXiv:2402.04059*, 2024.
- [27] M. Ma, J. Ren, L. Zhao, D. Testuggine, and X. Peng, "Are multi-modal transformers robust to missing modality?" in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 18 177–18 186.
- [28] W. Xue, T. Zhou, Q. Wen, J. Gao, B. Ding, and R. Jin, "Card: Channel aligned robust blend transformer for time series forecasting, 2024," *URL https://arxiv.org/abs/2305.12095*.
- [29] S. Ö. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 8, 2021, pp. 6679–6687.
- [30] X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "Tabtransformer: tabular data modeling using contextual embeddings (2020)," *arXiv preprint arXiv:2012.06678*, 2012.
- [31] G. Somepalli, M. Goldblum, A. Schwarzschild, C. B. Bruss, and T. Goldstein, "Saint: Improved neural networks for tabular data via row attention and contrastive pre-training," *arXiv preprint arXiv:2106.01342*, 2021.
- [32] S. Bloch, R. de Jong, and E. C. Sertöz, "Heights on curves and limits of hodge structures," *Journal of the London Mathematical Society*, vol. 108, no. 1, pp. 340–361, 2023.
- [33] M. Abd Elaziz, I. A. Fares, A. Dahou, and M. Shrahili, "Federated learning framework for iot intrusion detection using tab transformer and nature-inspired hyperparameter optimization," *Frontiers in Big Data*, vol. 8, p. 1526480, 2025.
- [34] H. Li and Y. Li, "A novel explanatory tabular neural network to predicting traffic incident duration using traffic safety big data," *Mathematics*, vol. 11, no. 13, p. 2915, 2023.
- [35] T. Liu, H. Wang, Y. Guo, Y. Ye, B. Weng, X. Li, J. Chen, S. Xie, G. Zhong, Z. Song *et al.*, "Tabular prior-data fitted network in real-world ct radiomics: benign vs. malignant renal tumor classification," *Quantitative Imaging in Medicine and Surgery*, vol. 15, no. 11, pp. 10 847–10 861, 2025.
- [36] C. Plizzari, G. Goletto, A. Furnari, S. Bansal, F. Ragusa, G. M. Farinella, D. Damen, and T. Tommasi, "An outlook into the future of egocentric vision," *International Journal of Computer Vision*, vol. 132, no. 11, pp. 4880–4936, 2024.
- [37] Q. Luo, Y. Ye, S. Liang, Z. Zhang, Y. Qin, Y. Lu, Y. Wu, X. Cong, Y. Lin, Y. Zhang *et al.*, "Repoagent: An llm-powered open-source framework for repository-level code documentation generation," *arXiv preprint arXiv:2402.16667*, 2024.
- [38] K. Helli, D. Schnurr, N. Hollmann, S. Müller, and F. Hutter, "Drift-resilient tabpfn: In-context learning temporal distribution shifts on tabular data," *Advances in Neural Information Processing Systems*, vol. 37, pp. 98 742–98 781, 2024.
- [39] J. Liu, D. Shen, Y. Zhang, W. B. Dolan, L. Carin, and W. Chen, "What makes good in-context examples for gpt-3?" in *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd workshop on knowledge extraction and integration for deep learning architectures*, 2022, pp. 100–114.
- [40] O. Rubin, J. Herzig, and J. Berant, "Learning to retrieve prompts for in-context learning," in *Proceedings of the 2022 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2022, pp. 2655–2671.
- [41] X. Wang, X. Han, W. Huang, D. Dong, and M. R. Scott, "Multi-similarity loss with general pair weighting for deep metric learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5022–5030.
- [42] Q. Qian, L. Shang, B. Sun, J. Hu, H. Li, and R. Jin, "Softtriple loss: Deep metric learning without triplet sampling," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6450–6458.
- [43] S. Kim, D. Kim, M. Cho, and S. Kwak, "Proxy anchor loss for deep metric learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3238–3247.
- [44] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.
- [45] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.
- [46] Y.-C. Chen, L. Li, L. Yu, A. El Kholy, F. Ahmed, Z. Gan, Y. Cheng, and J. Liu, "Uniter: Universal image-text representation learning," in *European conference on computer vision*. Springer, 2020, pp. 104–120.
- [47] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PmLR, 2021, pp. 8748–8763.
- [48] S. Sun, Y.-C. Chen, L. Li, S. Wang, Y. Fang, and J. Liu, "Lightningdot: Pre-training visual-semantic embeddings for real-time image-text retrieval," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 982–997.
- [49] Z. Wang, W. Wei, G. Cong, X.-L. Li, X.-L. Mao, and M. Qiu, "Global context enhanced graph neural networks for session-based recommendation," in *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, 2020, pp. 169–178.
- [50] C. Mavromatis and G. Karypis, "Gnn-rag: Graph neural retrieval for efficient large language model reasoning on knowledge graphs," in *Findings of the Association for Computational Linguistics: ACL 2025*, 2025, pp. 16 682–16 699.
- [51] Y. Gorishniy, I. Rubachev, N. Kartashev, D. Shlenskii, A. Kotelnikov, and A. Babenko, "Tabr: Tabular deep learning meets nearest neighbors," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: <https://openreview.net/forum?id=rhgIgTSSxW>
- [52] J. Goldberger, G. E. Hinton, S. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," *Advances in neural information processing systems*, vol. 17, 2004.
- [53] X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "Tabtransformer: Tabular data modeling using contextual embeddings," *arXiv preprint arXiv:2012.06678*, 2020.
- [54] G. Somepalli, A. Schwarzschild, M. Goldblum, C. B. Bruss, and T. Goldstein, "Saint: Improved neural networks for tabular data via row attention and contrastive pre-training," in *NeurIPS 2022 First Table Representation Workshop*.
- [55] S. Popov, S. Morozov, and A. Babenko, "Neural oblivious decision ensembles for deep learning on tabular data," in *International Conference on Learning Representations*.
- [56] R. Wang, R. Shivanna, D. Cheng, S. Jain, D. Lin, L. Hong, and E. Chi, "Dcn v2: Improved deep and cross network and practical lessons for web-scale learning to rank systems," in *Proceedings of the web conference 2021*, 2021, pp. 1785–1797.
- [57] K.-Y. Chen, P.-H. Chiang, H.-R. Chou, T.-W. Chen, and D. T.-H. Chang, "Trompt: towards a better deep neural network for tabular data," in *Proceedings of the 40th International Conference on Machine Learning*, 2023, pp. 4392–4434.
- [58] J. Chen, J. Yan, Q. Chen, D. Z. Chen, J. Wu, and J. Sun, "Can a deep learning model be a sure bet for tabular prediction?" in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 288–296.
- [59] J. Carbonell and J. Goldstein, "The use of mmr, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 1998, pp. 335–336.

- [60] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases," in *International conference on foundations of data organization and algorithms*. Springer, 1993, pp. 69–84.
- [61] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2003, pp. 2–11.
- [62] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.



Baobing Zhang is a Research Fellow at the School of Engineering and Computer Science, University of Hertfordshire, UK. He received his PhD in Artificial Intelligence from Brunel University London and his M.Sc. in Mathematical Sciences from Queen Mary University of London, after earning a B.Sc. in Computer Science and Engineering from Hebei University of Technology. His research interests include automated machine learning and uncertainty estimation, deep Gaussian processes, image processing and IoT sensing. Dr Zhang serves as a reviewer for IEEE IJCNN and the International Journal of Fuzzy Systems.



Shrish Rafique is a PhD researcher in computer science in the Robotics Research Group at the University of Hertfordshire (UK). She is jointly enrolled at the University of Hertfordshire (UK) and Western Sydney University (Australia), under a dual-award PhD programme. Her research focuses on human-robot interaction, particularly the detection of people, objects, and activities in ambient assisted living environments. Her work applies artificial intelligence to support home-based healthcare and independent living, including human presence detection, contextual awareness, and hospital@home systems. Additionally, she holds a Master's degree in Artificial Intelligence.



Mohamad Reza Shahabian Alashti is a Post-doctoral Research Fellow in Robotics and Deep Learning with the University of Hertfordshire, U.K. He received the Ph.D. degree in Computer Science from the University of Hertfordshire, where his research focused on human activity recognition in ambient assisted living environments. His research interests include ambient intelligence, multimodal sensing, machine learning, deep learning, human-robot interaction, assistive robotics, and healthcare technologies. He has contributed to projects in smart home monitoring, socially assistive robotics, wearable sensing, and AI-enabled healthcare systems, including work related to the Hospital@Home and SWAG projects.



Shadiya Alingal Meethal is a Research Fellow in Human-Robot Interaction at the School of Physics Engineering and Computer Science, University of Hertfordshire, U.K. She obtained her Ph.D. in Computer Science from the University of Hertfordshire, where her research focused on neural correlates of fatigue during Human-Robot Interaction. Her research interests include Rehabilitation Robotics, machine learning, deep learning and physiological signal processing. She is currently involved in the SWAG project and the Hospital@Home project, supporting the development of assistive robotics and AI-enabled solutions for home-based care.



Vignesh Velmurugan is a Lecturer in Robotics Engineering at the University of Hertfordshire. Prior to this, he worked as a post-doctoral research fellow within the Kaspar project, involving the development of robotic companions for educational purposes. He earned his PhD from the University of Hertfordshire, concentrating on rehabilitation robotics for stroke survivors, and continues to explore this area of study. Additionally, he holds a master's degree in robotics from Ecole Centrale de Nantes and Warsaw University of Technology.



Angela Dickinson is an experienced academic with over 25 years experience in undertaking health and social care research with older people. Angela has extensive expertise in a range of research methods including ethnographic qualitative research, visual methods, interviewing, focus groups, survey, diaries, and mixed methods. She has led and worked on a large number of studies and evaluations funded by UK Research Councils, NIHR, local government, Food Standards Agency and voluntary organisations.



Patrick Holthaus, is a Reader in Interactive Assistive Technology (Associate Professor Research) in the Robotics Research Group at the University of Hertfordshire (UK), teaching foundational and advanced topics in computer science with a research focus on nonverbal interactive signals, social credibility and trust in assistive and companion robots. He is further interested in interaction architectures and behaviour coordination in artificially intelligent embodied agents. He is a Fellow of AdvanceHE and registered as a Chartered Engineer. Dr Holthaus has

authored more than 80 peer-reviewed scientific papers in the field of human-robot interaction, in which he drives forward best methods and practices as the coordinator of the UKRI-funded UK-HRI topic group. He co-leads the Horizon Europe project SWAG and Dinwoodie-funded hospital@home teams at the University of Hertfordshire. As manager of the Robot House, a unique facility for human-robot interaction, Dr Holthaus brings together real-world applications and fundamental robotics research. His research involving the Kaspar robot has a direct impact on the lives and learning of children with autism and learning difficulties.



Farshid Amirabdollahian is a leading academic and a professor of Human-Robot Interaction at the University of Hertfordshire's School of Physics, Engineering and Computer Science in the UK. He earned his PhD in Cybernetics from the University of Reading in 2003, focusing his doctoral research on robot-mediated therapies for stroke recovery. He currently serves as Director of the Robotics Research Group and oversees initiatives such as Robot House and the humanoid Kaspar project. His research focuses on adaptive robotics, assistive and rehabilitation technologies, and machine learning approaches to improve how robots interact with people. Prof Amirabdollahian has a long academic career with previous research and leadership roles in UK research centres and collaborative EU projects.



Gabriella Lakatos is a Reader in Social- and Ethorobotics (Associate Professor Research) at the School of Physics, Engineering, and Computer Science at the University of Hertfordshire (UK), and a member of the Robotics Research Group at the same university. She completed her PhD in the field of Ethology at the Eotvos Lorand University, Hungary, after which she transferred her skills to the field of Etho-robotics and Human-Robot Interactions. Her research interest lies in biologically inspired acceptable and credible robot behaviour

design, trust in human-robot interactions, and robot-assisted therapy for vulnerable populations. Her various research projects were funded by Horizon Europe, EPSRC, UKRI: TAS & UK-RAS, the Baily Thomas Charitable Fund, and the Dinwoodie Charitable Company. She currently co-leads the Horizon Europe project SWAG and Dinwoodie-funded Hospital@Home teams at the University of Hertfordshire. Dr Lakatos is a Fellow of the Higher Education Academy, her teaching focusing on Responsible Technology and Advanced Research Topics. She is an editorial board member of two international journals. She has been co-organiser of several iterations of the Trust, Acceptance and Social Cues in Human-Robot Interaction (SCRITA) workshop at the RO-MAN conference and a guest editor of special issues on the same topic in two international journals.