

Magnetic Fingerprinting for Indoor Localization in a Social Interaction Setting

Sotiris Vacanas

Magnetic Fingerprinting for Indoor Localization in a Social Interaction Setting

by

Sotiris Vacanas

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on June 12, 2026.

Student number:	5466709
Project duration:	November 2025 – June 2026
Thesis committee:	Dr. Hayley Hung, TU Delft, supervisor Dr. Stephanie Tan, TU Delft, supervisor Dr. Manon Kok, TU Delft, examiner

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Preface

This thesis is the result of my graduation project carried out at the HOMI-Lab of the Pattern Recognition and Bioinformatics group of Delft University of Technology. The work investigates whether magnetic fingerprinting — an infrastructure-free localization approach — can be used to determine the position of people in a social interaction setting, where movement is unconstrained and driven by interpersonal behavior rather than navigation goals.

I would like to express my sincere gratitude to my supervisors, Dr. Hayley Hung and Dr. Stephanie Tan, for their guidance, feedback, and encouragement throughout this project. Their expertise and thoughtful input were invaluable in shaping both the research direction and the quality of this work.

Research reported in this work was partially or completely facilitated by computational resources and support of the Delft AI Cluster (DAIC) at TU Delft (RRID: SCR_025091) [4], but remains the sole responsibility of the author, not the DAIC team.

*Sotiris Vacanas
Delft, June 2026*

Abstract

Indoor localization in social interaction settings presents unique challenges: participants move freely, stand still for extended periods, and follow no predetermined trajectories. This thesis investigates whether magnetic fingerprinting — an infrastructure-free approach exploiting the spatially unique distortions of the indoor geomagnetic field — can provide reliable zone-level localization under these conditions. An LSTM network is trained on windowed magnetometer sequences collected from body-worn IMUs (chest badge and drinking cup) during a 30-minute social mingling event involving 32 participants. Four square grid resolutions (0.25 m, 0.5 m, 0.75 m, 1.0 m), a hexagonal grid variant, and a zone-merging strategy are systematically evaluated across 16 feature configurations per resolution. Results show that combined dual-sensor features consistently outperform single-sensor variants, and that coarser grids yield higher exact accuracy while mean physical error remains stable across resolutions at approximately 68–78 cm. At 0.5 m resolution, the best configuration places 71% of predictions within 50 cm of the true location, approaching the accuracy of a UWB baseline system while requiring no installed infrastructure. Zone merging and hexagonal tessellation do not provide consistent improvements over the plain square grid, suggesting that magnetic ambiguity rather than data imbalance or cell geometry is the dominant source of error. The findings demonstrate that infrastructure-free magnetic fingerprinting is a practically viable approach for coarse spatial awareness in socially dynamic indoor environments.

Contents

Use of Generative AI	1
1 Introduction	3
2 Background & Motivation	5
3 Related Work	7
3.1 Classical Magnetic Fingerprinting	7
3.2 Sequence-Based Magnetic Fingerprinting	7
3.3 Deep Learning for Feature Representation	8
3.4 Mapping Efficiency and Data Collection Constraints	8
3.5 Multimodal Magnetic Fusion	8
3.6 Summary and Design Implications	9
4 Data	11
4.1 Collection Setting	11
4.2 Ground Truth Acquisition	11
4.3 Data Preprocessing	11
5 Method	13
5.1 Zone Definition	13
5.2 Feature Extraction and Windowing	13
5.3 Model Architecture	14
6 Experimental Setup	17
6.1 Spatial Partitioning Variants	17
6.1.1 Hexagonal Grid	17
6.2 Hyperparameter Optimisation	18
6.3 Evaluation Metrics	18
7 Results	23
7.1 Overview	23
7.2 Effect of Cell Size	23
7.3 Feature Set Comparison	24
7.4 Effect of World-Frame Rotation	25
7.5 Baseline Comparison: Locatify UWB	26
7.6 Spatial Error Distribution	26
7.7 Hyperparameter Analysis	26
7.8 Effect of Zone Merging	27
7.9 Hexagonal Grid	28
8 Discussion	29
8.1 Localization Feasibility	29
8.2 Challenges Specific to Social Settings	29
8.3 Sensor Placement and Feature Design	29
8.4 Limitations and Future Work	30

Use of Generative AI

Claude Code (Anthropic, model: claude-sonnet-4-6) was used during the preparation of this thesis. Its use was limited to the following:

- **Writing assistance:** improving clarity, academic style, and readability of draft text.
- **Text structuring:** suggesting section organisation and restructuring the methodology chapter.
- **Formatting assistance:** converting content into $\text{\LaTeX}$ tables and equations.
- **Code assistance:** generating boilerplate code for data processing and analysis scripts.

All research questions, experimental design, data collection, analysis, results, and interpretations are the author's own. All AI-assisted text was reviewed, verified, and revised by the author before inclusion. The author retains full responsibility for the accuracy and integrity of the content of this thesis.

Introduction

Indoor localization remains a challenging problem in today's world due to the limited availability and reliability of Global Navigation Satellite Systems (GNSS) in enclosed environments. While numerous alternative approaches have been proposed, including Wi-Fi fingerprinting, Bluetooth beacons, and inertial navigation, many of these solutions rely on dedicated infrastructure, suffer from environmental sensitivity, or accumulate drift over time.

An alternative approach used for indoor localization is the ambient geomagnetic field. The Earth's magnetic field provides an interesting alternative that requires no dedicated installed hardware — such as routers, beacons, or anchors — and is therefore referred to as an infrastructure-free solution for indoor localization. Despite the fact that inside indoor environments, the magnetic field becomes distorted by the presence of magnetic materials such as steel beams and electronic devices, these effects are stable over a long period of time, allowing one to model the magnetic field in indoor environments. Early work has demonstrated that magnetic field measurements can be used to create a fingerprint of locations inside buildings which allows for indoor positioning [20, 3].

With the advent of Machine Learning over the past years, more methods have been introduced to better model the complex and subtle patterns present in magnetic field measurements. Neural networks, in particular, have shown particular promise in capturing non-linear relationships between magnetic fingerprints and location [19, 2, 26]. However, existing approaches differ widely in their assumptions regarding data collection, spatial partitioning, trajectory constraints, as well as the use of auxiliary sensors such as inertial measurement units (IMUs). Furthermore, many systems rely on using magnetic measurements in combination with pedestrian dead reckoning, particle filter frameworks, or Kalman filters, making it difficult to isolate the contribution of the magnetic fingerprint itself [25, 22].

This thesis investigates indoor localization using magnetic fingerprinting in a social interaction setting, where human movement is inherently unconstrained, irregular, and influenced by interpersonal behavior. Unlike traditional navigation scenarios, social settings involve spontaneous changes in direction, frequent stops, small movements, body rotations, and close proximity between individuals. These factors introduce additional variability in sensor orientation, motion patterns, and magnetic readings, making localization more challenging but also more representative of real-world indoor behavior.

Rather than focusing on long-term navigation or trajectory reconstruction, the proposed approach formulates localization as a supervised learning problem in which raw magnetometer readings are directly mapped to discrete spatial zones. This design choice allows the system to operate without motion models, step detection, or explicit tracking assumptions, and enables a clearer analysis of how well magnetic fingerprints alone can support localization under socially dynamic conditions.

The work is guided by the following research question(s):

- Can magnetic fingerprinting alone enable reliable indoor zone-level localization in socially unconstrained environments?
- What performance trade-offs can be observed as zone sizes vary?

By addressing these questions through systematic experimentation on a newly collected dataset, we aim to clarify the achievable performance and limitations of learning-based magnetic fingerprinting for indoor localization in realistic socially dynamic environments.

Background & Motivation

The core challenge of indoor localization is determining *where* a person is inside a building. Unlike outdoor positioning, which can rely on GPS (Global Positioning System) signals from satellites, GPS signals are too weak to penetrate building structures reliably, leaving the problem unsolved by standard tools. The general strategy across most indoor approaches is *fingerprinting*: during an offline phase, measurements of some physical signal (e.g. radio strength, magnetic field) are collected at known locations and stored as a reference map; during the online phase, a new measurement from an unknown location is compared against this map to find the closest match and report a position estimate. What distinguishes different approaches is the physical signal used and how the comparison is made.

Despite significant research progress in the field, practical indoor localization systems remain difficult to deploy at scale. Infrastructure-based approaches represent the dominant class of indoor localization solutions. Wi-Fi fingerprinting works by recording the received signal strength (RSS) from nearby access points at many labelled reference locations, building a map that associates each position with its characteristic RSS signature; at runtime, a new RSS measurement is matched against this map to estimate position. While widely deployed, the technique is sensitive to furniture rearrangements, moving people, and changes in access point configuration, all of which alter the signal environment over time. Bluetooth Low Energy (BLE) beacons operate on a similar fingerprinting principle at shorter range and require dedicated beacon hardware to be installed and maintained. Ultra-wideband (UWB) systems achieve sub-30 cm accuracy by measuring the precise time-of-flight of radio pulses between a mobile tag and a set of fixed anchors, but require carefully placed and calibrated infrastructure per venue.

Vision-based methods use cameras or depth sensors to localise a device by matching observed visual features against a pre-built map (visual SLAM) or by recognising known landmarks. Although capable of high accuracy, they raise significant privacy concerns when cameras are deployed in occupied spaces, and performance degrades under poor lighting, occlusion, or featureless environments.

Inertial navigation estimates position by double-integrating accelerometer readings and integrating gyroscope readings to track orientation, requiring no external signals or infrastructure [9]. However, small sensor noise and bias accumulate with each integration step, causing position error to grow unboundedly over time—a phenomenon known as drift. Without periodic correction from an external reference (such as a known landmark or GPS fix), the estimated position diverges rapidly from the true one, making pure inertial navigation impractical for extended indoor use.

Magnetic fingerprinting provides a compelling alternative: it is passive, infrastructure-free and uses sensors that are already widely available. However, the indoor magnetic field exhibits relatively low spatial information density, and measurements are highly sensitive to device orientation, user motion, and sensor placement. Typical fingerprint matching approaches such as Dynamic Time-Warping (DTW) and model-based magnetic SLAM are not well suited to social interaction settings for distinct reasons. DTW-based matching compares a query trajectory against stored reference sequences: it therefore requires the user to follow a path that was previously surveyed, at a comparable speed and orientation. In social settings, participants frequently stand still, change direction arbitrarily, or revisit the same location from different body orientations, making such trajectory-level comparisons unreliable. Magnetic SLAM, as demonstrated by Kok and Solin [11], simultaneously estimates position and builds a mag-

netic map of the environment, but depends on continuous motion to accumulate new observations and propagate position updates. Extended periods of stillness—common during conversations—provide no new trajectory information, stalling the localisation process. Both families of methods are additionally sensitive to sensor orientation and walking speed, which vary substantially in unconstrained social interactions.

Recent surveys have shown that neural-network-based geomagnetic localization shows promise, albeit with some limitations that emphasize the importance of proper feature representation, orientation handling, and inconsistent ground truth acquisition methods [18, 15]. Many existing studies rely on predetermined trajectories [27, 14], static reference points, or simulated trajectories [7], which limits their applicability to real-world use cases.

In contrast, social interaction settings allow people to move naturally without any constraints. Unlike structured data collection protocols—where a surveyor deliberately walks a pre-planned route to ensure uniform spatial coverage—social participants move according to interpersonal goals: they approach conversational partners, form groups, and then stand still for extended periods. This behavior produces trajectories that are short, clustered, and highly repetitive rather than space-filling. Figure 5.1 illustrates this concretely: the occupancy heatmaps show that observations accumulate heavily in a few central zones while peripheral zones are rarely visited. Because each zone must accumulate enough samples for the model to learn a reliable fingerprint, infrequently visited zones are likely to be misclassified, and the classifier may become biased toward the dominant central zones. Addressing these challenges is an important step toward deploying indoor localization systems in realistic environments where human behavior, rather than navigation efficiency, dictates movement patterns.

During data collection, participants were also equipped with a Locatify badge that provides position estimates using a UWB-based localization system. These badges record two-dimensional coordinates and are specified by the manufacturer to achieve an accuracy of up to 30 cm. To assess real-world accuracy in the experimental setting, the badge-reported positions were compared against independent ground-truth positions derived from ceiling-mounted cameras (described in Section 4.2). This validation revealed a mean position error of approximately 50 cm, already exceeding the manufacturer specification, suggesting that UWB accuracy is sensitive to the specific room geometry and anchor placement and may degrade in a small, crowded space with many moving bodies causing multipath signal reflections.

Related Work

Indoor localization has attracted significant research interest due to the limitations of GNSS signals in enclosed environments. Among infrastructure-free approaches, magnetic fingerprinting has emerged as a promising alternative because it uses naturally occurring magnetic distortions caused by building structures. Due to these distortions exhibiting spatial variability, as well as consistent over time, they are suitable for discriminating between locations [16].

3.1. Classical Magnetic Fingerprinting

Magnetic fingerprinting methods typically operate in two phases: an offline survey phase and an online phase for pattern matching. During the survey phase, a person walks through the target area carrying a magnetometer and records magnetic field measurements at known, labelled locations, building a reference map that associates each location with its characteristic magnetic signature. This map is then used during the online phase to match new measurements against stored fingerprints and estimate position. Parralejo et al. [16] conducted a comprehensive evaluation of neural network architectures for magnetic fingerprinting and demonstrated that shallow multilayer perceptrons (MLPs) achieved competitive performance compared to deeper convolutional networks. Their study also highlighted the influence of device heterogeneity on localization accuracy, emphasizing a critical practical challenge in magnetic-based systems.

Similarly, Ashraf et al. [1] investigated deep neural networks (DNNs) for magnetic fingerprint classification using heterogeneous smartphones. Their results confirmed that magnetic fingerprints are discriminative but sensitive to device-dependent variations, which can introduce positioning bias.

However, a common characteristic across these studies is the use of structured data collection protocols, typically involving controlled sampling at labeled reference locations.

3.2. Sequence-Based Magnetic Fingerprinting

Unlike Wi-Fi fingerprints, magnetic signals exhibit strong temporal correlation when collected along walking trajectories. Recent studies exploit this property using recurrent neural networks (RNNs) and long short-term memory (LSTM) models.

The work performed by [6] in 2017 showed that applying RNNs to magnetic sequences yielded promising results, with an average localization error of 1m.

Furthermore, Platte et al. [17] proposed an LSTM-based method operating on sequential geomagnetic vectors fused with gravitational vectors while also eliminating the need for accurate 2D coordinates by simply having the users who collected the data label where the data came from. Their approach demonstrated that temporal modeling significantly improves recognition rates compared to static classification.

Similarly, Liu et al. [12] used spiking neural networks to detect distinctive magnetic landmarks—locations with a uniquely recognisable field signature—and used these detections to periodically correct the accumulated drift of a pedestrian dead reckoning (PDR) system. PDR estimates position by counting steps and tracking heading from inertial sensors, but accumulates error over time; magnetic

landmark detection provides a correction signal whenever the user passes a known landmark. This illustrates how learning-based models can capture characteristic patterns in the magnetic sequence even for a correction task rather than full localisation. Sequence modeling has also been explored in RNN-based geomagnetic localization frameworks, where recurrent architectures outperform traditional fingerprint matching under dynamic motion conditions.

Additionally, Ding et al. [5] proposed a CNN-LSTM hybrid in which a one-dimensional CNN first processes the raw magnetic sequence fingerprint to suppress noise and extract local patterns, producing a flattened feature vector that is then passed into a three-layer LSTM for temporal modelling. The motivation is that raw magnetic sequences are susceptible to environmental noise, making it hard for an LSTM alone to localise accurately from unfiltered input. Their method achieves mean errors of 0.26 m, 0.64 m, and 0.93 m across three environments, outperforming a plain RNN-LSTM baseline by 34–57%. While this architecture is a natural candidate, Klipp et al. [8] directly compared CNN and LSTM networks on magnetic localization and found that LSTM-based models achieved higher precision. This, combined with the relatively small input dimensionality in our setting (at most 8 channels per timestep), makes a plain LSTM a well-justified and interpretable choice over a more complex CNN-LSTM pipeline.

These studies collectively suggest that incorporating temporal dependencies enhances robustness and reduces ambiguity inherent in static magnetic fingerprints. Concretely, Jang et al. [6] reported a mean localization error of 1 m using an RNN on magnetic sequences, while Platte et al. [17] demonstrated that their LSTM-based approach substantially improved recognition rates over static nearest-neighbour matching on the same dataset. Ding et al. [5] further showed that their CNN-LSTM hybrid reduced localization error compared to K-NN and single-layer neural network baselines, with the sequential component being the primary driver of the improvement.

3.3. Deep Learning for Feature Representation

Convolutional neural networks (CNNs) have been applied increasingly to magnetic fingerprinting. Systems such as MINLOC employ CNN-based feature extraction to capture spatial patterns embedded in magnetic signals. Parralejo et al. [16] showed that convolutional architectures provide competitive performance, but may not significantly outperform well-optimized shallow models when the dataset size is limited.

Deep learning approaches improve generalization by learning nonlinear relationships between magnetic vectors and spatial locations. However, they also introduce challenges including overfitting, computational complexity, and training data requirements.

3.4. Mapping Efficiency and Data Collection Constraints

Magnetic fingerprinting typically requires extensive data collection. To address this limitation researchers have proposed a pedometer-free approach. Maficon [23] removes the assumption of constant walking speed and aligns magnetic sequences using trajectory similarity instead of step-based constraints. This improves practical usability while maintaining distinguishable fingerprints.

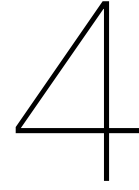
Crowdsourcing-based systems further attempt to reduce manual survey effort. The SVD-CSP [13] framework leverages IMU and magnetic signals to align walking trajectories and construct radio maps using sparse crowd-sensed data. While such approaches reduce manual survey effort, they rely on coherent walking trajectories generated through pedestrian dead reckoning (PDR), thereby introducing movement constraints during data acquisition.

3.5. Multimodal Magnetic Fusion

While pure magnetic fingerprinting is infrastructure-free, some systems incorporate complementary modalities to improve accuracy. For instance, integration of WiFi, geomagnetic, and light sensors via deep feedforward networks [21] has shown improved performance compared to single-modality systems. However, such fusion increases infrastructure dependence and other constraints, while mitigating one of the primary advantages of magnetic-only systems.

3.6. Summary and Design Implications

The body of related work leads to several concrete design decisions for the present study. First, sequence-based models—particularly LSTMs—consistently outperform static fingerprint matching by exploiting temporal structure in magnetic signals [6, 17], motivating the use of a sliding-window LSTM rather than a point-wise classifier. Second, dual-sensor or multi-axis feature sets provide richer spatial information than single-axis or magnitude-only inputs [16, 2], informing the evaluation of multiple feature subsets including combined badge and cup configurations. Third, existing work almost universally relies on structured walking trajectories for data collection and model training; our setting—where participants move freely according to social goals—is fundamentally different and requires an approach that does not assume continuous or directed motion. Finally, the consistent use of predetermined paths or simulated data in prior studies [27, 7] limits how much their results generalise to real human-centered environments, reinforcing the need for evaluation on naturally collected data.



Data

4.1. Collection Setting

Data were collected during a structured 30-minute social mingling event (hereafter Mingling 1), involving 32 participants in an indoor room measuring approximately 5×6 metres. Each participant wore two inertial measurement units (IMUs) sampling at 50 Hz: a badge clipped to the chest, and a sensor embedded in a drinking cup. Both devices recorded three-axis magnetometer (MAG), accelerometer, and gyroscope readings throughout the session. Mingling 1 is used throughout this work.

4.2. Ground Truth Acquisition

Spatial ground truth was obtained from three ceiling-mounted cameras (Camera 06, Camera 08, Camera 10) recording at 60 fps. Human poses were estimated in each camera's image using ViTPose [24], a vision transformer-based pose estimator that predicts the 2D locations of body keypoints (e.g. ankles, knees, hips, shoulders) by regressing heatmaps over the image. Each predicted keypoint is accompanied by a confidence score derived from the peak value of its heatmap, reflecting the model's certainty about the keypoint's location. Keypoints with a confidence score below a threshold of 0.3 were discarded as unreliable detections and not used in subsequent position estimation.

Floor positions were recovered by back-projecting the estimated ankle keypoints — the most reliable foot-level indicator of ground contact — to the room plane using each camera's calibrated extrinsic parameters, transforming all detections into a shared world frame defined by Camera 06. When a participant was visible to more than one camera in the same frame, the per-camera position estimates were merged as a confidence-weighted average, where each camera's contribution was weighted by the mean confidence of its ankle keypoints. This approach gives greater influence to cameras that produced high-confidence detections and suppresses estimates from cameras with occluded or uncertain views. Orientation was taken from the camera yielding the highest mean keypoint confidence for that participant in that frame. This produced one position and orientation record per participant per frame at 60 Hz. Although orientation was recorded, only the 2D floor position is used in this work to assign zone labels; orientation is not used as a training signal or evaluation target.

4.3. Data Preprocessing

Before zone labels could be assigned, the IMU streams and the camera-derived ground truth were brought into a common temporal reference frame. The two data sources operate at different sampling rates: the IMUs record at 50 Hz while the cameras produce position estimates at 60 Hz. Camera ground-truth positions were therefore linearly interpolated to the IMU timestamps, yielding one position record per IMU sample per participant.

Frames in which a participant was not detected by any camera were treated as missing. Consecutive missing segments shorter than 0.5 s were filled by linear interpolation between the surrounding valid estimates; longer gaps were left unlabelled and the corresponding IMU windows were excluded from training and evaluation.

The scalar field magnitude was clipped to a maximum of $200\,\mu\text{T}$ to suppress sensor saturation artefacts; no lower-bound filter was applied. No hard-iron or soft-iron calibration was applied, as the goal of the study is to assess raw magnetometer signals under realistic body-worn conditions.

5

Method

5.1. Zone Definition

The room was discretised into a regular grid of square cells. Four cell sizes were evaluated — 0.25 m, 0.5 m, 0.75 m, and 1.0 m — yielding grids of varying spatial resolution. Zone labels were assigned to each ground-truth position by mapping the (x, y) coordinate to its enclosing grid cell. Cells with no observed occupancy across the session were discarded. Positions outside the room boundary were discarded as outlier back-projections.

Figure 5.1 shows the occupancy heatmaps for each cell size, where colour intensity indicates the total number of ground-truth observations in each zone across all participants and frames. The distribution is markedly uneven: zones near the centre of the room accumulate substantially more observations than those along the walls, reflecting the tendency of participants to congregate in conversational clusters away from the periphery. This imbalance is a natural consequence of the unconstrained social setting and is expected to affect the model’s ability to learn reliable fingerprints for infrequently visited zones. Evaluating coarser cell sizes (0.5 m, 0.75 m, 1.0 m) is partly motivated by this: larger cells aggregate more observations per zone, reducing the severity of the imbalance and potentially enabling more reliable per-zone learning, at the cost of reduced spatial precision.

5.2. Feature Extraction and Windowing

The geomagnetic field measured by body-worn IMUs produces a locally unique temporal pattern as a participant moves through the room [6]. Rather than treating individual magnetometer readings as static fingerprints, we exploit this temporal structure by applying a sliding window to each participant’s time series, following the sequential classification approach of [17].

Both the window length (2–10 seconds) and stride (0.5–2 seconds) were treated as hyperparameters and tuned jointly with the model. Each window produces an 8-dimensional feature vector per timestep: MAG X, Y, Z, and magnitude from the badge, followed by the same four from the cup. The

scalar field magnitude $\|\mathbf{B}\| = \sqrt{B_x^2 + B_y^2 + B_z^2}$ is included alongside the directional axes because it is invariant to sensor orientation: rotating the device changes the distribution of the field across the X, Y, and Z axes but leaves the total magnitude unchanged. This makes the magnitude a stable, orientation-independent indicator of the local field strength, which varies spatially due to the distortions introduced by ferromagnetic building materials. Including it therefore provides a complementary signal to the orientation-sensitive XYZ components.

Eight feature subsets were evaluated, summarised in Table 5.1.

These are Badge XYZ, Badge XYZ + mag, Badge mag only, the corresponding three cup variants (Cup XYZ, Cup XYZ + mag, Cup mag only), and two combined variants (Badge + Cup XYZ and Badge + Cup XYZ + mag), as listed in Table 5.1. Each subset was also evaluated in two rotation variants: raw sensor-frame axes (norot) and axes rotated into the world frame via DMP quaternion (rot), yielding 16 feature configurations per cell size and 64 configurations in total. The Digital Motion Processor (DMP) is an onboard processing unit embedded in the IMU chip that continuously fuses accelerometer, gyroscope, and magnetometer readings to estimate the sensor’s orientation in real time. This fusion is

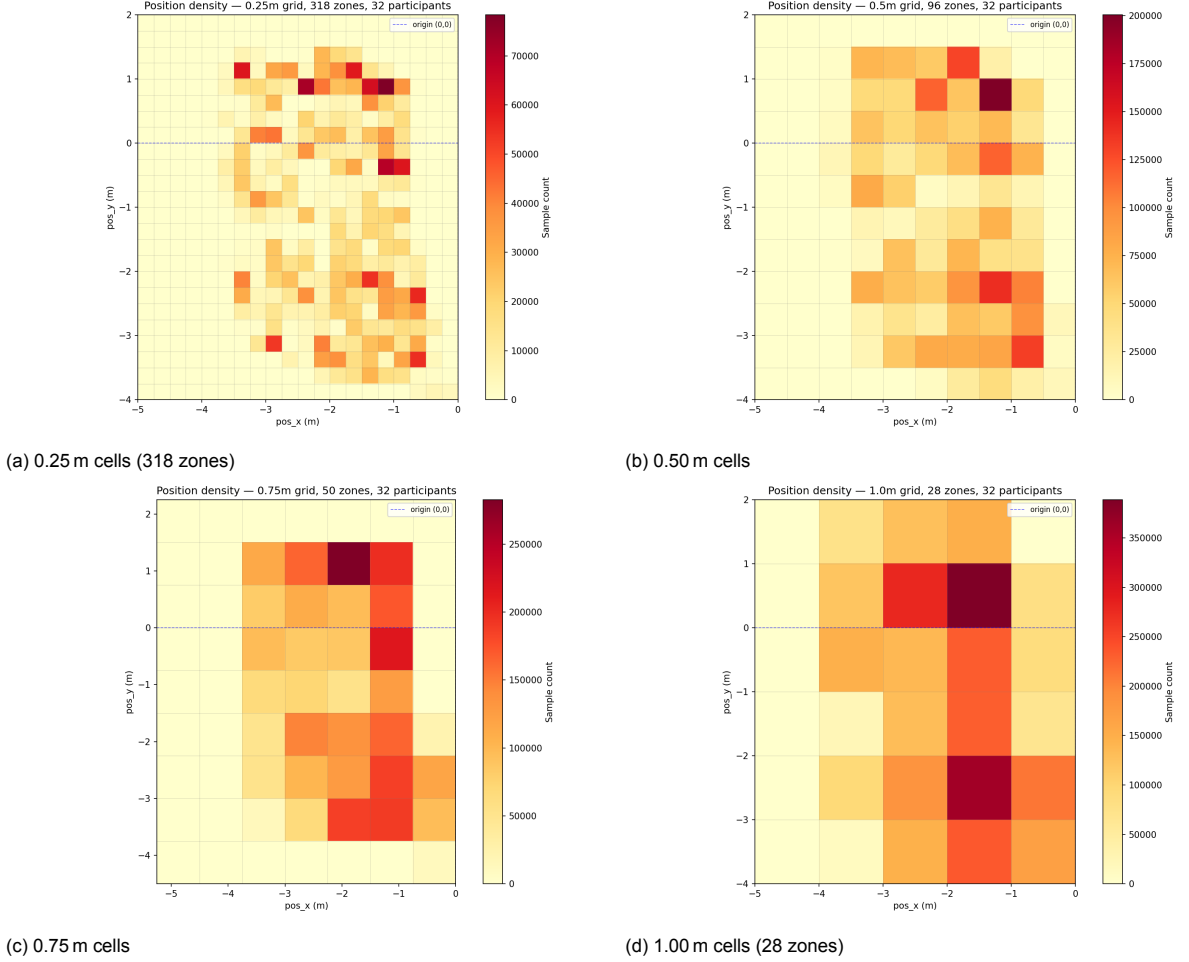


Figure 5.1: Zone occupancy heatmaps for each evaluated cell size. Colour intensity indicates cumulative ground-truth observations per zone across all participants. White cells were never occupied and are excluded from the dataset.

output as a quaternion — a four-component representation of 3D rotation — which can be used to rotate the raw sensor-frame magnetometer axes into a fixed world frame, removing the orientation-dependent component of the signal and leaving only the location-specific ambient field.

A window was retained only if a tunable fraction of its timesteps (label threshold, tuned in the range 0.5–0.95) had a valid zone label. The majority zone across the valid timesteps was assigned as the window label. For each participant, the first 80% of samples by time were used for training and the remaining 20% for testing, preserving temporal order to avoid data leakage.

5.3. Model Architecture

Motivated by prior work demonstrating the effectiveness of LSTM networks for magnetic-based indoor localization [8, 17], zone classification was performed by an LSTM network taking the full window as input and producing a softmax distribution over K zones from the final hidden state. A CNN-LSTM hybrid [5] was considered but not adopted for two reasons. First, Klipp et al. [8] directly compared CNN and LSTM architectures on magnetic localization and found that LSTM achieved higher precision. Second, the denoising CNN stage in [5] was motivated by the local temporal structure present in walking trajectories; in social interaction settings, participants are frequently stationary, reducing the local pattern structure that a 1D-CNN is designed to exploit. A plain LSTM is therefore the more appropriate inductive bias for this setting, and avoids the additional hyperparameters (filter count, kernel size) that a CNN stage would introduce into an already large search space. The model was trained using the Adam optimiser with a cross-entropy loss.

Table 5.1: Feature subsets evaluated. Each subset was tested with and without world-frame rotation (16 configurations per cell size).

Label	Description	Sensor(s)	Channels
Badge XYZ	Magnetometer X, Y, Z axes	Badge	3
Badge XYZ + mag	Axes and field magnitude	Badge	4
Badge mag only	Field magnitude only	Badge	1
Cup XYZ	Magnetometer X, Y, Z axes	Cup	3
Cup XYZ + mag	Axes and field magnitude	Cup	4
Cup mag only	Field magnitude only	Cup	1
Badge + Cup XYZ	Axes from both sensors	Badge + Cup	6
Badge + Cup XYZ + mag	Axes and magnitude from both sensors	Badge + Cup	8

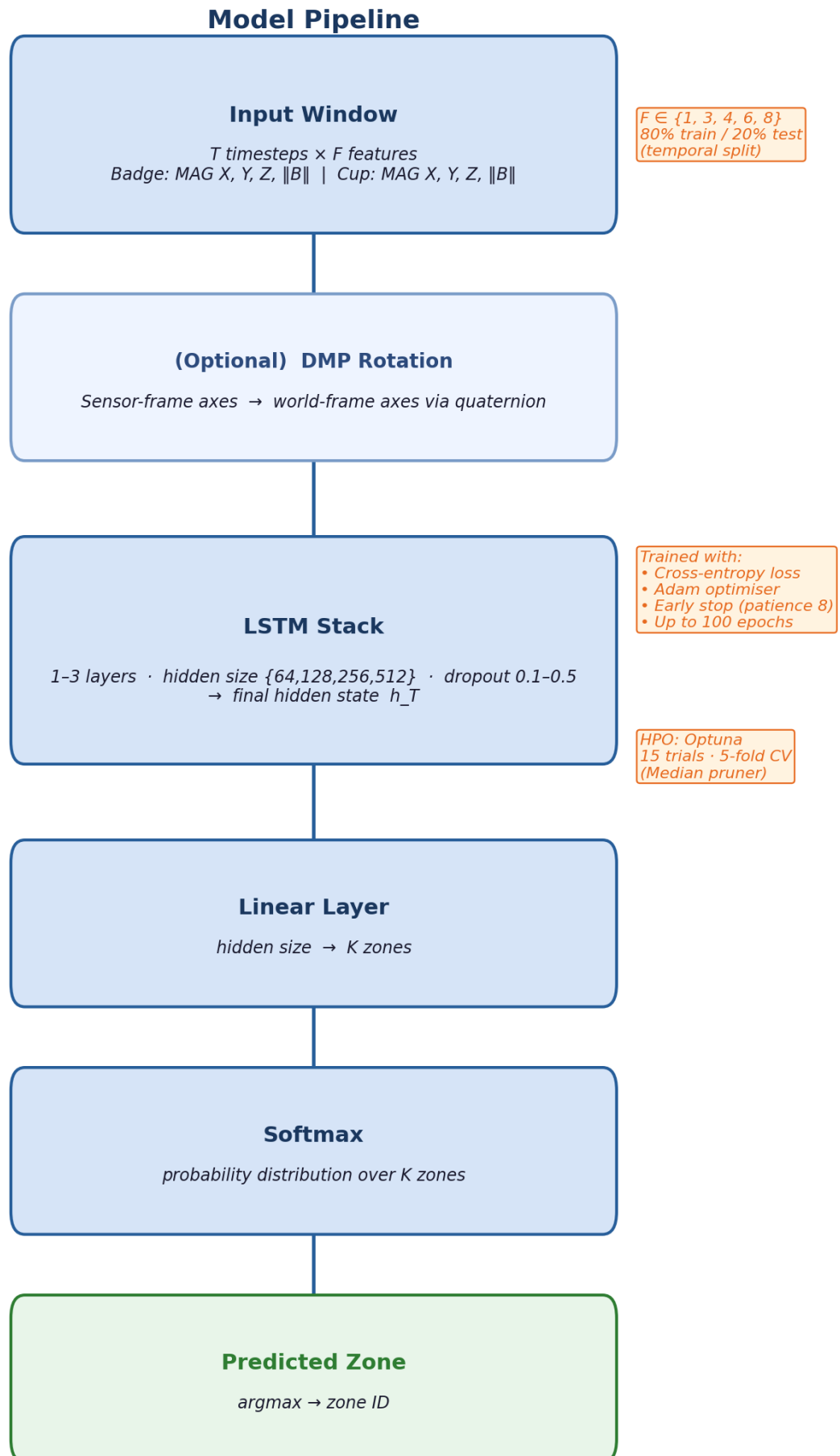


Figure 5.2: Overview of the LSTM classification pipeline. Raw magnetometer windows from the badge and cup sensors are extracted, optionally rotated into the world frame, and fed into an LSTM network whose final hidden state is mapped to a softmax distribution over K zones.

Experimental Setup

6.1. Spatial Partitioning Variants

The unconstrained nature of social interaction produces a markedly uneven distribution of ground-truth observations across zones. Peripheral zones visited infrequently accumulate too few samples to train a reliable per-zone fingerprint. To investigate whether enforcing a minimum sample count per zone improves classification, a region-growing merging procedure was applied as a post-processing step on the plain grid partition.

Zones with fewer than 200 ground-truth observations are iteratively merged with their 8-connected grid neighbours. In each pass, all sparse zones are processed in ascending order of sample count; each is absorbed into the adjacent neighbour with the greatest number of samples. If a sparse zone has no occupied neighbours (completely surrounded by empty cells), it is merged with the nearest occupied zone by grid distance. Passes repeat until every surviving zone meets the 200-sample threshold. Grid cells that were never visited are assigned to the nearest surviving zone root by grid distance, as in the plain grid case.

Table 6.1 shows the number of zones before and after merging for each cell size.

Table 6.1: Zone counts before and after region-growing merging (minimum 200 samples per zone).

Cell size	Zones (plain grid)	Zones (merged)	Zones removed
0.25 m	318	249	69 (22%)
0.50 m	96	76	20 (21%)
0.75 m	50	37	13 (26%)
1.00 m	28	24	4 (14%)

The merged zone partition is used as a drop-in replacement for the plain grid partition: the same LSTM model architecture, feature sets, rotation variants, and Optuna tuning procedure are applied unchanged, with only the zone labels and zone count K updated to reflect the merged boundaries.

6.1.1. Hexagonal Grid

As a second partitioning variant, the room was tessellated with a flat-top hexagonal grid. Hexagonal cells have the geometric property that all six neighbours are equidistant from the cell centre, eliminating the asymmetry between axis-aligned and diagonal adjacency that exists in square grids. The cell size parameter is the *circumradius* — the distance from the cell centre to a vertex — in metres. Circumradii of 0.25 m and 0.5 m were evaluated. Because hex cell area scales as $(3\sqrt{3}/2)s^2$, a circumradius of 0.25 m gives a cell area of approximately 0.16 m^2 (equivalent in area to a 0.40 m square cell), and a circumradius of 0.5 m gives approximately 0.65 m^2 (equivalent to a 0.81 m square cell). Consequently the hexagonal grids are considerably coarser than square grids of the same nominal size: at circumradius 0.25 m the partition yields 72 active zones, and at 0.5 m it yields 28 active zones. Spatial error is measured using the hex distance — the minimum number of hex-neighbour steps between the predicted and true cell — which is the natural analogue of Chebyshev distance for the hexagonal lattice. The

same LSTM architecture, all 16 feature configurations, and the Optuna tuning procedure were applied unchanged.

6.2. Hyperparameter Optimisation

Hyperparameters were tuned using Optuna with 20 trials per configuration and 5-fold cross-validation on the training set. The search space covered: hidden size {64, 128, 256, 512}, number of LSTM layers (1–3), dropout (0.1–0.5), learning rate (10^{-4} – 10^{-2}), batch size {32, 64, 128}, window length, stride, and label threshold. Unpromising trials were pruned early using the Median Pruner with 5 startup trials and 3 warmup steps. The best hyperparameters from tuning were used to retrain a final model on the full training set for up to 100 epochs with early stopping (patience 8).

6.3. Evaluation Metrics

Three metrics are reported for each configuration. **Test accuracy** is the proportion of windows whose predicted zone exactly matches the true zone. **Spatial tolerance** measures the proportion of predictions falling within k grid cells of the true zone, using the Chebyshev distance between the predicted zone (x_p, y_p) and the true zone (x_t, y_t) :

$$d_\infty = \max(|x_p - x_t|, |y_p - y_t|) \quad (6.1)$$

A prediction is counted as within k cells if $d_\infty \leq k$. This metric treats diagonal neighbours the same as axis-aligned neighbours, reflecting the fact that a diagonal step corresponds to the same physical cell width as an axis-aligned step. **Mean cell error** reports the average Chebyshev distance across all test windows, which can be converted to a physical distance by multiplying by the cell size.

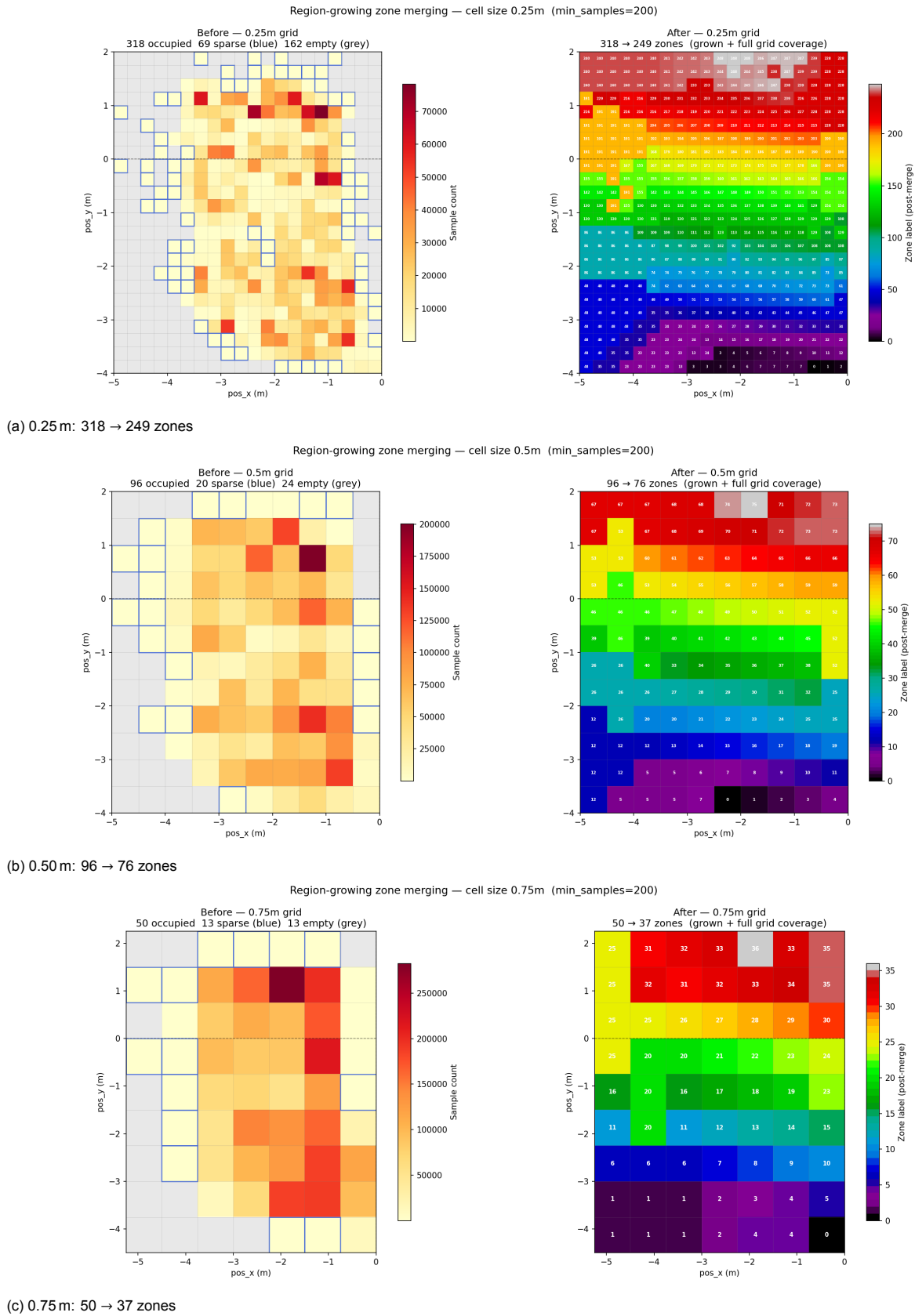
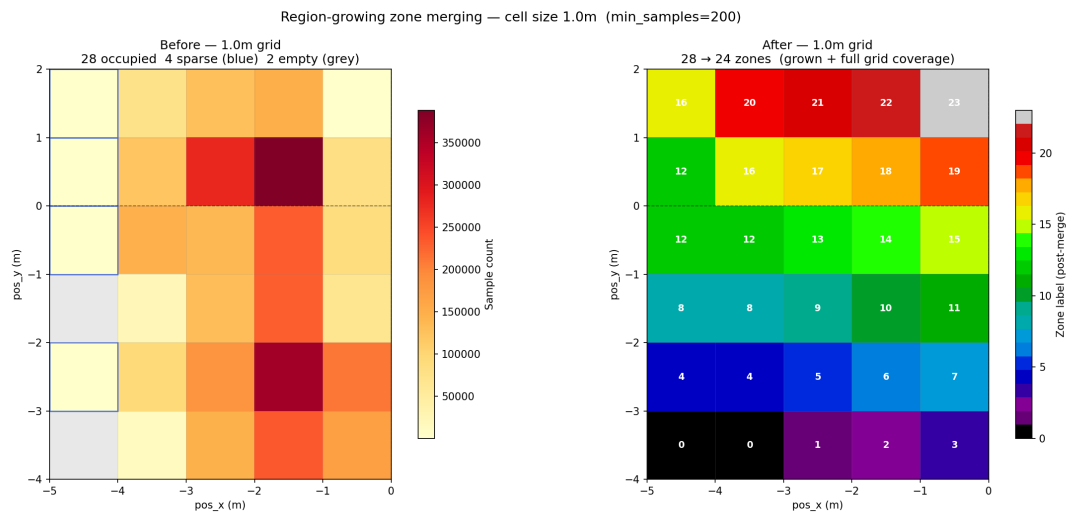


Figure 6.1: Zone partitions before (left half) and after (right half) region-growing merging (0.25m–0.75m). Merged zones are shown with shared colours indicating which original zones were absorbed together.



(a) 1.00 m: 28 → 24 zones

Figure 6.2: Zone partitions before (left half) and after (right half) region-growing merging (1.00 m).

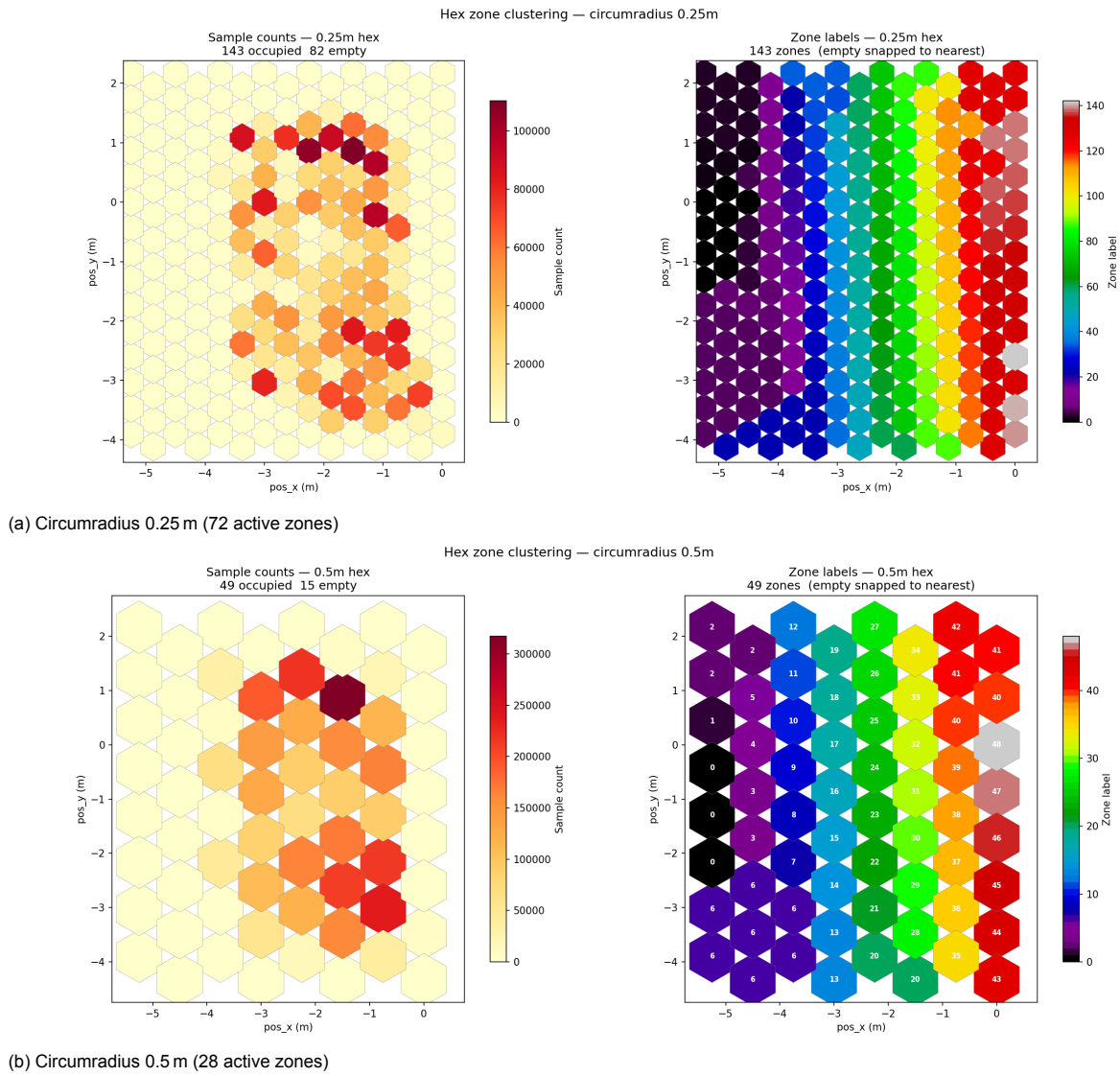


Figure 6.3: Flat-top hexagonal zone partitions for the two evaluated circumradii. Unoccupied cells are excluded from the active zone set.

Results

7.1. Overview

Results are reported for all 64 configurations: 16 feature configurations across each of the four cell sizes (0.25 m, 0.5 m, 0.75 m, and 1.0 m). Metrics are as defined in Section 6: test accuracy, within- k -cells spatial tolerance (Chebyshev), and mean cell error.

7.2. Effect of Cell Size

Table 7.1 summarises the mean and maximum test accuracy across all completed feature configurations for each evaluated cell size.

Table 7.1: Test accuracy and spatial tolerance by cell size (mean and max across all 16 configurations per cell size).

Cell size	Test accuracy		Within 1 cell		Within 2 cells		Mean cell error	
	mean	max	mean	max	mean	max	mean	best
0.25 m (16 configs)	0.160	0.239	0.304	0.450	0.466	0.664	4.56	2.86
0.50 m (16 configs)	0.230	0.360	0.511	0.710	0.690	0.867	2.25	1.36
0.75 m (16 configs)	0.299	0.417	0.649	0.814	0.771	0.856	1.52	1.03
1.00 m (16 configs)	0.325	0.459	0.735	0.860	0.831	0.929	1.17	0.78

As expected, exact zone classification is considerably harder at finer spatial resolution. At 0.25 m, the grid comprises 318 active zones, and the best configuration (Badge + Cup XYZ + mag, unrotated) achieves 23.9% test accuracy. At 0.5 m, the task is simpler and the best configuration (Badge + Cup XYZ + mag, unrotated) achieves 36.0%. In both cases, performance is substantially above the chance level (approximately 0.3% at 0.25 m).

However, exact zone accuracy understates practical localization utility. Expressed in physical units, the within-1-cell criterion corresponds to 25 cm at the 0.25 m resolution and 50 cm at the 0.5 m resolution. At 0.5 m, the best configuration places 71.0% of predictions within 50 cm of the correct location and 86.7% within 1 m. At 0.25 m, 45.0% of predictions fall within 25 cm of the true zone and 66.4% within 50 cm. These tolerance figures are more informative for the intended social interaction use case, where coarse zone-level awareness is often sufficient.

At 0.75 m, the classification task is considerably easier given the coarser grid. The best configuration (Badge + Cup XYZ + mag, rotated) achieves 41.7% test accuracy, with 78.3% of predictions within one cell (75 cm) and 84.2% within two cells (1.5 m). The mean physical error of this configuration is approximately 80 cm. Across all 16 configurations at this resolution the mean test accuracy rises to 29.9%, a clear step up from the 0.5 m tier.

At 1.0 m the grid comprises 28 active zones and the best configuration (Badge + Cup XYZ, unrotated) achieves 45.9% test accuracy — the highest exact-match accuracy across all cell sizes evaluated. Within-cell tolerance further improves: 86.0% of predictions fall within one cell (1 m) and 91.2% within two cells (2 m), with a mean physical error of approximately 78 cm. As the grid becomes coarser,

between-zone magnetic contrast increases and the number of classes decreases, both of which facilitate higher exact accuracy. However, the reduced spatial resolution limits utility for fine-grained localization tasks.

7.3. Feature Set Comparison

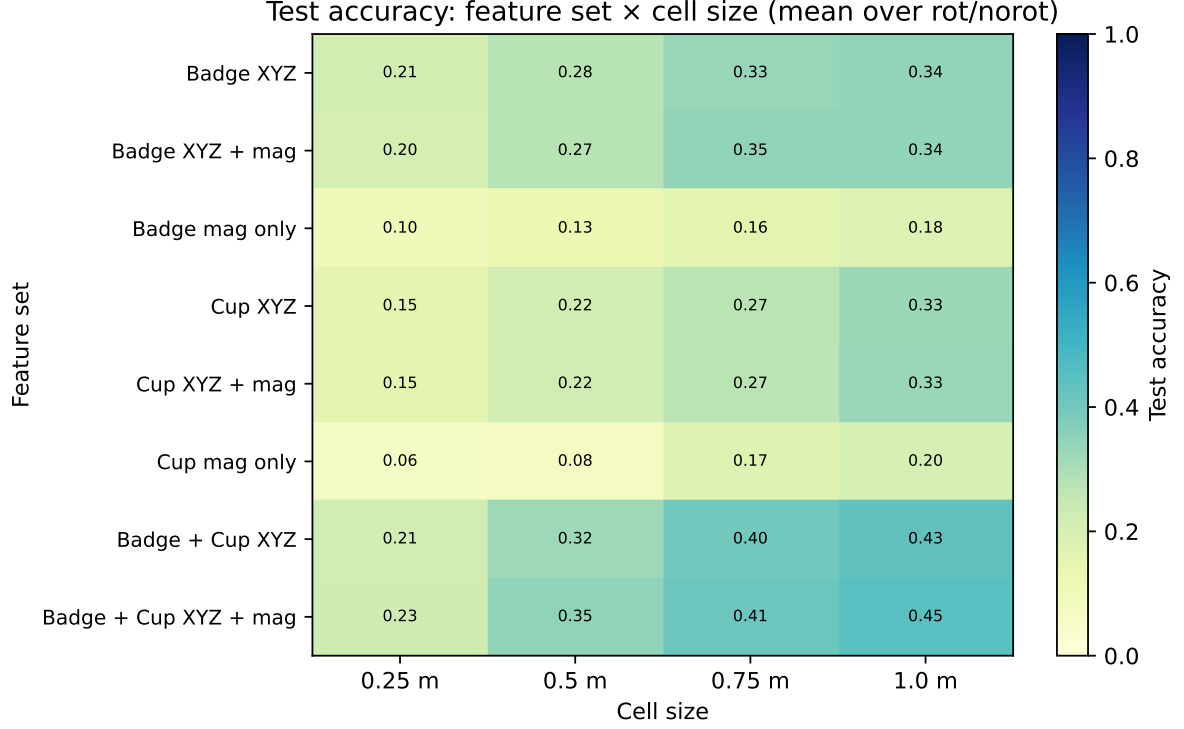


Figure 7.1: Test accuracy (exact zone match) across all feature subsets and cell sizes. Each cell shows the mean accuracy averaged over the rotated and unrotated variants.

Figure 7.1 shows the test accuracy across feature subsets and cell sizes. Here, *XYZ* denotes the three directional magnetometer axes, *mag* the scalar field magnitude, *badge* the chest-mounted sensor, and *cup* the drinking-cup sensor. The combined badge and cup configurations consistently outperform single-sensor variants across all four cell sizes. At 0.25 m, Badge + Cup XYZ + mag and Badge + Cup XYZ reach 23.9% and 22.2% respectively, compared to 20.6% for the best badge-only variant. At 0.5 m, the advantage widens: Badge + Cup XYZ + mag (unrotated) achieves 36.0% and Badge + Cup XYZ (unrotated) achieves 35.5%, both substantially outperforming the best single-sensor variant (Badge XYZ + mag, unrotated, 27.8%). At 0.75 m and 1.0 m the same pattern holds, with combined configurations reaching 41.7% and 45.9% respectively, against best single-sensor values of 36.9% and 36.1%. This confirms that the two sensors capture complementary spatial information, likely reflecting their different body-worn positions (chest badge versus drinking cup held at varying heights and orientations).

Badge-only configurations consistently outperform cup-only counterparts at finer resolutions, with the gap narrowing at coarser cell sizes. At 0.25 m, the best cup variant (Cup XYZ + mag) reaches 14.7%, roughly a third of the best badge result. At 0.5 m the gap shrinks (Cup XYZ + mag: 22.9% vs. Badge XYZ + mag: 27.8%), and at 1.0 m the best cup variant (Cup XYZ + mag, unrotated, 36.1%) marginally exceeds the best badge-only result (Badge XYZ, unrotated, 35.0%). The chest-mounted badge occupies a more stable position relative to the body than the cup, which is held and moved freely during social interaction. This positional stability likely results in more consistent magnetometer readings at fine spatial scales.

Magnitude-only features (Badge mag only and Cup mag only) perform poorly at fine resolutions (6.1%–11.6% at 0.25 m), improving somewhat at coarser cell sizes (up to 20.3% at 1.0 m) but consistently ranking last within each cell-size tier. While the field magnitude provides some spatial discrim-

inability, discarding the directional XYZ components removes substantial location-specific information. This supports the use of full three-axis measurements, as adopted by prior work [17, 6].

7.4. Effect of World-Frame Rotation

Across all 64 configurations, the unrotated (raw sensor-frame) variant achieves a marginally higher mean test accuracy than the rotated variant (26.1% vs. 24.6%). This difference is small and inconsistent across feature subsets and cell sizes, suggesting that DMP-based world-frame rotation does not provide a systematic benefit for zone classification in this setting.

This finding is somewhat counterintuitive. World-frame rotation is intended to remove orientation-dependent components of the magnetometer signal, leaving only the location-specific ambient field. In practice, the benefit may be limited because: (i) the LSTM has capacity to learn orientation invariance from temporal context directly; (ii) the DMP quaternion introduces its own noise and drift; and (iii) in social interaction settings, participants' orientations at a given location are naturally diverse across the session, providing the model with multi-orientation coverage that reduces the dependence on any single orientation.

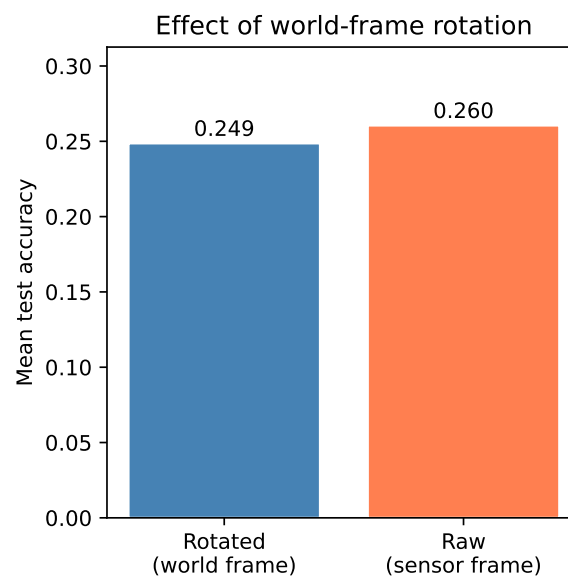


Figure 7.2: Mean test accuracy for world-frame rotated vs. raw sensor-frame configurations, averaged across all 32 completed feature and cell-size combinations.

7.5. Baseline Comparison: Locatify UWB

Each participant in the experiment was equipped with a Locatify UWB badge, which provides continuous 2D position estimates using infrastructure mounted in the room. The manufacturer specifies a nominal accuracy of up to 30 cm. To assess real-world accuracy in the experimental setting, the badge positions were validated against the camera-derived ground truth via reprojection, yielding a mean position error of approximately 50 cm across participants.

Table 7.2 places the magnetic fingerprinting results in context alongside the Locatify baseline. For a fair comparison, we report the mean physical error for each approach, converting the mean cell error of the best magnetic configuration at each cell size into metres.

Table 7.2: Comparison of the best magnetic fingerprinting configuration against the Locatify UWB baseline. Physical error for the magnetic system is computed as mean cell error $\times$ cell size.

System	Infrastructure	Mean position error	Within 1 cell
Locatify UWB badge	Yes (UWB anchors)	~50 cm	—
Magnetic (0.25 m, best)	No	~75 cm	44.9%
Magnetic (0.50 m, best)	No	~68 cm	71.1%
Magnetic (0.75 m, best)	No	~77 cm	81.4%
Magnetic (1.00 m, best)	No	~78 cm	86.0%

The Locatify system achieves a lower mean error, as expected from a dedicated UWB positioning system with known anchor positions. However, several contextual factors temper this comparison. First, the Locatify badges require fixed UWB infrastructure, which must be installed and calibrated per venue. Second, validation against the camera ground truth revealed systematic inconsistencies in the recorded positions, suggesting the real-world accuracy of the UWB system is lower than the manufacturer specification in this setting. Third, the magnetic fingerprinting system operates entirely infrastructure-free, requiring only the IMUs already worn by participants.

At the 0.25 m resolution, the best magnetic configuration achieves a mean error of approximately 75 cm — 25 cm worse than the Locatify baseline — while placing 44.9% of predictions within 25 cm and 66.4% within 50 cm. At the 0.5 m resolution, the best configuration (Badge + Cup XYZ, unrotated) achieves a mean error of approximately 68 cm, placing 71.1% of predictions within 50 cm and 86.7% within 1 m, closing to within the margin of the UWB baseline while operating entirely infrastructure-free. At 0.75 m, the best configuration (Badge + Cup XYZ + mag, unrotated) achieves a mean error of approximately 77 cm, with 81.4% of predictions within one cell (75 cm) and 85.4% within two cells (1.5 m). At 1.0 m, the best configuration (Badge + Cup XYZ, unrotated) achieves a mean error of approximately 78 cm, with 86.0% of predictions within one cell (1 m) and 91.2% within two cells (2 m). This suggests that for applications requiring coarse spatial awareness — such as detecting social proximity or identifying conversational groups — infrastructure-free magnetic fingerprinting offers a practically viable alternative to UWB-based systems across all evaluated resolutions.

7.6. Spatial Error Distribution

The mean cell error at 0.5 m ranges from 1.4 to 3.5 cells across configurations, corresponding to physical distances of approximately 0.7–1.75 m. At 0.25 m the range is 2.9–6.5 cells (0.72–1.62 m). At 0.75 m the range narrows to 1.03–2.15 cells (0.77–1.61 m), and at 1.0 m to 0.78–1.65 cells (0.78–1.65 m). The similarity in physical error magnitude across all four cell sizes reflects a consistent underlying spatial ambiguity in the magnetic fingerprint, regardless of how the space is discretised.

7.7. Hyperparameter Analysis

Optuna selected hyperparameters independently for each of the 64 configurations. Table 7.3 summarises the distributions of the most informative parameters across all completed trials.

The largest hidden size (512) was selected in 47% of configurations and 256 in a further 30%, indicating that higher model capacity consistently benefits the classification task. At 0.25 m there are 318 output classes, and even at 1.0 m there are 28 classes, so the model must learn a rich internal representation regardless of resolution. The 64-unit hidden size was selected in only 6% of cases, confirming that very small models are insufficient.

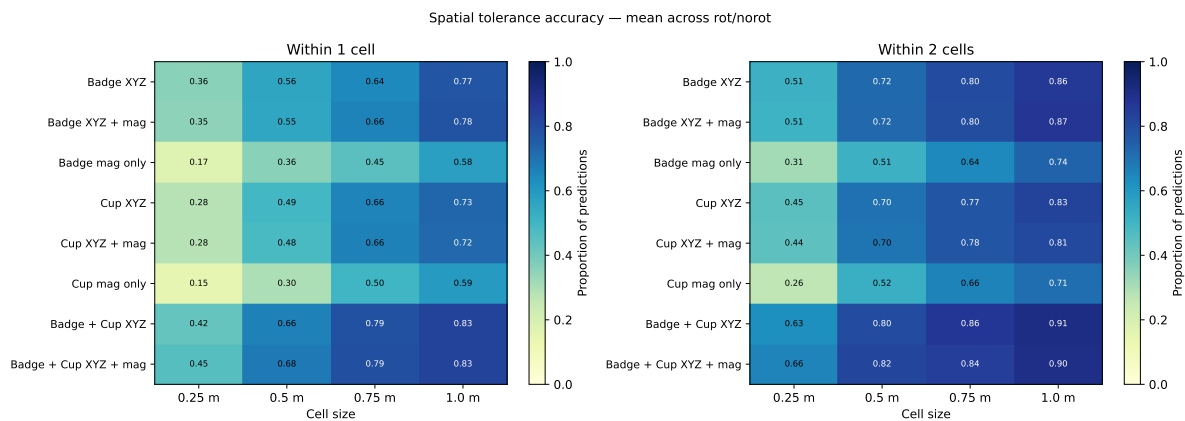


Figure 7.3: Mean proportion of predictions falling within one cell (left) and two cells (right) of the true zone, across feature subsets and cell sizes. Values are averaged over rot and norot variants.

Table 7.3: Distribution of best hyperparameters selected by Optuna across all 64 configurations. Search ranges are given in parentheses.

Parameter	Search space	Selected distribution
Hidden size	{64, 128, 256, 512}	512: 47%, 256: 30%, 128: 17%, 64: 6%
LSTM layers	1–3	1: 41%, 2: 41%, 3: 19%
Window length (s)	2–10	2 s: 38%, 3–4 s: 36%, ≥ 5 s: 27%
Batch size	{32, 64, 128}	128: 50%, 32: 31%, 64: 19%
Dropout	0.1–0.5	mean 0.31, range 0.10–0.49
Learning rate	10^{-4} – 10^{-2}	mean 2.9×10^{-4} , max 1.2×10^{-3}
Stride (s)	0.5–2.0	mean 1.23
Label threshold	0.5–0.95	mean 0.74, range 0.50–0.95

Single- and two-layer LSTMs were equally preferred (41% each), with three layers selected in only 19% of configurations. At finer resolutions (0.25 m), two-layer models were more common, whereas at coarser resolutions (0.75 m and 1.0 m) single-layer models dominated, consistent with the view that additional recurrent depth is more beneficial when the discrimination task is harder.

The 2-second window was selected most frequently overall (38% of configurations), and was dominant at 0.25 m (11 of 16 configurations). At coarser resolutions, longer windows of 3–5 s were more prevalent, suggesting that coarser zones benefit from more temporal context to accumulate discriminative signal. Windows of 8 s or longer were rare (<8%), indicating that very long sequences introduce noise that outweighs additional coverage. Learning rates were consistently low (mean 2.9×10^{-4} ; no configuration exceeded 1.2×10^{-3} despite the search range extending to 10^{-2}), and larger batch sizes (128) were preferred in half of all configurations.

7.8. Effect of Zone Merging

Table 7.4 compares the plain-grid and merged-zone results across all cell sizes.

Zone merging does not produce a consistent improvement over the plain grid partition. At fine resolutions (0.25 m and 0.50 m), the merged variant underperforms: at 0.25 m the best configuration drops from 23.9% to 22.2%, and at 0.50 m from 36.0% to 33.1%. The within-cell tolerance metrics follow the same trend, with the merged variant producing lower within-1-cell and within-2-cells rates at both fine resolutions. At 0.75 m and 1.0 m the differences are negligible: the merged best configuration marginally exceeds the plain-grid best (42.1% vs. 41.7% at 0.75 m and 46.9% vs. 45.9% at 1.0 m), but mean accuracy across all 16 configurations remains essentially the same.

Notably, the identity of the best-performing feature set shifts at 0.25 m: the plain grid favours Badge + Cup XYZ + mag (dual-sensor combined), whereas the merged grid favours Badge XYZ + mag (badge only). This suggests that merging alters the spatial boundary structure in ways that interact differently with different feature combinations, rather than providing a neutral redistribution of training samples.

Table 7.4: Test accuracy and spatial tolerance for plain-grid versus merged-zone configurations.

Cell size	Variant	Test accuracy		Within 1 cell		Within 2 cells	
		mean	max	mean	max	mean	max
0.25 m	Plain	0.160	0.239	0.304	0.450	0.466	0.664
	Merged	0.159	0.222	0.262	0.407	0.455	0.661
0.50 m	Plain	0.230	0.360	0.511	0.710	0.690	0.867
	Merged	0.222	0.331	0.459	0.647	0.668	0.834
0.75 m	Plain	0.299	0.417	0.649	0.814	0.771	0.856
	Merged	0.304	0.421	0.628	0.804	0.733	0.846
1.00 m	Plain	0.325	0.459	0.735	0.860	0.831	0.929
	Merged	0.329	0.469	0.714	0.834	0.825	0.921

The absence of a clear benefit from zone merging is consistent with the observation that the dominant source of localization error is magnetic ambiguity rather than data imbalance. Sparse zones are intrinsically difficult to localise because they are visited infrequently and their magnetic signatures may closely resemble neighbouring zones; merging them into larger groups reduces the number of classes but does not resolve the underlying ambiguity. At fine resolutions, merging additionally reduces spatial discriminability by conflating zones that may have had distinct magnetic profiles.

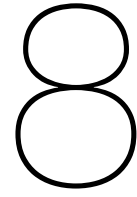
7.9. Hexagonal Grid

Table 7.5 summarises performance of the hexagonal grid partition for the two complete cell sizes (0.25 m and 0.5 m). Spatial tolerance is measured in hex-neighbour steps (hex distance), the natural analogue of Chebyshev distance for the hexagonal lattice.

Table 7.5: Hexagonal grid results (mean and max across 16 configurations per cell size). Spatial tolerance is in hex-distance steps.

Cell size (active zones)	Test accuracy		Within 1 hex		Within 2 hexes	
	mean	max	mean	max	mean	max
0.25 m (72 zones)	0.186	0.282	0.394	0.579	0.590	0.816
0.50 m (28 zones)	0.314	0.460	0.635	0.796	0.764	0.864

At both circumradii the best-performing feature set is Badge + Cup XYZ + mag, consistent with the square grid findings. Because the circumradius parameter yields cells substantially larger than a square of the same nominal size, the zone counts are most directly comparable with coarser square grids. At circumradius 0.5 m the hexagonal partition produces 28 active zones — identical to the 1.0 m square grid — and achieves essentially the same classification performance: the best configuration reaches 46.0% exact accuracy versus 45.9% for the best 1.0 m square configuration, with 79.6% of predictions within one hex step. At circumradius 0.25 m the hexagonal partition yields 72 active zones, falling between the 50 zones of the 0.75 m square grid and the 96 zones of the 0.5 m square grid. The best configuration achieves 28.2% accuracy, compared to 41.7% for the best 0.75 m square configuration at a similar zone count. Taken together, the hexagonal topology does not confer a consistent accuracy advantage over the square grid at the spatial scales evaluated; classification difficulty appears to be governed primarily by the number of zones and their magnetic discriminability rather than by cell geometry.



Discussion

8.1. Localization Feasibility

The results demonstrate that magnetic fingerprinting using body-worn magnetometers is feasible for zone-level localization in a socially dynamic indoor setting, even without motion models, step detection, or infrastructure. The LSTM model, trained solely on windowed magnetometer sequences, consistently learns to associate temporal signal patterns with spatial zones at levels far exceeding random chance. This supports the core premise of the approach and partially answers the first research question: magnetic fingerprinting alone can provide meaningful spatial information in socially unconstrained environments.

That said, exact zone accuracy remains modest, particularly at fine spatial resolution. The low absolute accuracy at 0.25 m (best 23.9%) reflects the difficulty of disambiguating 318 zones in a small room where adjacent zones may exhibit highly similar magnetic signatures. Across all four cell sizes, exact accuracy ranges from 23.9% (0.25 m) to 45.9% (1.0 m), confirming the expected monotonic relationship between cell size and classification ease. The within-cell tolerance metrics reveal that the model's spatial errors are generally small in physical terms: at 0.5 m resolution, over 71% of predictions fall within 50 cm of the true location, rising to 86% at 1.0 m. This level of precision may be sufficient for coarse-grained social proximity analysis, even if precise zone assignment remains unreliable at finer resolutions.

8.2. Challenges Specific to Social Settings

Several characteristics of the social interaction setting contribute to the observed difficulty. First, participants frequently remain stationary during conversations, limiting the temporal diversity of magnetic sequences observed at each location. An LSTM window of 2–10 s captures only local temporal context, which offers little discriminability when the participant has not moved. Second, the session lasts 30 minutes with 32 participants, of whom only the Mingling 1 session is used for training. This provides a relatively small training set relative to the number of zones, especially at fine resolutions. Third, the 80/20 temporal split means the test set covers the final portion of the session, during which participants may occupy different zones from those seen frequently in training, introducing a form of distribution shift.

8.3. Sensor Placement and Feature Design

The consistent advantage of the badge over the cup sensor suggests that sensor placement is a critical design decision in body-worn magnetic localization. The chest-mounted badge maintains a relatively fixed orientation relative to the body during normal interaction, while the cup is subject to arbitrary handling gestures. Future work could explore sensor fusion with explicit orientation filtering (as proposed in [10]), or dedicated calibration of the cup sensor to reduce handling-induced variability.

The marginal performance difference between Badge XYZ and Badge XYZ + mag (adding field magnitude) suggests that the magnitude channel provides a small but consistent benefit in some configurations, while not uniformly improving performance across all conditions. This is consistent with

prior work showing that field magnitude alone is insufficient for localization but contributes positively when combined with directional components [2].

8.4. Limitations and Future Work

All 64 configurations have been evaluated, providing a complete picture of the accuracy-resolution trade-off. The results confirm that coarser grids yield higher exact accuracy while the mean physical error remains broadly stable across cell sizes (0.68–0.78 m for the best configuration at each resolution), suggesting that the dominant source of error is a consistent spatial ambiguity in the magnetic fingerprint rather than a discretisation artefact.

A further limitation is that the evaluation is conducted within a single session (Mingling 1), with temporal splitting for train/test separation. Cross-session generalization — applying a model trained on Mingling 1 to Mingling 2 data — would provide a stronger test of the fingerprint’s temporal stability and is a natural direction for future evaluation.

The evaluation is also limited to a plain LSTM architecture. A CNN-LSTM hybrid [5] was not evaluated; while the 1D-CNN denoising stage could in principle help with noisy body-worn magnetic sequences, the stationarity characteristic of social interaction reduces the local temporal structure that such a stage is designed to capture. Nonetheless, evaluating a CNN-LSTM variant—particularly for configurations with longer windows where noise accumulation is greater—remains a natural direction for future work.

Finally, the hyperparameter search was limited to 20 Optuna trials per configuration due to computational constraints. A larger search budget or Bayesian warm-starting across related configurations may yield further performance improvements, particularly for configurations with larger input dimensionality where the LSTM has more parameters to optimise.

Bibliography

- [1] I. Ashraf, S. Hur, and Y. Park. "Application of deep convolutional neural networks and smart-phone sensors for indoor localization". In: *Applied Sciences (Switzerland)* 9.11 (2019). DOI: 10.3390/app9112337. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067253502&doi=10.3390%2Fapp9112337&partnerID=40&md5=0e65e2d46f7387f4e6d1cfd87f9a4c5b>.
- [2] Biplav Bhattarai et al. "Geomagnetic field based indoor landmark classification using deep learning". In: *IEEE Access* 7 (2019), pp. 33943–33956.
- [3] Jaehwan Chung, Mark Donahoe, and Chris Schmandt. "Indoor location sensing using geomagnetism". In: *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*. 2011.
- [4] Delft AI Cluster (DAIC). *The Delft AI Cluster (DAIC)*, RRID:SCR_025091. 2024. DOI: 10.4233/rrid:scr_025091. URL: https://doi.org/10.4233/rrid:scr_025091.
- [5] X. Ding, M. Zhu, and B. Xiao. "Accurate Indoor Localization Using Magnetic Sequence Fingerprints with Deep Learning". In: vol. 13155 LNCS. 2022, pp. 65–84. DOI: 10.1007/978-3-030-95384-3_5. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126217697&doi=10.1007%2F978-3-030-95384-3_5&partnerID=40&md5=b6bebf9956b138f50acfe06a63473ddf.
- [6] H.J. Jang, J.M. Shin, and L. Choi. "Geomagnetic field based indoor localization using recurrent neural networks". In: vol. 2018-January. 2017, pp. 1–6. DOI: 10.1109/GLOCOM.2017.8254556. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85046428600&doi=10.1109%2FGLOCOM.2017.8254556&partnerID=40&md5=066fafb969c8fe73ef03f2>.
- [7] H.J. Jun Bae and L. Choi. "Large-Scale Indoor Positioning using Geomagnetic Field with Deep Neural Networks". In: vol. 2019-May. 2019. DOI: 10.1109/ICC.2019.8761118. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85070235257&doi=10.1109%2FICC.2019.8761118&partnerID=40&md5=2d74a4a047dd0401658c20992b34ed32>.
- [8] K. Klipp et al. "Comparing CNN and LSTM Networks for Magnetic Localization of IoT Devices and Pedestrian Tracking". In: 2025, pp. 549–555. DOI: 10.1109/ICCAI66501.2025.00088. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105015847792&doi=10.1109%2FICCAI66501.2025.00088&partnerID=40&md5=b5bcbdd4ed3f98db87c6109195af9f4a7>.
- [9] Manon Kok, Jeroen D. Hol, and Thomas B. Schön. "Using Inertial Sensors for Position and Orientation Estimation". In: *Found. Trends Signal Process.* 11.1–2 (Nov. 2017), pp. 1–153. ISSN: 1932-8346. DOI: 10.1561/20000000094. URL: <https://doi-org.tudelft.idm.oclc.org/10.1561/20000000094>.
- [10] Manon Kok and Thomas B. Schön. "Magnetometer Calibration Using Inertial Sensors". In: *IEEE Sensors Journal* 16.14 (2016), pp. 5679–5689. DOI: 10.1109/JSEN.2016.2569160.
- [11] Manon Kok and Arno Solin. "Scalable Magnetic Field SLAM in 3D Using Gaussian Process Maps". In: *2018 21st International Conference on Information Fusion (FUSION)*. 2018, pp. 1353–1360. DOI: 10.23919/ICIF.2018.8455789.
- [12] X. Liu et al. "Robust Pedestrian Dead Reckoning Integrating Magnetic Field Signals and Digital Terrestrial Multimedia Broadcasting Signals". In: *Remote Sensing* 15.13 (2023). DOI: 10.3390/rs15133229. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85165173641&doi=10.3390%2Fr15133229&partnerID=40&md5=f08a5a97df8e2687014b8426bc7e08f>.
- [13] X. Liu et al. "Singular Value Decomposition Based Indoor Localization Using Small Scale Crowd Sensing Data". In: *IEEE Transactions on Mobile Computing (2025)*. DOI: 10.1109/TMC.2025.3610882. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105016817095&doi=10.1109%2FTMC.2025.3610882&partnerID=40&md5=db6e62f97ae05194429fd71b86bb60aa>.

- [14] Miroslav Opiela et al. "Building Parts Classification using Neural Network". In: *Proceedings of the Work-in-Progress Papers at the 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN-WiP 2023)*. Vol. 3581. CEUR-WS.org. Nuremberg, Germany: CEUR Workshop Proceedings, 2023, pp. —. URL: https://ceur-ws.org/Vol-3581/164_WiP.pdf.
- [15] Guanglie Ouyang and Karim Abed-Meraim. "A Survey of Magnetic-Field-Based Indoor Localization". In: *Electronics* 11.6 (2022). ISSN: 2079-9292. DOI: 10.3390/electronics11060864. URL: <https://www.mdpi.com/2079-9292/11/6/864>.
- [16] F. Parralejo et al. "Neural Network Evaluation for Magnetic Fingerprinting Positioning". In: 2025. DOI: 10.1109/ICCT-Europe63283.2025.11157679. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105019048315&doi=10.1109%2FICCT-Europe63283.2025.11157679&partnerID=40&md5=192c3d24ade204d4cce91ae20f9e18c0>.
- [17] B. Platte et al. "Simplified Indoor Localization Data Acquisition by Use of Recurrent LSTM Networks on Sequential Geomagnetic Vectors". In: vol. 1152 AISC. 2020, pp. 115–121. DOI: 10.1007/978-3-030-44267-5_17. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085166451&doi=10.1007%2F978-3-030-44267-5_17&partnerID=40&md5=343803aa6512a423ea80384184cb9b57.
- [18] Benny Platte, Maximilian Eibl, and Marc Ritter. "Geomagnetic Localization with Neural Networks: A Comparative Survey". In: *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 2025.
- [19] K Y Qiu, H Huang, and A El-Rabbany. "Geomagnetic field-based indoor positioning using back-propagation neural networks". In: *ISPRS Archives*. 2020.
- [20] William Storms, Jason Shockley, and John Raquet. "Magnetic field navigation in an indoor environment". In: *Ubiquitous Positioning Indoor Navigation and Location Based Service*. 2010.
- [21] C. Sun et al. "Indoor Localization Based on Integration of Wi-Fi with Geomagnetic and Light Sensors on an Android Device Using a DFF Network". In: *Electronics (Switzerland)* 12.24 (2023). DOI: 10.3390/electronics12245032. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85180218493&doi=10.3390%2Felectronics12245032&partnerID=40&md5=31bf0309f12da5c9c546e78d015ff9b5>.
- [22] Jian Wang et al. "An Improved PDR/Magnetometer/Floor Map Integration Algorithm for Ubiquitous Positioning Using the Adaptive Unscented Kalman Filter". In: *ISPRS International Journal of Geo-Information* 4.4 (2015), pp. 2638–2659. ISSN: 2220-9964. DOI: 10.3390/ijgi4042638. URL: <https://www.mdpi.com/2220-9964/4/4/2638>.
- [23] H. Wu, J. Tan, and S.-H.G. Chan. "Pedometer-free Geomagnetic Fingerprinting with Casual Walking Speed". In: *ACM Transactions on Sensor Networks* 18.1 (2021). DOI: 10.1145/3470850. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85145589740&doi=10.1145%2F3470850&partnerID=40&md5=15c9fd74758f85d567071afa5e54f5>.
- [24] Yufei Xu et al. "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation". In: *Advances in Neural Information Processing Systems*. 2022.
- [25] Shuo Yan et al. "Deep learning-based geomagnetic navigation method integrated with dead reckoning". In: *Remote Sensing* 15.17 (2023).
- [26] Meng Zhang et al. "Real-time indoor localization using smartphone magnetic with LSTM networks". In: *Neural Computing and Applications* 33 (2021), pp. 10093–10110.
- [27] M. Zmitri, H. Fourati, and C. Prieur. "BiLSTM Network-Based Extended Kalman Filter for Magnetic Field Gradient Aided Indoor Navigation". In: *IEEE Sensors Journal* 22.6 (2022), pp. 4781–4789. DOI: 10.1109/JSEN.2021.3091862. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85112447036&doi=10.1109%2FJSEN.2021.3091862&partnerID=40&md5=f2e9b98ec50611f75a016597b4253aef>.