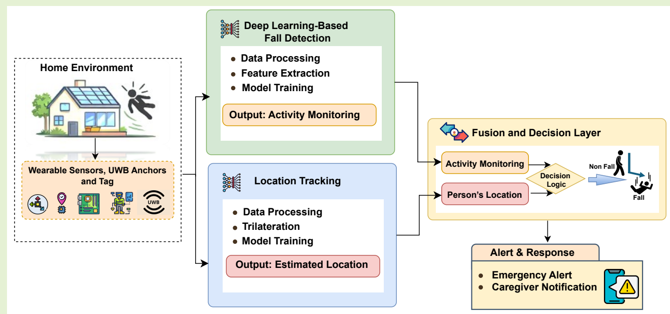
# Position-Aware Indoor Human Activity Recognition and Fall Detection

Muhammad Moazam Shahid, Pedro Machado, Jordan J. Bird, Salisu Wada Yahaya, Sozo Inoue, Ahmad Lotfi and Isibor Kennedy Ihianle

***Abstract*— With increasing life expectancy, particularly in developed nations, the proportion of elderly individuals is rising rapidly, necessitating advanced systems for continuous monitoring and timely intervention to support independent living and enhance safety in assisted care environments. Falls are among the leading causes of hospitalisations and deaths related to injuries in this demographic, highlighting the urgent need for intelligent fall detection systems. However, most existing solutions struggle with real-world deployment due to incomplete anomaly modelling and a lack of contextual location awareness. This paper introduces a novel position-aware indoor activity recognition and fall detection approach that uses spatial and motion data to detect falls with high accuracy and contextual relevance. The system integrates Ultra-Wideband (UWB) positioning technology with a Multilayer Perceptron (MLP) model to achieve indoor localisation. Furthermore, accelerometer and gyroscope data are used for activity monitoring, which is processed using a hybrid deep learning architecture that combines a Variational Autoencoder (VAE), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks. This architecture takes advantage of temporal and spatial feature extraction for improved fall detection. The localisation module achieves over 96% accuracy. For activity recognition, the VAE CNN-LSTM model achieving fall detection accuracy exceeding 97%. A late fusion decision layer combines spatial and activity-level insights to enable precise detection and localisation of fall events within indoor environments. The proposed system is validated in a real-world smart home setting and demonstrates strong performance in terms of accuracy, scalability, and adaptability.**




***Index Terms*— Fall Detection, Indoor Localisation, Ultra-Wideband, Machine Learning, Variational Autoencoder, CNN-LSTM, Late Fusion Decision Layer**

## I. Introduction

Global life expectancy has risen steadily in recent decades, leading to a significant increase in the elderly population, particularly in developed countries where many individuals live beyond 60 years [1]–[4]. Projections indicate that by 2030, one in six people worldwide will be 60 years or older [1], [5], [6]. This demographic change poses considerable challenges for healthcare and care infrastructures, as a growing shortage of caregivers leaves many elderly individuals without adequate physical and emotional support [7]–[9].

Older adults, especially those living alone, face significant health and safety risks, with falls being a primary concern [10]–[12]. Falls are a prevalent and serious threat, often occurring during routine activities and causing injuries such as fractures and head trauma [13], [14]. These injuries often require hospitalisation and extended recovery, severely affecting quality of life [11]–[14]. Alarmingly, fall-related mortality rates are increasing, with around one in four older adults (approximately 14 million people) experiencing a fall annually [13], [14]. About 20% of these falls result in serious injuries, placing a substantial burden on healthcare systems [14], [15]. Between 1999 and 2020, falls were the leading cause of injury-related deaths among adults 65 and older, accounting for more than 478,000 deaths [16], [17].

To mitigate such risks, several fall detection systems have been developed to enable prompt intervention and support independent living. Among these, sensor-based wearable technologies have shown considerable promise [18]–[20]. Devices such as wristbands and smart belts equipped with accelerometers and gyroscopes can monitor physical activity and detect sudden deviations indicative of a fall. The integration of multiple sensing modalities, including barometric, pressure, and motion sensors, further improves the accuracy and reliability of detection [19], [21], [22]. In addition, non-intrusive ambient sensor systems have been proposed, which use mo-

Muhammad Moazam Shahid (muhammad.shahid2022@my.ntu.ac.uk), Pedro Machado (pedro.machado@ntu.ac.uk), Jordan J. Bird (jordan.bird@ntu.ac.uk), Salisu Wada Yahaya (salisu.yahaya@ntu.ac.uk), Ahmad Lotfi (ahmad.lotfi@ntu.ac.uk), Isibor Kennedy Ihianle (isibor.ihianle@ntu.ac.uk) are with the Department of Computer Science, Nottingham Trent University, UK.; Sozo Inoue (sozo@brain.kyutech.ac.jp) is with the Kyushu Institute of Technology, Japan.

tion, acoustic, and vibration data to identify falls within a home setting [23], [24]. Despite their potential, environmental sensors often fail in terms of accuracy, cost-efficiency, and ease of deployment compared to wearable solutions [25].

Traditional fall detection methods rely on threshold-based algorithms to interpret sensor signals. Although simple to implement, these techniques often lack the flexibility and adaptability required in dynamic real-world environments. Recent advances have introduced Machine Learning (ML) and Deep Learning (DL) approaches, significantly improving classification accuracy by modelling complex activity patterns. ML-based models typically depend on manual feature extraction and are implemented using classifiers such as Random Forest (RF), Logistic Regression (LR), and Decision Trees (DT) [22], [26], [27]. However, they require substantial domain knowledge for effective feature selection. In contrast, DL models including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) can automatically learn hierarchical representations directly from raw sensor data, enhancing robustness and scalability [28], [29]. Hybrid models that combine different architectures, such as CNN-GRU [30], LSTM-CNN [28], and CBLSTM [31], have been particularly effective in enhancing performance through the fusion of temporal and spatial features.

Vision-based fall detection systems, which use cameras to infer posture and motion, have also shown promising performance [32]–[34]. However, these systems face significant challenges, including privacy concerns, high computational requirements, and reduced reliability under varying lighting and occlusion conditions. Furthermore, many existing fall detection systems are trained on datasets collected in controlled laboratory environments. These conditions do not reflect the variability and complexity of real-world scenarios, thereby limiting the generalisation of the model. In addition, real fall events are inherently rare and difficult to capture, contributing to data scarcity and class imbalance problems in training.

A major limitation in existing fall detection solutions is the lack of contextual awareness, specifically, knowledge of where a fall occurs. This spatial information is critical for enabling timely and effective intervention, particularly for individuals living independently or in assisted living. To address this gap, we propose a position-aware fall detection approach that integrates indoor localisation using Ultra-Wideband (UWB) technology with anomaly-based activity recognition. The proposed approach integrates real-time spatial tracking with activity monitoring to enhance anomaly detection. It uses UWB-based Time of Arrival (TOA) modelling and a Multilayer Perceptron (MLP) classifier to perform precise, real-time localisation. Beacons placed at fixed locations communicate with wearable tags to continuously track the user's location within the home. Concurrently, a CNN-LSTM-based Variational Autoencoder (VAE) model processes accelerometer and gyroscope signals to detect falls with high accuracy. A late fusion decision layer then combines outputs from both the localisation and fall detection modules, enabling the system to report not only the occurrence but also the precise location of a fall. This multimodal integration significantly enhances reliability and facilitates actionable real-time monitoring in indoor environments. The contributions of this paper are threefold.

- We present a novel hybrid framework that tightly integrates UWB-based indoor localisation with deep learning-based fall detection.
- We introduce and validate a custom real-world dataset consisting of daily activities and staged falls, improving generalisability in practical settings.
- We implement a late fusion decision layer that enhances accuracy by combining spatial and activity-level predictions, enabling real-time, context-aware monitoring.

The remainder of this paper is structured as follows: Section II reviews relevant literature; Section III describes the proposed methodology; Section IV presents experimental results; Section V, discussion and Section VI outlines future work and concludes the paper.

## II. Related Work

### Indoor Localisation with Ultra-Wideband (UWB)

UWB technology has gained significant traction for indoor localisation due to its high temporal resolution and robustness against multipath interference [35]–[37]. Traditional UWB localisation methods typically use signal characteristics like TOA, Time Difference of Arrival (TDoA), Time of Flight (ToF), or Angle of Arrival (AoA) [38]–[40]. While effective in controlled settings, these methods often perform poorly in Non-Line-of-Sight (NLOS) conditions where signal obstruction and multipath propagation are common [40].

To improve localisation performance, numerous studies have explored machine learning (ML)-based models that exploit statistical patterns within UWB Channel Impulse Response (CIR) data. For example, [41] introduced a grid search-optimised support vector machine (SVM) combined with Principal Component Analysis (PCA) for CIR-based localisation. This approach demonstrated superior performance compared to traditional k-Nearest Neighbour (k-NN) and Backpropagation Neural Network (BPNN) methods by reducing dimensionality and enhancing kernel selection. Similarly, an SVM-based localisation classifier in [42] achieved approximately 92% precision, proving to be adaptable to multipath interference in nonlinear sight (NLOS) settings. Further refining this, a two-stage SVM framework proposed in [43] reached 93.7% accuracy in distinguishing between Line-of-Sight (LOS) and NLOS conditions.

Despite these advancements, SVM-based approaches are inherently sensitive to kernel choice and hyperparameter settings. This sensitivity limits their adaptability to dynamic environments, and grid-based tuning significantly increases computational overhead. To mitigate these issues, regression-based models have been investigated. For example, a logistic regression model introduced in [44] for TDoA topologies effectively suppressed NLOS-induced errors by up to 80%. However, these classical ML approaches remain constrained by the high variability of indoor radio environments and often struggle with generalisation across diverse room geometries and materials.

More recently, deep learning models have emerged as a promising alternative for UWB localisation, offering end-to-end learning directly from raw CIR data. For example, [45] proposed a CNN that extracts TOA features directly from CIR profiles, achieving a mean absolute localisation error of 17.3 cm in NLOS environments. However, such architectures often introduce challenges related to overfitting, interpretability, and high computational load, making them impractical for real-time deployment on resource-constrained devices. A recent study [46] further emphasises that most DL models struggle to generalise without extensive environmental-specific training.

Although a wide range of models have explored UWB localisation using signal processing, machine learning, and deep learning techniques, key challenges remain in achieving a consistent trade-off between accuracy, computational efficiency, and robustness in diverse real-world settings. To extend and build on these, we introduce a position-aware approach integrating environmental context into the localisation process while maintaining a lightweight architecture suitable for deployment on resource-constrained platforms.

In addition to UWB-based techniques, several studies have explored multimodal sensing for indoor localisation, particularly within ambient assisted living environments. [47] demonstrated that combining BLE beacons, inertial sensing, and behavioural interaction data can improve zone-level localisation accuracy in activity-rich settings. Similarly, [48] provides a standardised benchmark for evaluating diverse localisation technologies, including BLE, Wi-Fi, and hybrid systems, under realistic conditions. These works highlight the strengths of multimodal approaches for broad-coverage indoor positioning. However, many such systems prioritise coverage over precision, typically achieving room or zone-level granularity. In contrast, our study focuses on UWB and inertial measurements because these modalities deliver the high temporal resolution and centimetre-level accuracy required for fine-grained fall localisation in home environments, thus positioning our work within the class of precision-oriented localisation systems.

### *Fall Detection: Challenges and Approaches*

The development of reliable fall detection systems is limited by several recurring challenges. One of the most significant issues is the lack of representative datasets. Fall events, especially among elderly populations, are infrequent, unpredictable, and ethically complex to simulate [49], [50]. Consequently, many studies rely on publicly available datasets, such as SisFall [28], which are typically collected in controlled laboratory settings with young participants performing scripted falls. Although these data sets facilitate benchmarking, they fail to capture the variability, spontaneity, and subtlety of real-world fall events.

Another significant challenge in developing fall detection models is data imbalance. Falls occur much less frequently than Activities of Daily Living (ADL), which creates highly imbalanced datasets that can bias model training. Although data balancing techniques such as oversampling and undersampling [51] are commonly used to address class-level imbalance, they often overlook intraclass variability. For example, in the SisFall dataset, which includes 15 distinct fall types and 19 ADL categories, the number of samples between subclasses is unevenly distributed, introducing further bias and potentially degrading model performance. Furthermore, many of these data sets lack essential contextual metadata such as the age of the participant, physical condition, and environmental characteristics, all of which are needed to build generalisable fall detection systems.

To address the limitations associated with fall data scarcity, researchers have increasingly turned to unsupervised and semi-supervised learning approaches. VAEs and other autoencoder-based anomaly detection approaches have demonstrated considerable potential to model normal patterns of ADL) and identifying deviations indicative of falls without requiring large amounts of annotated fall data [52], [53]. These methods are particularly well-suited to real-world deployments, where the collection of authentic fall events is both ethically and practically constrained. However, the effectiveness of such models depends heavily on careful architectural design and training strategies to balance sensitivity and specificity, especially in noisy or heterogeneous sensor environments.

Beyond data challenges, the deployment of fall detection systems on embedded platforms presents unique hurdles. Wearable and edge devices have strict limitations on memory, power, and processing capabilities. For example, the EmotionNet model [54], despite being designed for embedded deployment, requires 761 MB of storage for a five-layer CNN. Such requirements are prohibitive for continuous monitoring systems intended for older adults or low-power environments. Efficient architectures such as pruning, quantised models, or hybrid CNN-LSTM approaches [28], [29] are necessary to ensure real-time responsiveness without compromising accuracy. Given these constraints, recent studies have explored hybrid approaches that combine convolutional layers for spatial feature extraction, recurrent layers for temporal modelling, and attention mechanisms for context awareness [30], [31]. However, these models are still largely evaluated on synthetic datasets. Our work contributes a more holistic approach by combining localisation data with fall detection to improve situational awareness.

In contrast to previous studies, we propose a position-aware fall detection approach that integrates a lightweight UWB-based localisation module with a deep anomaly detection pipeline using CNNs, LSTMs, and VAEs. Our model is optimised for edge deployment and trained on a custom dataset collected in a realistic indoor environment with carefully balanced fall and ADL data. This approach not only addresses the limitations of existing datasets and models, but also enhances robustness and contextual awareness, key requirements for practical ambient-assisted living systems.

## III. Proposed Approach

Falls are abnormal activities that deviate significantly from the routine patterns of daily activities. Detecting these anomalies requires models capable of learning normal behavioural patterns and identifying deviations that indicate fall events. Since the occurrence of such events is often closely tied to

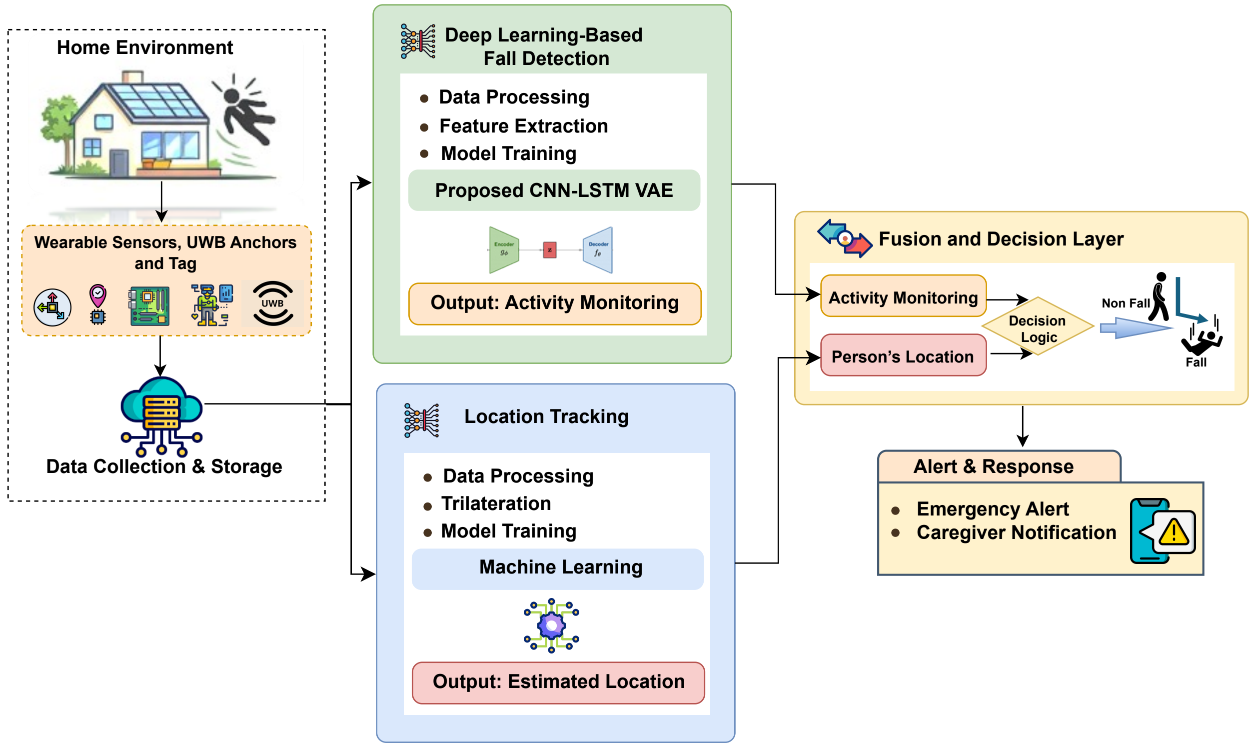



Fig. 1. Overview of the proposed position-aware fall detection approach, which integrates sensor data from a home environment with cloud-based storage and processing to combine indoor localisation and deep learning-based VAE-based fall detection, with results fused at a late fusion decision layer to accurately identify and localise fall events.

their spatial context, the proposed approach integrates indoor positioning with fall detection, extracting both location and motion features from multiple sensors. The proposed approach achieves fall detection and precise localisation, enabling real-time, context-aware monitoring by fusing these features. Figure 1 illustrates the architecture of the proposed approach, which integrates a late fusion layer to combine spatial and activity data, allowing the model to communicate fall events alongside their spatial locations within an indoor environment.

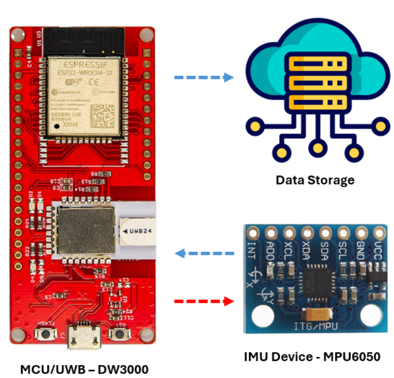



Fig. 2. Hardware setup consisting of an MCU with integrated UWB (DW3000) for communication, interfaced with an IMU device (MPU6050) for motion sensing, and connected to cloud-based data storage.

The architecture consists of two modules: one for indoor localisation and the other for fall detection, both of which operate independently before combining their outputs at a late fusion decision layer. The indoor location module utilises UWB signals for accurate positioning. A tag attached to an individual transmits a signal to beacons, and the response time of these signals is processed using trilateration and Time of Flight (TOF) techniques to estimate the precise location of the individual within the environment. The hardware setup supporting this system, shown in Figure 2, includes an MCU with UWB capabilities and an IMU sensor to collect real-time motion data. For fall detection, data collected from the gyroscope and accelerometer sensors are analysed to extract motion features of activities. A VAE-based CNN-LSTM model is utilised to reconstruct normal activities, while fall events produce significant reconstruction errors due to their anomalous nature. The late fusion layer integrates the spatial and motion data into a unified and cohesive system to detect not only the occurrence of a fall but also the precise location where it took place. This multimodal fusion enhances the system's reliability and provides actionable insights for real-time monitoring. The subsequent sections detail the components of the proposed approach.

## A. Indoor Location Tracking Module

To effectively detect a person's location within a home environment, the Indoor Location Tracking Module utilises UWB. Unlike traditional indoor positioning techniques such as Bluetooth and ZigBee, which suffer from limited accuracy, interference issues, and vulnerability to multipath effects, UWB technology offers precision and reliability [35], [55]. The ability of UWB to transmit short pulses with high bandwidth minimises multipath errors and enables accurate distance estimation in indoor spaces [36], [37]. This makes it ideal for real-time positioning in complex environments such as homes.

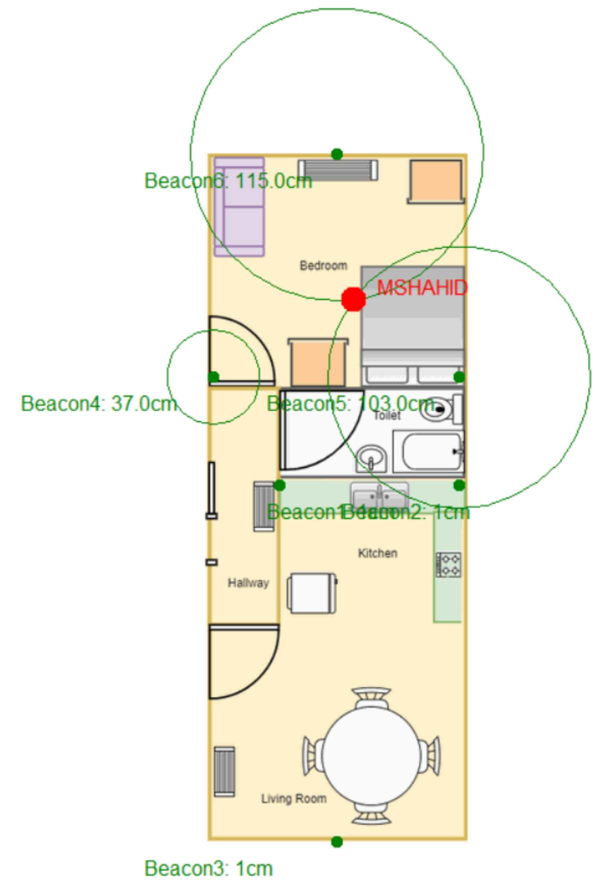



Fig. 3. Floor plan of the experimental environment with beacons (green dots) in fixed locations and a tag (red dot) emitting pulses for localisation in the living room and bedroom.

Localisation is achieved by strategically deploying multiple beacons within the home environment. A tag attached to the individual communicates with these beacons using UWB signals. The TOA of these signals is used to estimate the distances between the tag and each beacon. These distance measurements are critical for determining precise positions [36], [37].

Figure 3 illustrates the layout of the experimental environment, where beacons (denoted by green dots) are placed at known fixed coordinates across different rooms. The tag (red dot) worn by the individual emits UWB pulses, which are received by the beacons. Six beacons are deployed, three in the living room and three in the bedroom, allowing independent localisation within each space. When the tag is within range of a set of beacons, only those beacons actively calculate the TOA, while others are ignored. This selective activation minimises interference and computational complexity. The distance between the tag and each beacon is determined by calculating the TOA of each signal.

The beacon-tag communication process begins with the tag periodically sending UWB pulses. Each beacon calculates the time difference of signal arrival and estimates its distance from the tag. These distances are then passed to a localisation algorithm. The localisation algorithm relies on the trilateration technique, which computes the tag's position based on the distances from multiple beacons. For instance, in the proposed setup, separate beacon networks are deployed for different rooms, allowing each network to operate independently. When the tag (individual) enters a room, the corresponding beacon set actively calculates its position, while the others remain inactive. To minimise localisation error, signal processing techniques ensure that the TOA model accounts for environmental factors, including potential non-Line-of-Sight (NLOS) conditions caused by obstructions such as furniture or walls.

*1) Localisation by Trilateration:* Indoor localisation is performed using UWB communication, which transmits short-duration pulses across a broad frequency range (3.1 to 10.6 GHz), providing resistance to interference. The distance between the wearable tag and each fixed anchor beacon is calculated using Two-Way Ranging (TWR), which determines the ToF of the radio waves travelling at the speed of light, $c \approx 3 \times 10^8$ m/s. The distance is calculated as:

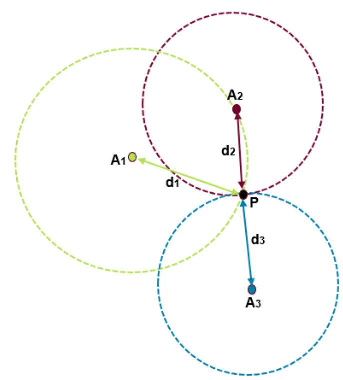



Fig. 4. Trilateration for tag position estimation using distance measurements from three fixed beacons, the intersection of the circles determines the tag's position.

$$Distance = \frac{ToF \times c}{2} \tag{1}$$

To estimate the tag's position, trilateration is applied using distance measurements from a minimum of three fixed beacons, as illustrated in Figure 4. The tag's coordinates $(x, y)$ are determined by the intersection of the circles defined by the distance $d_i$ from each anchor $(x_i, y_i)$:

$$d_i^2 = (x - x_i)^2 + (y - y_i)^2 \tag{2}$$

To solve for the unknown coordinates $(x, y)$, the non-linear distance equations are transformed into a solvable system of linear equations. By expanding the squared terms and subtracting the equation for one reference anchor (Anchor $N$) from the equations for the others, the squared coordinates $(x^2+y^2)$ are eliminated. This algebraic operation simplifies the system into the linear matrix form $A{\cdot}\mathbf{x} = \mathbf{b}$, where $\mathbf{x} = [x, y]^T$ is the vector of the unknown coordinates.

The final linear system derived from these algebraic steps can be expressed as:

$$A.\mathbf{x} = \mathbf{b} \tag{3}$$

The least-squares solution for the coordinates $\mathbf{x}$ is then calculated as: $\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}$. This method effectively transforms the complex non-linear problem into a computationally efficient linear solution, essential for real-time tracking.

*2) ML Models for Position Prediction:* To implement accurate position prediction, ML models are used to process the highly non-linear data obtained from UWB sensors. The aim is to predict an individual's position within a defined space by modelling the problem as a classification task, dividing the space into distinct segments. For example, in a home environment, as shown in Figure 3, the space is divided into segments such as the kitchen, living room, bed, and couch areas. Each segment is mapped to its corresponding UWB profile, ensuring that the distance ranges from the anchors remain consistent as the tag moves within the space. Using $N$ beacons $i$, $i \in 1, 2, 3..., N$ the distance measurements received from the tag can be expressed as a series $D_{i-1}, D_{i-2}, ..., D_{i-M_i}$, where $M_i$ represents the distance between the tag and anchor $i$. These distance measurements, derived from the TOA distance model, are converted into a profile map for the space [56]. The profile map for a location $(x, y)$ is represented as:

$$D_{x,y} = \begin{bmatrix} D_{1-1}, D_{1-2}, \ldots, D_{1-M_1}, \\ D_{2-1}, D_{2-2}, \ldots, D_{2-M_2}, \\ ..., \\ D_{N-1}, D_{N-2}, \ldots, D_{N-M_N} \end{bmatrix} \tag{4}$$

At a specific time $t$, the distance data, denoted as $\mathbf{D}_t = [d_1, d_2, \ldots, d_n]$, serves as the input feature vector to the ML model. The model predicts the user's position, and its output is expressed as the coordinates of the predicted position at time $t$:

$$(x_t, y_t) = \text{argmax}_{(x,y)}\, f_{\text{model}}(\mathbf{D}_t)$$

where $\text{argmax}_{(x,y)}$ selects the segment or region with the highest probability, and $(x, y)$ represents the centre coordinates of that segment. The choice of ML models focuses on their ability to learn spatial relationships and non-linear patterns from the input data, while balancing computational efficiency and predictive performance. Four algorithms - K-Nearest Neighbors (kNN), Support Vector Machine (SVM), RF, and MLP were evaluated for their suitability in the context of position prediction. kNN served as a simple, instance-based baseline [57], SVM exploited the kernel trick to handle non-linear separation of data points [58] and the RF enhanced robustness against noisy UWB measurements [59]. The MLP is deployed to utilise its deep learning capabilities for modelling the complex non-linear relationships inherent in the UWB data [60]. Each model was evaluated based on predictive accuracy and generalisability.

### B. Fall Detection Module

The Fall Detection module is designed to accurately detect falls by analysing motion patterns derived from wearable sensors. It relies on acceleration data $\{A_x, A_y, A_z\}$ and angular velocity data $\{G_x, G_y, G_z\}$ obtained from the accelerometer and gyroscope, which are collected through a wearable node typically worn on the waist. The ability to detect such rapid changes in motion is critical for distinguishing between normal activities and potential falls, enabling timely intervention and support. To extract meaningful features from this data, feature extraction is implemented to analyse the temporal characteristics of the acceleration and angular velocity signals. This ensures that the system remains sensitive to the subtle yet critical motion patterns that define a fall.

*1) Feature Extraction:* The feature extraction process transforms raw sensor data into a set of informative descriptors, enabling the model to effectively analyse motion patterns associated with falls. A sliding window approach is employed, dividing the continuous time-series data into fixed-size overlapping segments, with a specified window size and shift. This segmentation facilitates the analysis of temporal variations within each segment, capturing patterns and trends essential for distinguishing between normal activities and falls. From the raw accelerometer and gyroscope data, a total of 17 statistical features are extracted within each window. These features include both basic and advanced statistical metrics to comprehensively describe the characteristics of the motion data. Basic statistical features include measures of central tendency and variability, such as mean, median, minimum, maximum, and standard deviation. These features provide foundational insights into the distribution and spread of sensor readings. Advanced statistical features, including kurtosis and skewness, are computed. These metrics help analyse the shape of the data, identifying sharp peaks or asymmetries that may indicate abrupt movements or changes in orientation. Mean absolute deviation and interquartile range are also derived to capture the variability of motion patterns. Table I provides a summary of the extracted features.

*2) Fall Detection Model:* The proposed fall detection model integrates a CNN-LSTM VAE enhanced with an attention mechanism, as illustrated in Figure 5. This architecture is tailored to differentiate falls from normal activities by prioritising critical segments of time-series sensor data. It combines convolutional layers for spatial feature extraction, LSTM layers for sequence modelling, and an attention mechanism to focus on essential temporal regions, see [61], [62] for more details.

The convolutional layers extract spatial patterns from the sensor data, capturing localised variations effectively. These layers apply transformations using filters and kernel sizes with padding to preserve input dimensions. After each convolutional layer, max pooling is applied to reduce dimensionality and retain significant features, pooling over a window of size 2. This combination of operations ensures that spatial features relevant to fall detection are extracted while reducing computational complexity. The VAE architecture also incorporates LSTM layers optimised to model temporal dependencies and capture sequential patterns in the data. Each LSTM cell uses a gating mechanism, where the hidden state $h_t$ is updated based on the output gate $o_t$ and the candidate cell state $C_t$, resulting in the following equation:

$$h_t = o_t \cdot \tanh(C_t)$$

This operation allows the model to effectively process time-series dependencies and learn from the sequential structure of the data. The encoder generates a latent representation

TABLE I
EXTRACTED FEATURES FOR FALL DETECTION MODULE

| Feature | Description |
|---|---|
| Central Tendency and Dispersion | Mean, median, minimum, maximum, and standard deviation of acceleration $\{A_x, A_y, A_z\}$ and angular velocity $\{G_x, G_y, G_z\}$. |
| Shape of Distribution | Kurtosis and Skewness, capturing the symmetry and peakedness of the sensor data distribution. |
| Variability Metrics | Mean absolute deviation, interquartile range, and variance to quantify data dispersion. |
| Dynamic Changes | Mean and median of absolute differences, as well as the sum of absolute differences between consecutive values, capturing variations in motion. |
| Magnitude of Movement | Root mean square (RMS) of acceleration and angular velocity signals, representing the intensity of movements. |

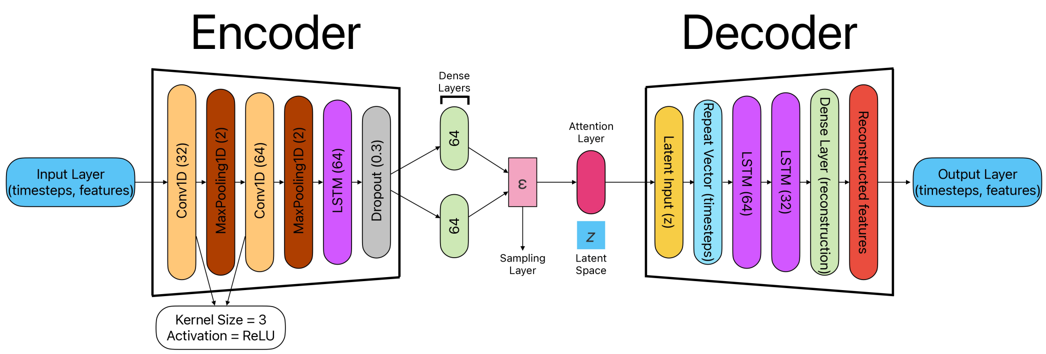



Fig. 5. Proposed CNN-LSTM VAE Architecture with Attention Layer

characterised by two parameters: $z_{mean}$ and $z_{log\_var}$, representing the mean and log variance of the latent variable distribution, respectively. The reparameterisation trick enables differentiable sampling:

$$z = z_{mean} + \exp(0.5 \cdot z_{log\_var}) \cdot \epsilon$$

where $\epsilon \sim N(0, 1)$. This latent representation enhances the model's ability to detect anomalies, such as fall events, by distinguishing them from normal activities. An attention mechanism is incorporated after the encoding stage to dynamically assign weights to each timestep, highlighting crucial portions of the sequence. The attention score $e_t$ for a given timestep is computed as:

$$e_t = \tanh(W \cdot h_t + b)$$

where $W$ and $b$ are the attention layer's weights and biases. These scores are then normalised using the softmax function:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t'} \exp(e_{t'})}$$

The context vector $c$ is computed as the weighted sum of the LSTM outputs:

$$c = \sum_t \alpha_t \cdot h_t$$

This weighted context vector emphasises the most relevant parts of the time series for fall detection. The VAE loss function combines two components: the reconstruction loss and the Kullback-Leibler divergence. The reconstruction loss evaluates the decoder's ability to reconstruct the input:

$$L_{reconstruction} = \frac{1}{N} \sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2$$

The KL divergence penalises deviations of the latent distribution from a standard Gaussian:

$$L_{KL} = -0.5 \sum_{j=1}^{d} \left(1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2\right)$$

The total VAE loss is the sum of the reconstruction loss and the KL divergence:

$$L_{VAE} = L_{reconstruction} + L_{KL}$$

The outputs are passed through a late fusion layer, which integrates positional predictions with the extracted temporal and spatial features.

*3) Late Fusion:* In the proposed approach, late fusion is implemented at the decision level by combining classifier outputs from the fall detection model and the UWB-based positioning system. Each modality is first processed independently: the fall detection model provides binary outputs ($F \in \{0, 1\}$), indicating "No Fall" or "Fall," while the UWB localisation system assigns positional data ($L \in$ Locations, e.g., Bedroom, Living Room). The fusion process integrates these results into a unified decision, enabling simultaneous activity recognition and spatial identification.

The fusion operation is formulated as a weighted sum of classifier scores:

$$S_{\text{fusion}} = \alpha S_{\text{fall}} + \beta S_{\text{location}} \tag{5}$$

where $S_{\text{fall}}$ and $S_{\text{location}}$ represent the normalised scores from the fall detection and localisation models, respectively, and $\alpha, \beta$ are weighting coefficients chosen to balance their contributions. The fused feature representation is then passed through a linear layer to transform the combined scores into a unified decision. The final output maps the detection status and spatial location, expressed as:

$$\text{Output} = \begin{cases} \text{Fall in BR} & \text{if } F = 1 \text{ and } L = \text{BR}, \\ \text{No Fall in LR} & \text{if } F = 0 \text{ and } L = \text{LR}, \\ \text{Unknown Location} & \text{if location unavailable.} \end{cases} \quad (6)$$

where BR and LR represent Bedroom and Living Room locations, respectively. The decision layer is designed to address two primary challenges to ensure accurate synchronisation of multimodal data and provide interpretable outputs for real-time monitoring. Synchronisation ensures that temporal alignment between activity detection and location data is maintained. Errors in localisation, such as missing or invalid data, are handled by defaulting to "Unknown Location," thus avoiding false outputs.

The fusion system enhances contextual understanding by associating detected activities with spatial locations, enabling precise identification of fall events and their occurrence zones. This approach improves practical utility in real-world scenarios, such as assisted living, where rapid and location-specific responses are critical.

## IV. Experiment and Performance Evaluation

### A. Data and Dataset Collection

A custom dataset was collected to meet the dual objectives of localisation and fall detection. Data collection occurred in a simulated home environment, where participants performed ADL. The data collection phase involved recording ADL in a simulated home environment, using accelerometer and gyroscope sensors to capture six degrees of freedom (6DOF) motion data. The dataset comprised over 33,273 instances, primarily featuring normal activities with staged fall events injected to introduce anomaly patterns. A range of activity types, timings, and participant behaviours ensured data diversity, enhancing model generalisation. The dataset was recorded over 10 days, with normal activities such as walking, sitting, standing, and position transitions, along with staged falls. These falls were varied (forward, backwards, sideward, and collapsing to a seated position) to avoid identifiable patterns, ensuring robust anomaly detection. For localisation, beacon-based tags, combined with accelerometer and gyroscope data, were processed using sensor fusion algorithms to achieve accurate indoor positioning and activity tracking. Table II provides an overview of the dataset, including activity types, data collection durations, and recorded features:

### B. Performance Evaluation and Results

The experiments conducted aimed to validate the proposed localisation and fall detection phases, focusing on their ability to accurately track user positions and identify fall events. The focus was on integrating indoor localisation with fall detection using late fusion techniques, with experiments designed to test various configurations and evaluate their impact on model performance. Feature extraction included statistical metrics such as central tendency, dispersion, and shape of distribution, variability metrics, as detailed in Table 1. The evaluation focused on precision, recall, accuracy, and F1 score, given their relevance in assessing anomaly detection and classification tasks. Precision measured the proportion of correctly identified falls out of all predicted falls, while recall captured the ability to detect actual falls from the dataset. Accuracy reflected the overall classification performance, and F1 score balanced precision and recall to account for imbalanced data. To ensure that the reported performance metrics were not dependent on a specific train-test split, all models were evaluated using k-fold cross-validation. The mean performance across folds is reported for both localisation and fall detection. The following experiments evaluate the performance of the proposed approach. The first experiment assessed indoor location tracking and fall detection, evaluating the system's ability to accurately track user positions and detect falls. The second experiment investigated the effect of window size and training dataset size on model performance. The final experiment focused on threshold calculation for reconstruction errors in fall detection, examining how adjustments to the threshold impacted fall classification accuracy.

*1) Location Experiment and Results:* **Evaluation of Different Window Sizes:** This experiment evaluated the impact of varying window sizes on indoor positioning accuracy using four ML models MLP, SVM, RF and kNN. Window sizes of 3, 5, and 10 with a fixed overlap of 0.2 were tested to determine optimal configurations for accurate position classification. Table V summarises the performance of each model under different windowing configurations. MLP consistently outperformed other models with all window sizes. Smaller window sizes showed moderate performance improvements. At a window size of 3, MLP achieved an F1 score of 0.61 and an accuracy of 0.67, outperforming other models. Increasing the window size to 5 resulted in a slight drop in performance across all models, with MLP achieving 0.65 accuracy. At a window size of 10, performance degraded further; for instance, SVM and MLP recorded accuracy scores of 0.53 and 0.61, respectively. Larger windows reduced sensitivity to temporal variations, limiting the ability to capture dynamic positioning patterns effectively. To address the observed limitations, the models were tested without applying any window size or overlap, using all data instances directly. This configuration yielded significant performance gains, with MLP achieving 0.94 accuracy, with RF and SVM as 0.91 and 0.88, respectively. The results suggest that while windowing can enhance performance in some cases, removing temporal segmentation may be more effective for datasets characterised by dense, high-resolution temporal data.

**Impact of Training Dataset Composition on Indoor Localisation Performance**: Following the optimisation of window size in the previous experiment, this investigation examines the influence of training dataset size and composition on model performance for indoor localisation. For this experiment, two distinct datasets were constructed to assess the effect of training data selection on classification accuracy across multiple machine learning models. The first dataset, referred to as "All Positions," includes the complete set of positional instances captured during data collection. While comprehensive, this dataset risks redundancy due to overlapping or closely spaced position samples. The second

TABLE II
DATASET OVERVIEW

| Activity type | Instances | Duration(s) | Features Collected | Notes |
|---|---|---|---|---|
| Normal ADL | 29,002 | 10 - 60 | Accelerometer, Gyroscope | Includes walking, sitting etc. |
| Staged Falls | 4,271 | 10 - 20 | Accelerometer, Gyroscope | Simulated various fall scenarios. |
| Localization Points | N/A | Continuous | Beacon-based tags | Captured location data for activity tracking. |

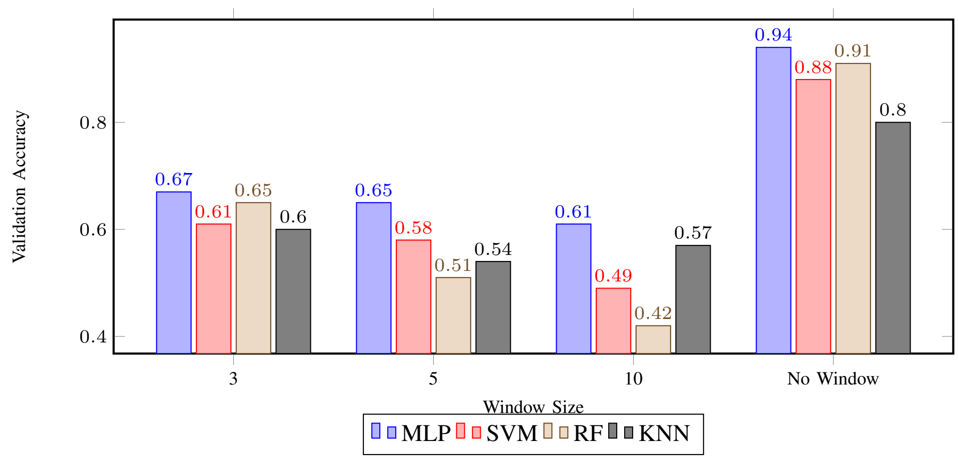



Fig. 6. Validation Accuracy of Different Models Across Window Sizes - This should be for the different rooms/segments

TABLE III
PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS ACROSS DIFFERENT WINDOW SIZES FOR INDOOR POSITIONING ACCURACY.

| Window sizes | MLP | SVM | RF | KNN |
|---|---|---|---|---|
| 3 | 0.67 | 0.61 | 0.65 | 0.60 |
| 5 | 0.65 | 0.58 | 0.51 | 0.54 |
| 10 | 0.61 | 0.49 | 0.42 | 0.57 |
| No Window | 0.94 | 0.88 | 0.91 | 0.80 |

dataset, "Selected Positions," comprises a balanced subset of representative position instances, minimising redundancy while retaining sufficient variability to capture spatial patterns effectively. This approach allows investigation into whether a more curated dataset enhances generalisation and reduces model overfitting.

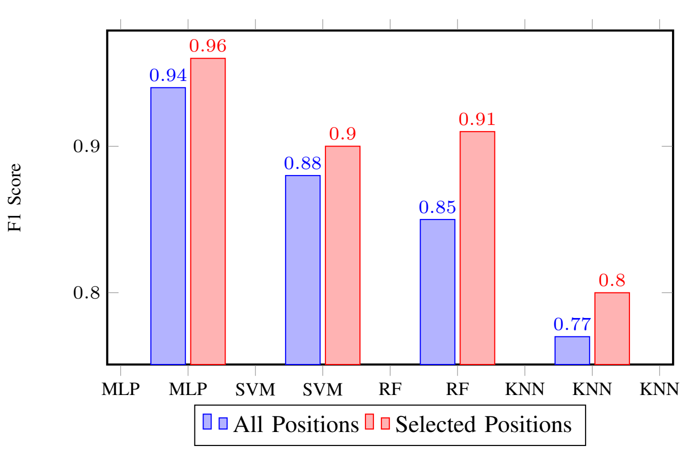



Fig. 7. Validation Accuracy of Different Models Across Window Sizes

Experimental results demonstrate that dataset composition significantly affects model performance. For instance, the Multilayer Perceptron (MLP) model trained on the "All Positions" dataset achieved an accuracy of 0.94 and an F1 score of 0.93. However, when trained on the "Selected Positions" dataset, MLP performance improved, achieving an accuracy of 0.96 and an F1 score of 0.98, reflecting the performance in both precision and recall. Similarly, the Support Vector Machine (SVM) model improved from an accuracy of 0.88 on the "All Positions" dataset to over 0.90 when trained on the balanced dataset. This performance were consistent across other models, underscoring the effectiveness of selective training data in boosting performance.

The findings reinforce the importance of optimising not only pre-processing techniques, such as windowing but also the structure and distribution of training data. The balanced samples enhance a model's ability to generalise spatial patterns while mitigating issues related to redundant information. Furthermore, MLP outperformed other models across dataset configurations, demonstrating sensitivity to positional variations.

*2) Fall Detection Experiment and Results:* We evaluated five unsupervised fall detection models designed to enhance generalisation and anomaly detection, all trained exclusively on normal activity data, identifying falls via reconstruction error. These architectures combined CNNs, LSTM, BiLSTMs, VAEs, and attention mechanisms to explore their influence on learning temporal patterns and detecting fall deviations. The baseline architecture, a CNN-LSTM Autoencoder, used convolutional layers for spatial feature extraction and LSTM layers for temporal dependencies, though it showed limited precision and sensitivity. We implemented a CNN-BiLSTM

Autoencoder with an attention layer to improve temporal context learning. The idea was to leverage BiLSTM network for both forward and backwards dependencies, while the attention mechanism focused on critical moments in the sequence.

We also implemented a VAE integrated into a CNN-LSTM architecture to learn a latent distribution for normal activities, improving reconstruction fidelity and fall differentiation. Furthermore, an attention mechanism was added to the VAE CNN-LSTM architecture to boost its performance. Experiments were conducted on these architectures evaluate the impact of sequential modelling, latent space encoding, and attention-based weighting on fall anomaly detection.

**Effect of Window Sizes**: This experiment examined the impact of sliding window sizes 3, 5, and 10 on the performance of five fall detection architectures, each trained with a fixed overlap of 0.2. The models evaluated included both standard and variational autoencoders with attention configurations, using combinations of CNN, LSTM, and BiLSTM layers. Model evaluation was based on accuracy, precision, recall, and F1 score. As shown in Table IV, window size had a clear impact on performance. Across all architectures, a window size of 5 consistently yielded the best results. Most notably, the VAE CNN + LSTM with attention achieved the highest overall performance at this setting, with an accuracy of 0.98, precision of 0.95, recall of 1.00, and F1 score of 0.97. These results indicate a strong ability to generalise normal activity patterns and detect anomalies accurately. In contrast, performance declined at window sizes 3 and 10. With a smaller window (size 3), the models lacked sufficient temporal context, limiting their ability to capture complex activity dynamics. Conversely, with a larger window (size 10), performance degradation was attributed to the inclusion of redundant or noisy data, which likely diluted the reconstruction error signal used for fall detection. Performance trends are further illustrated in Figure 10, which visualises the precision, recall, F1 score, and precision of the best performing model VAE CNN + LSTM with attention, in the three window sizes. The peak at window size 5 reinforces its suitability as the optimal configuration for this task.

TABLE IV
PERFORMANCE METRICS OF VARIOUS FALL DETECTION ARCHITECTURES ACROSS WINDOW SIZES. VAE + CNN + LSTM WITH A WINDOW SIZE OF 5 ACHIEVED THE BEST OVERALL PERFORMANCE.

| Model | Window Sizes | Acc | Prec | Rec | F1 |
|---|---|---|---|---|---|
| AE CNN + LSTM | 3 | 0.81 | 0.78 | 0.85 | 0.81 |
| | 5 | 0.84 | 0.82 | 0.88 | 0.85 |
| | 10 | 0.79 | 0.76 | 0.82 | 0.79 |
| AE CNN + BiLSTM (A) | 3 | 0.83 | 0.80 | 0.87 | 0.83 |
| | 5 | 0.86 | 0.84 | 0.89 | 0.86 |
| | 10 | 0.81 | 0.79 | 0.85 | 0.82 |
| VAE CNN + LSTM | 3 | 0.78 | 0.74 | 0.79 | 0.76 |
| | 5 | 0.83 | 0.80 | 0.85 | 0.82 |
| | 10 | 0.76 | 0.72 | 0.78 | 0.75 |
| VAE CNN + LSTM (A) | 3 | 0.94 | 0.91 | 0.97 | 0.94 |
| | 5 | **0.98** | **0.95** | **1.00** | **0.97** |
| | 10 | 0.83 | 0.86 | 0.80 | 0.81 |
| VAE CNN + BiLSTM (A) | 3 | 0.90 | 0.88 | 0.92 | 0.90 |
| | 5 | 0.96 | 0.94 | 0.98 | 0.96 |
| | 10 | 0.88 | 0.86 | 0.91 | 0.88 |

**Effect of Training Data Composition**: Following the identification of window size 5 as optimal in Experiment 1, this experiment investigates the impact of training dataset composition on fall detection performance. Using the fixed optimal window size, we evaluate how the size and selection of training data influence model effectiveness across five architectures, including both standard and variational autoencoders, with and without attention layers. Two distinct training datasets were constructed. The first, All Normal Activities, includes the entire pool of available normal activity sequences, while the second, Selected Normal Activities, comprises a balanced and representative subset designed to avoid redundancy and overfitting. This setup enables us to assess whether a curated dataset improves the model's ability to differentiate between normal and fall events.

As illustrated in Figure 10, models trained on the Selected Normal Activities dataset consistently outperformed those trained on the full dataset. For example, the Autoencoder CNN-LSTM model without attention achieved a low F1 score of 0.24 when trained on the full set, indicating poor recall and overfitting to normal patterns. However, when trained on the selected subset, its F1 score improved significantly to 0.87, with perfect recall, demonstrating enhanced fall sensitivity through balanced data representation. The most notable improvement was observed in the VAE + CNN + LSTM model with an attention layer. When trained on the selected dataset, it achieved an accuracy of 97.8%. In contrast, the same model trained on the full dataset showed a noticeable drop in performance, reinforcing the value of data curation.

These results clearly demonstrate that training data composition is as critical as architectural design. Even with optimal window sizing, using a well-balanced, representative training set significantly enhances model generalisation and anomaly detection performance. Together with findings from Experiment 1, this experiment confirms that the best-performing configuration for practical fall detection is a VAE + CNN + LSTM with attention, trained on a carefully selected subset of normal activity data.

**Effect of Threshold Selection on Reconstruction Errors** In this experiment, we investigate how varying the anomaly detection thresholds based on reconstruction error percentiles impacts fall detection performance. Specifically, we consider three threshold levels $90^{th}$, $95^{th}$, and $99^{th}$ percentiles tested across the five model architectures to evaluate sensitivity and specificity trade-offs using F1 score as the primary metric. As shown in Table V and visualised in Figure 10, model performance was highly sensitive to the chosen threshold. At the 90th percentile, most models demonstrated balanced precision and recall. The VAE + CNN + LSTM with attention achieved the highest F1 score of 0.98, with a perfect recall of 1.00, indicating robust fall detection without missing critical events. The summary in Table VI supports this interpretation, showing that models with both variational and attention mechanisms are more resilient to moderate threshold changes but still suffer at extreme values. Similarly, other attention-based models, including the VAE + CNN + BiLSTM, also performed well at this threshold. The 95th percentile provided a strong

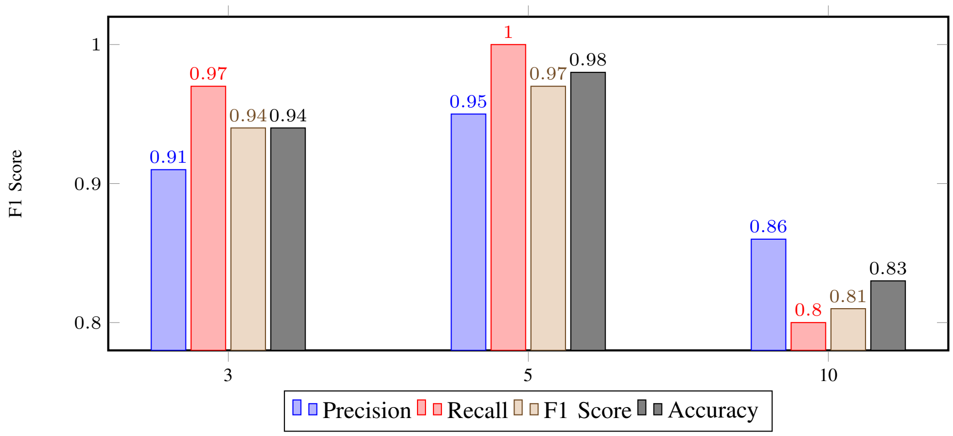



Fig. 8. Performance Metrics of Best Model Across Window Sizes

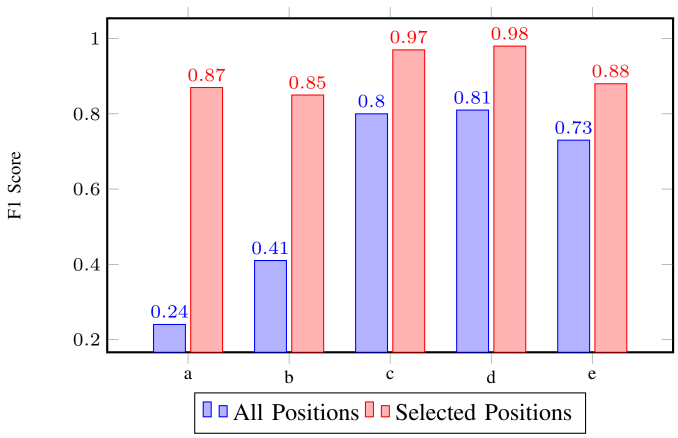



Fig. 9. Validation Accuracy of Different Models Across Window Sizes where a = AE CNN-LSTM; b = AE CNN-BiLSTM (A); c = VAE CNN-LSTM; d = VAE CNN-LSTM (A); e = VAE CNN-BiLSTM (A)

compromise for models prioritising higher precision. For example, the AE + CNN + LSTM model, which struggled at lower thresholds, improved significantly, reaching an F1 score of 0.95, with both precision and recall near optimal. Notably, the VAE + CNN + BiLSTM model achieved its highest F1 score (0.96) at this threshold, suggesting enhanced control over false positives while maintaining detection sensitivity. However, performance declined sharply at the 99th percentile for all models. The VAE + CNN + LSTM with attention, which had performed best at lower thresholds, dropped to an F1 score of 0.18, with recall falling below 0.08. This pattern was consistent across architectures, indicating that an overly restrictive threshold reduces the model's ability to detect falls, resulting in high false negatives—unacceptable in real-world applications where missed detections are critical.

## V. Discussion

The experimental results validate the effectiveness of our proposed hybrid system for indoor localisation and fall detection. The modular architecture allows for independent optimisation of both localisation and fall detection components. Integrating these via a decision-level fusion strategy significantly enhances the system's ability to detect falls while simultaneously identifying their precise location within an indoor environment. A demonstration of the proposed indoor localisation and fall detection system is provided in online videos. [1] For localisation, the MLP model performed best, particularly when trained on the full-resolution dataset without window segmentation. Further improvements in generalisation and reduced overfitting were observed by training on a carefully curated subset of positional instances. The result suggests the significance of high-resolution and balanced datasets for accurately representing user movement across distinct spatial zones, which is essential for precise location mapping. Generalising UWB localisation across different rooms and buildings typically requires environment-aware calibration. Emerging techniques such as environmental fingerprinting, model-based compensation, and domain adaptation offer promising foundations for extending localisation models to unseen indoor layouts. Incorporating these strategies in future work will enhance the scalability and cross-environment robustness of the proposed system beyond the current testbed. In fall detection, the VAE + CNN + LSTM model with an attention mechanism consistently performed best. This combination of probabilistic encoding (VAE) and temporal weighting (attention) enabled robust reconstruction of normal activity patterns while accurately identifying anomalies, such as falls. The attention mechanism enhanced temporal focus, and the VAE's latent encoding improved generalisation. Our results also showed that training on a representative subset of normal activities, rather than the entire dataset, further boosted the model's precision and recall. Threshold tuning critically influenced detection reliability. A lower threshold of 90th percentile maximised sensitivity, but higher thresholds up to 99th percentile led to increased false negatives. This highlights the necessity of careful calibration to balance detection sensitivity and specificity, especially in safety-critical applications where minimising false negatives is paramount.

The decision-level fusion integrates fall detection

[1] Video demonstrations of the system are available online at: https://www.youtube.com/watch?v=yMouJz2BcMU&list=PLxQnOSnSGFrmhxxujs3QD_T_2T-phWmsQ&index=3 and https://www.youtube.com/watch?v=DAhLQok13YM&list=PLxQnOSnSGFrmhxxujs3QD_T_2T-phWmsQ&index=1.

| Threshold | AE CNN-LSTM | AE CNN-BiLSTM (A) | VAE CNN-LSTM | VAE CNN-LSTM (A) | VAE CNN-BiLSTM (A) |
|---|---|---|---|---|---|
| 90 | 0.87 | 0.85 | 0.97 | 0.98 | 0.88 |
| 95 | 0.95 | 0.94 | 0.88 | 0.85 | 0.96 |
| 99 | 0.88 | 0.89 | 0.14 | 0.18 | 0.12 |

TABLE V
MODEL'S F1 SCORE ACROSS DIFFERENT THRESHOLD

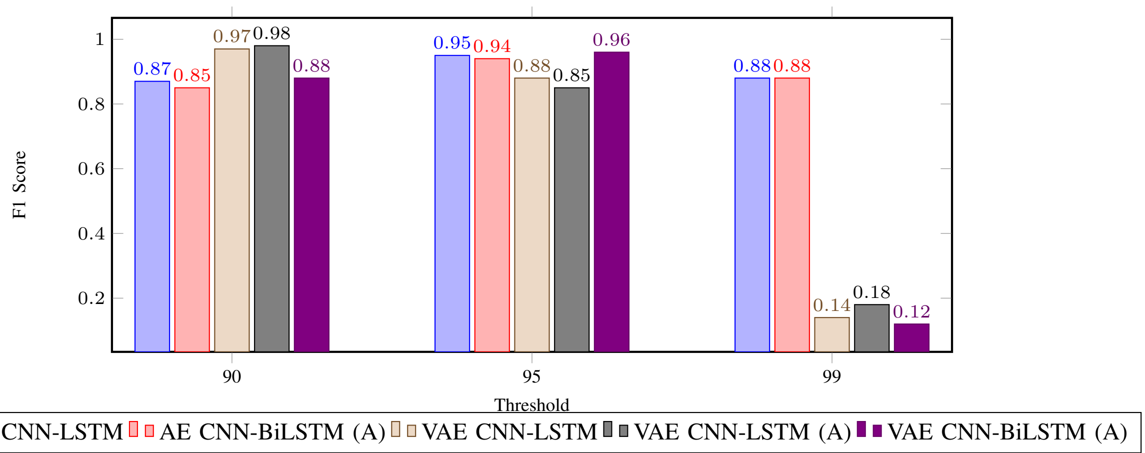



Fig. 10. Performance Metrics of Best Model Across Window Sizes

TABLE VI
PERFORMANCE COMPARISONS

| Architecture | Attention Layer | Accuracy | Recall | F1 Score | Performance Comments |
|---|---|---|---|---|---|
| AE CNN + LSTM | No | Low | Low | 0.87 | Struggles to capture complex patterns in fall sequences |
| AE CNN + BiLSTM | Yes | Moderate | Moderate | 0.85 | Attention layer improved performance but was insufficient |
| VAE + CNN + LSTM | No | High | High | 0.97 | Variational layer improved pattern recognition |
| VAE + CNN + LSTM | Yes | Highest | 1.00 | 0.98 | Attention layer enhanced temporal pattern detection |
| VAE + CNN + BiLSTM | Yes | Moderate | Moderate | 0.88 | Using Bi-LSTM Layers with VAE yielded lower performance |

outputs ($F \in \{0, 1\}$) with localisation predictions ($L \in \{LivingRoom, Kitchen, BedArea, CouchArea, FloorArea\}$). This is achieved through a weighted score formulation:

$$S_{\text{fusion}} = \alpha S_{\text{fall}} + \beta S_{\text{location}}$$

Where, $\alpha$ and $\beta$ control the influence of each module on the final decision. The system's output not only confirms if a fall has occurred but also pinpoints its exact location, enabling spatially contextualised alerts. As depicted in Figure 11, fall events were detected across various zones and times of day. This spatial-temporal distribution demonstrates the system's capability to track and characterise high-risk areas, providing valuable insights for caregivers and automated monitoring systems.

This level of granularity facilitates proactive risk management, such as reinforcing safety measures in specific high-risk zones, for example, activities in the kitchen in the early morning, and the couch area in the evening. Furthermore, by maintaining a decoupled architecture for detection and localisation, the system ensures modularity. This allows for independent updates or improvements to either module, thereby enhancing scalability and long-term applicability in real-world settings.

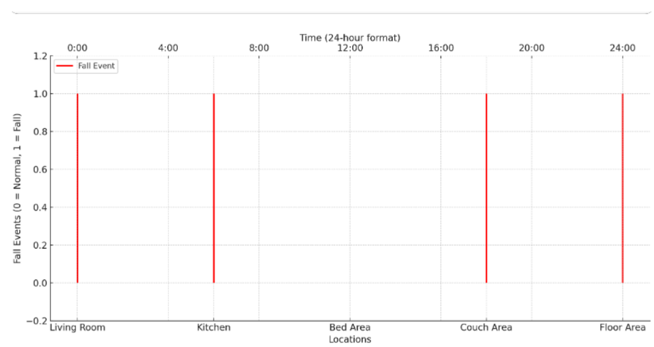



Fig. 11. Late fusion output showing spatial-temporal fall event detection across locations in a 24-hour format. Red lines represent fall events.

## VI. CONCLUSION

In this work, we proposed a position-aware fall detection approach that combines indoor localisation with anomaly-based activity recognition to enable real-time monitoring of elderly individuals within home environments. By fusing spatial and motion data at the decision level, the system not only detects the occurrence of a fall but also identifies its precise location, addressing a key limitation in existing fall detection approaches.

We evaluated several architectural configurations for fall detection, including standard CNN–LSTM autoencoders and enhanced variants incorporating BiLSTMs, variational encoding, and attention mechanisms. Among these, the VAE + CNN + LSTM model with attention achieved the highest performance in terms of precision, recall, and F1 score when trained on curated normal activity sequences and optimised window parameters. For indoor localisation, UWB-based trilateration combined with machine learning models proved effective, with the MLP outperforming other classifiers in accuracy and generalisation.

The proposed approach was validated in a realistic, home-like environment using a custom dataset capturing a wide range of daily activities and staged falls. While the results demonstrate strong accuracy and robustness, several limitations remain. First, the UWB localisation framework was evaluated within a single controlled environment, and its generalisability to different residential layouts, building materials, and furniture configurations is not yet established. Future work will therefore explore environment-specific calibration, domain adaptation, and multi-room modelling to improve cross-environment robustness. Second, the fall detection model was trained on staged falls collected over a short period. Real-world fall events among older adults are rare, highly variable, and ethically challenging to capture, limiting the ecological validity of staged datasets. Longitudinal daily-living data and semi-supervised refinement will be essential to strengthen real-world applicability. Third, the hybrid VAE + CNN + LSTM Attention architecture introduces some computational footprint. Although evaluated offline on high-performance hardware, practical deployment on edge or wearable platforms will require quantifying model complexity and applying optimisation techniques to ensure low-latency, on-device operation.

The spatial-temporal mapping of fall events generated by the system provides actionable insights that can support caregivers in prioritising interventions and adapting home environments to reduce risk. Looking forward, we plan to extend the framework to multi-occupant scenarios, enhance localisation precision through additional sensing modalities, and investigate real-time edge deployment to reduce latency and improve scalability. Long-term in-home studies will further evaluate sustained system performance, user adaptation, and overall usability in ambient assisted living contexts.

## REFERENCES


[1] V. L. Bengtson, “Is the “contract across generations” changing? effects of population aging on obligations and expectations across age groups,” in *The changing contract across generations*. Routledge, 2024, pp. 3–23.

[2] R. Lesthaeghe, “The second demographic transition, 1986–2020: sub-replacement fertility and rising cohabitation—a global update,” *Genus*, vol. 76, pp. 1–38, 2020.

[3] G. Livingston, J. Huntley, K. Y. Liu, S. G. Costafreda, G. Selbæk, S. Alladi, D. Ames, S. Banerjee, A. Burns, C. Brayne *et al.*, “Dementia prevention, intervention, and care: 2024 report of the lancet standing commission,” *The Lancet*, vol. 404, no. 10452, pp. 572–628, 2024.

[4] S. Dixson-Declève, O. Gaffney, J. Ghosh, J. Randers, J. Rockstrom, and P. E. Stoknes, *Earth for All: A survival guide for humanity*. new society Publishers, 2022.

[5] K. G. Kinsella and D. R. Phillips, *Global aging: The challenge of success*. Population Reference Bureau Washington, DC, 2005, vol. 60, no. 1.

[6] D. E. Bloom and D. L. Luca, “The global demography of aging: facts, explanations, future,” in *Handbook of the economics of population aging*. Elsevier, 2016, vol. 1, pp. 3–56.

[7] J. Strommen, H. Fuller, G. F. Sanders, and D. M. Elliott, “Challenges faced by family caregivers: Multiple perspectives on eldercare,” *Journal of Applied Gerontology*, vol. 39, no. 4, pp. 347–356, 2020.

[8] W. W. Y. Lam, K. Nielsen, C. A. Sprigg, and C. M. Kelly, “The demands and resources of working informal caregivers of older people: a systematic review,” *Work & Stress*, vol. 36, no. 1, pp. 105–127, 2022.

[9] V. Bressan, C. Visintini, and A. Palese, “What do family caregivers of people with dementia need? a mixed-method systematic review,” *Health & social care in the community*, vol. 28, no. 6, pp. 1942–1960, 2020.

[10] J. Tunstall, *Old and alone: A sociological study of old people*. Taylor & Francis, 2024.

[11] M. D. o. S. I. Committee on the Health and L. in Older Adults, *Social isolation and loneliness in older adults: Opportunities for the health care system*. National Academies Press, 2020.

[12] A. J. Perez, F. Siddiqui, S. Zeadally, and D. Lane, “A review of iot systems to enable independence for the elderly and disabled individuals,” *Internet of Things*, vol. 21, p. 100653, 2023.

[13] C. Sherrington, N. Fairhall, W. Kwok, G. Wallbank, A. Tiedemann, Z. A. Michaleff, C. A. Ng, and A. Bauman, “Evidence on physical activity and falls prevention for people aged 65+ years: systematic review to inform the who guidelines on physical activity and sedentary behaviour,” *International journal of behavioral nutrition and physical activity*, vol. 17, pp. 1–9, 2020.

[14] N. Salari, N. Darvishi, M. Ahmadipanah, S. Shohaimi, and M. Mohammadi, “Global prevalence of falls in the older adults: a comprehensive systematic review and meta-analysis,” *Journal of orthopaedic surgery and research*, vol. 17, no. 1, p. 334, 2022.

[15] R. Vaishya and A. Vaish, “Falls in older adults are serious,” *Indian journal of orthopaedics*, vol. 54, pp. 69–74, 2020.

[16] A. R. Santos-Lozada, “Trends in deaths from falls among adults aged 65 years or older in the us, 1999-2020,” *JAMA*, vol. 329, no. 18, pp. 1605–1607, May 2023.

[17] J. A. Haagsma, B. F. Olij, M. Majdan, E. F. Van Beeck, T. Vos, C. D. Castle, Z. V. Dingels, J. T. Fox, E. B. Hamilton, Z. Liu *et al.*, “Falls in older aged adults in 22 european countries: incidence, mortality and burden of disease from 1990 to 2017,” *Injury prevention*, vol. 26, no. Suppl 2, pp. i67–i74, 2020.

[18] X. Wang, J. Ellul, and G. Azzopardi, “Elderly fall detection systems: A literature survey,” *Frontiers in Robotics and AI*, vol. 7, p. 71, 2020.

[19] Y. Cheng, K. Wang, H. Xu, T. Li, Q. Jin, and D. Cui, “Recent developments in sensors for wearable device applications,” *Analytical and bioanalytical chemistry*, vol. 413, no. 24, pp. 6037–6057, 2021.

[20] S. Usmani, A. Saboor, M. Haris, M. A. Khan, and H. Park, “Latest research trends in fall detection and prevention using machine learning: A systematic review,” *Sensors*, vol. 21, no. 15, p. 5134, 2021.

[21] S. Subramaniam, A. I. Faisal, and M. J. Deen, “Wearable sensor systems for fall risk assessment: A review,” *Frontiers in digital health*, vol. 4, p. 921506, 2022.

[22] J. Liu, X. Li, S. Huang, R. Chao, Z. Cao, S. Wang, A. Wang, and L. Liu, “A review of wearable sensors based fall-related recognition systems,” *Engineering Applications of Artificial Intelligence*, vol. 121, p. 105993, 2023.

[23] X. Wang, E. Talavera, D. Karastoyanova, and G. Azzopardi, “Fall detection with a non-intrusive and first-person vision approach,” *IEEE Sensors Journal*, 2023.

[24] A. Ghosh, A. Chakraborty, D. Chakraborty, M. Saha, and S. Saha, “Ultrasense: A non-intrusive approach for human activity identification using heterogeneous ultrasonic sensor grid for smart home environment,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–22, 2023.

[25] N. L. Kazanskiy, S. N. Khonina, and M. A. Butt, “A review on flexible wearables-recent developments in non-invasive continuous health monitoring,” *Sensors and Actuators A: Physical*, p. 114993, 2024.

[26] J.-K. Kim, K. Lee, and S. G. Hong, “Detection of important features and comparison of datasets for fall detection based on wrist-wearable devices,” *Expert Systems with Applications*, vol. 234, p. 121034, 2023.

[27] M. Ş. Turan and B. Barshan, “Classification of fall directions via wearable motion sensors,” *Digital Signal Processing*, vol. 125, p. 103129, 2022.

[28] E. García, M. Villar, M. Fáñez, J. R. Villar, E. de la Cal, and S.-B. Cho, “Towards effective detection of elderly falls with cnn-lstm neural networks,” *Neurocomputing*, vol. 500, pp. 231–240, 2022.

[29] J. Yu, A. de Antonio, and E. Villalba-Mora, "Deep learning (cnn, rnn) applications for smart homes: a systematic review," *Computers*, vol. 11, no. 2, p. 26, 2022.

[30] A. Sultana, K. Deb, P. K. Dhar, and T. Koshiba, "Classification of indoor human fall events using deep learning," *Entropy*, vol. 23, no. 3, p. 328, 2021.

[31] I. K. Ihianle, A. O. Nwajana, S. H. Ebenuwa, R. I. Otuka, K. Owa, and M. O. Orisatoki, "A deep learning approach for human activities recognition from multimodal sensing devices," *IEEE Access*, vol. 8, pp. 179 028–179 038, 2020.

[32] J. Gutiérrez, V. Rodríguez, and S. Martin, "Comprehensive review of vision-based fall detection systems," *Sensors*, vol. 21, no. 3, p. 947, 2021.

[33] E. Alam, A. Sufian, P. Dutta, and M. Leo, "Vision-based human fall detection systems using deep learning: A review," *Computers in biology and medicine*, vol. 146, p. 105626, 2022.

[34] F. Zorriassatine, A. Naser, and A. Lotfi, "Gait anomaly detection with low cost and low resolution infrared sensor arrays," *IEEE Sensors Journal*, 2024.

[35] H. S. Fahama, K. Ansari-Asl, Y. S. Kavian, and M. N. Soorki, "An experimental comparison of rssi-based indoor localization techniques using zigbee technology," *IEEE Access*, 2023.

[36] M. Elsanhoury, P. Mäkelä, J. Koljonen, P. Välisuo, A. Shamsuzzoha, T. Mantere, M. Elmusrati, and H. Kuusniemi, "Precision positioning for smart logistics using ultra-wideband technology-based indoor navigation: A review," *Ieee Access*, vol. 10, pp. 44 413–44 445, 2022.

[37] F. Che, Q. Z. Ahmed, P. I. Lazaridis, P. Sureephong, and T. Alade, "Indoor positioning system (ips) using ultra-wide bandwidth (uwb)—for industrial internet of things (iiot)," *Sensors*, vol. 23, no. 12, p. 5710, 2023.

[38] L. Taponecco, A. A. D'Amico, and U. Mengali, "Joint toa and aoa estimation for uwb localization applications," *IEEE Transactions on Wireless Communications*, vol. 10, no. 7, pp. 2207–2217, 2011.

[39] S. Bottigliero, D. Milanesio, M. Saccani, and R. Maggiora, "A low-cost indoor real-time locating system based on tdoa estimation of uwb pulse sequences," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[40] C. Falsi, D. Dardari, L. Mucchi, and M. Z. Win, "Time of arrival estimation for uwb localizers in realistic environments," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, pp. 1–13, 2006.

[41] L. Zhang, Y. Li, Y. Gu, and W. Yang, "An efficient machine learning approach for indoor localization," *China Communications*, vol. 14, pp. 141–150, 2017.

[42] J. Kristensen, M. Ginard, O. Jensen, and M. Shen, "Non-line-of-sight identification for uwb indoor positioning systems using support vector machines," in *Proceedings of the IEEE MTT-S International Wireless Symposium (IWS)*. Guangzhou, China: IEEE, May 19–22 2019, pp. 1–3.

[43] M. Kołakowski and J. Modelski, "Detection of direct path component absence in nlos uwb channel," in *Proceedings of the 22nd International Microwave and Radar Conference (MIKON)*. Poznan, Poland: IEEE, May 15–17 2018.

[44] R. Zandian and U. Witkowski, "Differential nlos error detection in uwb-based localization systems using logistic regression," in *Proceedings of the 15th Workshop on Positioning, Navigation and Communications (WPNC)*. Bremen, Germany: IEEE, October 25–26 2018, pp. 1–6.

[45] A. Niitsoo, T. Edelhäußer, and C. Mutschler, "Convolutional neural networks for position estimation in tdoa-based locating systems," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Nantes, France: IEEE, September 24–27 2018, pp. 1–8.

[46] J. Gao, D. Wu, F. Yin, Q. Kong, L. Xu, and S. Cui, "Metaloc: Learning to learn wireless localization," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 12, pp. 3831–3847, 2023.

[47] N. Thakur and C. Y. Han, "Multimodal approaches for indoor localization for ambient assisted living in smart homes," *Information*, vol. 12, no. 3, p. 114, 2021.

[48] F. Potortì, S. Park, A. R. Jimenez Ruiz, P. Barsocchi, M. Girolami, A. Crivello, S. Y. Lee, J. H. Lim, J. Torres-Sospedra, F. Seco *et al.*, "Comparing the performance of indoor localization systems through the evaal framework," *Sensors*, vol. 17, no. 10, p. 2327, 2017.

[49] X. Wang, J. Ellul, and G. Azzopardi, "Elderly fall detection systems: A literature survey," *Frontiers in Robotics and AI*, vol. 7, pp. 1–23, June 2020.

[50] X. Yu, J. Jang, and S. Xiong, "A large-scale open motion dataset (kfall) and benchmark algorithms for detecting pre-impact fall of the elderly using wearable inertial sensors," *Frontiers in Aging Neuroscience*, vol. 13, pp. 1–14, July 2021.

[51] E. Casilari, J.-A. Santoyo-Ramon, and J.-M. Cano-Garcia, "Analysis of public datasets for wearable fall detection systems," *Sensors*, vol. 17, no. 7, p. 1513, June 2017.

[52] Y. Qiu, J. Meng, and B. Li, "Automated falls detection using visual anomaly detection and pose-based approaches: experimental review and evaluation," *J ISSN*, vol. 2766, p. 2276, 2024.

[53] F. Ebrahimi, J. Rousseau, and J. Meunier, "Mobility anomaly detection with intelligent video surveillance," in *World Congress in Computer Science, Computer Engineering & Applied Computing*. Springer, 2024, pp. 189–202.

[54] J. Lee, L. Wang, and A. Wong, "Emotionnet nano: An efficient deep convolutional neural network design for real-time facial expression recognition," *Frontiers in Artificial Intelligence*, vol. 3, pp. 1–9, January 2021.

[55] V. Kulkarni, K. Narayana, and S. K. Sahoo, "A survey on interference avoiding methods for wireless sensor networks working in the 2.4 ghz frequency band." *Journal of Engineering Science & Technology Review*, vol. 13, no. 3, 2020.

[56] A. Poulose and D. S. Han, "Uwb indoor localization using deep learning lstm networks," *Applied Sciences*, vol. 10, no. 18, p. 6290, 2020.

[57] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing k-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 1, p. 113, 2024.

[58] Y. Geng, Q. Li, G. Yang, and W. Qiu, "Support vector machine," in *Practical Machine Learning Illustrated with KNIME*. Springer, 2024, pp. 159–184.

[59] S. González, S. García, J. Del Ser, L. Rokach, and F. Herrera, "A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities," *Information Fusion*, vol. 64, pp. 205–237, 2020.

[60] S. Wang, L. Zhang, X. Wang, W. Huang, H. Wu, and A. Song, "Patchhar: A mlp-like architecture for efficient activity recognition using wearables," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

[61] F. F. Abir, M. E. Chowdhury, M. I. Tapotee, A. Mushtak, A. Khandakar, S. Mahmud, and A. Hasan, "Pcovnet+: A cnn-vae anomaly detection framework with lstm embeddings for smartwatch-based covid-19 detection," *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106130, 2023.

[62] C. Ma, Y. Wang, F. Li, H. Zhang, Y. Zhang, and H. Zhang, "Constructing attention-lstm-vae power load model based on multiple features," *Advances in Mathematical Physics*, vol. 2024, no. 1, p. 1041791, 2024.

**Muhammad Moazam Shahid** received his B.Sc. in Software Engineering with First Class Honours from the University of Management and Technology, Pakistan in 2021 and his M.Sc. in Computer Science with Distinction from Nottingham Trent University in 2023. He is currently a Research Assistant in the Department of Computer Science at Nottingham Trent University. His research interests include embedded systems, health monitoring, activity recognition, indoor positioning, and fall detection using wearable sensor data. His work focuses on developing intelligent, real-time monitoring solutions to support assisted living environments, particularly for elderly care.

**Pedro Machado** received his M.Sc. in Electrical and Computer Engineering from the University of Coimbra in 2012 and his Ph.D. in Computer Science from Nottingham Trent University in 2022. He is currently a Senior Lecturer in Computer Science and the MSc Artificial Intelligence Course Leader at Nottingham Trent University. His main research interests include neuromorphic engineering particularly spiking neural networks and neuromorphic hardware—edge computer vision and intelligent sensors for real-time applications, bio-inspired computing with a focus on retinal cell modelling and biological nervous systems, as well as robotics and autonomous systems, including applications in aquaculture such as monitoring endangered or invasive underwater species.

**Jordan J. Bird** received his BSc (2018) and PhD (2021) degrees in Computer Science from Aston University, Birmingham, UK. He is currently a Senior Lecturer and Deputy B11 REF Co-ordinator at Nottingham Trent University, UK. His research interests include Artificial Intelligence, Human-Computer Interaction, and AI in Education. He has been involved in several international research projects and serves as Associate Editor for the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

**Salisu Wada Yahaya** received his B.Sc. in Computer Science from Nasarawa State University, Keffi, and an M.Sc. in Cybernetics and Communication from Nottingham Trent University. He received his Ph.D. in Computer Science at Nottingham Trent University, where he now serves as a Senior Lecturer in Computer Science and is a member of NTU's Computational Intelligence and Applications research group. His research focuses on applying computational intelligence to human activity recognition, behaviour modelling, and abnormality detection particularly in the context of ambient assisted living and anomaly detection in daily routines.

**Professor Sozo Inoue** received his PhD in Engineering from Kyushu University in 2002. He is currently a Full Professor at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, and Director of the Care XDX Center. His research focuses on web and ubiquitous information systems, human behaviour recognition using smartphones, and medical applications of sensor data, particularly for nursing and caregiving. He has held visiting and advisory roles at institutions including Karlsruhe Institute of Technology, RIKEN, and the University of Los Andes.

**Ahmad Lotfi** (Senior Member, IEEE) received the B.Sc. degree in control systems from the Isfahan University of Technology, Iran, in 1989, the M.Tech. degree in control systems from Indian Institute of Technology Delhi, India, in 1991, and the Ph.D. degree in learning fuzzy systems from the University of Queensland, Brisbane, QLD, Australia, in 1995. He is currently a Professor of Computational Intelligence and Head of the Department of Computer Science at Nottingham Trent University, Nottingham, where he is also leading the Computational Intelligence and Applications (CIA) research group. Areas of his research interest include computational intelligence, ambient intelligence, assistive robotics, and smart homes.

**Isibor Kennedy Ihianle** received his Bachelor's and Ph.D. degrees in Computer Science from the University of East London. He is currently an Associate Professor in the Department of Computer Science at Nottingham Trent University, UK. His research focuses on the application of computational intelligence, artificial intelligence, and machine learning to real-world problems, particularly in the domains of health, well-being, and assistive technologies. He has developed novel intelligent systems for smart environments that support the daily activities and safety of elderly individuals and people with physical or cognitive impairments. Dr. Ihianle's work spans areas including activity recognition, fall detection, sensor data fusion, human-computer interaction, and context-aware systems.