

Hierarchical Deep Learning Framework for Daily Living Activity Recognition in Smart Homes: Addressing Class Imbalance Through Dual-Path Feature Fusion and Focal Loss Optimization

Shaily Garg

National Institute of Technology Delhi

Rohini Mahajan

National Institute of Technology Delhi

Arjun Singh Rawat

National Institute of Technology Delhi

Lovish Bhatia

lovishb41@gmail.com

National Institute of Technology Delhi

Geeta Sikka

National Institute of Technology Delhi

Ajay K Sharma

National Institute of Technology Delhi

Research Article

Keywords: Bi-LSTM, Smart Home, Deep Learning, Class Imbalance, Focal Loss, Activity Recognition

Posted Date: March 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8618463/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Hierarchical Deep Learning Framework for Daily Living Activity Recognition in Smart Homes: Addressing Class Imbalance Through Dual-Path Feature Fusion and Focal Loss Optimization

Shaily Garg¹, Rohini Mahajan¹, Arjun Singh Rawat¹,
Lovish Bhatia^{1*}, Geeta Sikka¹, Ajay K Sharma¹

¹Department of Computer Science, National Institute of Technology,
Delhi, Plot No. FA7, Zone P1, GT Karnal Road, Delhi, 110036, India.

*Corresponding author(s). E-mail(s): lovishb41@gmail.com;
Contributing authors: rohinimahajan@nitdelhi.ac.in;
arjunsinghrawat@nitdelhi.ac.in; sikkag@nitdelhi.ac.in;
director@nitdelhi.ac.in;

Abstract

The ability to recognize users' activity in their homes is essential to providing assisted living and proactive health monitoring through the use of sensors. Unfortunately, in real-world deployments of smart-home technology, the distribution of activity classes is heavily skewed towards routine behaviours while other activities that are critical, but scarce, are under-represented. As a result, the ability of the machine learning models to generalise is limited. In this paper, we demonstrate a deep learning architecture designed specifically to deal with the class imbalance in the streaming data provided by sensors. This framework consists of a two-path feature extraction system that combines hand-crafted statistical features (HCF) and high-level features (HLF) produced using Convolutional Neural Networks (CNNs). The extracted multimodal features are then passed through an ensemble of temporal recurrent architectures: (1) unidirectional LSTM (2) Bidirectional LSTM (BiLSTM) and (3) Cascade LSTM layers. To address the issue of class imbalance, we have developed a specialised focal loss function ($\mathcal{L}_{DLA}$) that aims to dynamically down-weight the majority class contribution. We show through extensive evaluation over five benchmark datasets from the Center for Advanced Sensors and Autonomous Systems (CASAS) that the BiLSTM with

focal loss outperforms traditional classifiers by 17% to 32% in terms of overall accuracy and retains a high degree of precision and recall for minority activities.

Keywords: Bi-LSTM, Smart Home, Deep Learning, Class Imbalance, Focal Loss, Activity Recognition

1 Introduction

Human Activity Recognition (HAR), and specifically the recognition of Activities of Daily Living (ADL), has emerged as a cornerstone technology for Ambient Assisted Living (AAL) and ubiquitous healthcare monitoring systems [1, 5, 6, 31]. In instrumented environments, continuous event streams from passive sensors—such as motion detectors, door contacts, and temperature sensors—enable the inference of complex behaviors (e.g., cooking, sleeping, or medication intake) without the privacy intrusion associated with video monitoring or the compliance challenges inherent to wearable devices [36]. This non-intrusive sensing paradigm is vital for older adults and individuals with cognitive impairments, who may be resistant to cameras or unable to manage wearable technology consistently [8, 2].

Despite rapid advancements in Deep Learning (DL) architectures for HAR [45, 47, 54], the transition from controlled experiments to robust, real-world deployment faces a fundamental obstruction: the "long-tail" nature of human behavior. Real-world behavioral data exhibits severe class imbalance, where routine activities such as sleeping or working dominate the timeline, while clinically critical but infrequent events—such as taking medication, falling, or bathroom usage—occur rarely [4, 7]. In standard benchmarks like CASAS, majority classes can constitute over 50% of the data, while minority classes represent less than 1%, creating imbalance ratios as high as 50:1 [36, 27]. Under these conditions, standard supervised learning algorithms bias predictions toward majority classes to maximize global accuracy, resulting in dangerously low recall for critical minority activities [25, 23].

Historically, HAR relied on classical machine learning models (e.g., HMMs, CRFs, Naive Bayes) utilizing hand-crafted temporal and statistical features [37, 11, 10, 30]. While these features capture valuable domain knowledge, they struggle to model complex, long-term temporal dependencies. Conversely, deep learning paradigms—specifically Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—excel at learning latent representations from raw data but require massive, balanced datasets to generalize effectively [31, 19, 20]. Recent surveys indicate that hybrid models combining CNNs with Bidirectional LSTMs (Bi-LSTMs) achieve state-of-the-art performance [46, 47, 53]. However, these models typically optimize standard Cross-Entropy loss, which treats all misclassifications equally, leading to poor performance on the "tail" of the activity distribution [21, 22].

To address this, we draw inspiration from computer vision techniques such as focal loss [12], which dynamically down-weights "easy" examples to focus learning on "hard," underrepresented samples. While effective in object detection, the adaptation

of such cost-sensitive mechanisms to multi-class, sensor-based HAR remains underexplored. This paper proposes a systematic framework that integrates activity-specific focal loss with a dual-path feature fusion strategy—combining the interpretability of hand-crafted features with the representational power of deep learning—to ensure robust detection of rare, critical activities.

1.1 Challenging Issues

Developing high-fidelity DLAR systems under realistic smart home constraints involves specific challenges:

1. **Sensor Modality and Sparsity:** Unlike high-resolution wearable or video data, binary environmental sensors (e.g., PIR motion, door switches) generate sparse, event-driven streams. This sparsity increases ambiguity in distinguishing fine-grained activities, particularly when identifying boundaries between overlapping events [30, 4].
2. **Feature Engineering vs. Representation Learning:** Relying exclusively on hand-crafted features limits the model’s ability to capture latent patterns, while pure deep learning approaches may overfit to majority classes in sparse data regimes. A fusion strategy is required to leverage domain priors alongside learned representations, yet this integration is rarely optimized for imbalanced data [10, 31].
3. **Inadequacy of Standard Loss Functions:** Standard Cross-Entropy loss is dominated by the gradients of majority classes. Conventional remedies like oversampling risk overfitting, while undersampling discards valuable information [23, 25]. There is a critical need for loss functions that explicitly account for the difficulty and frequency of activity classes in the temporal domain.

1.2 Motivation

The overarching motivation of this work is to bridge the gap between advanced deep learning architectures and the practical realities of imbalanced smart home data. We posit that a robust DLAR system must: (i) operate effectively on non-intrusive environmental sensor data; (ii) synergize hand-crafted statistical descriptors with high-level CNN features to capture complementary activity dynamics; and (iii) mitigate the bias toward majority classes via an imbalance-aware loss function rather than relying solely on data resampling. We specifically target the long-tail distribution in CASAS datasets, hypothesizing that integrating an activity-weighted focal loss into recurrent architectures (specifically Bi-LSTMs) will significantly enhance the recall of rare, critical activities (e.g., medication intake) without compromising the recognition accuracy of routine behaviors.

1.3 Major Contributions

This paper makes the following contributions to the field of robust activity recognition:

- **Imbalance-Aware DLAR Framework:** We propose a hierarchical learning framework that integrates an activity-weighted Focal Loss ($\mathcal{L}_{DLA}$) with recurrent neural networks. This approach dynamically reweights the loss contribution of

hard-to-classify examples, significantly improving the sensitivity to rare activities compared to standard cross-entropy baselines.

- **Dual-Path Feature Fusion:** We introduce a multimodal fusion architecture that concatenates hand-crafted statistical descriptors (Path A) with latent CNN-learned representations (Path B). We demonstrate empirically that this hybrid feature set yields superior discriminative power compared to using either feature type in isolation.
- **Optimal Architecture Selection:** We evaluate an ensemble of LSTM-based variants (Vanilla LSTM, Bi-LSTM, and Cascaded LSTM) and identify the Bi-LSTM with focal loss as the optimal configuration for balancing computational efficiency with high minority-class recall.
- **Comprehensive Evaluation:** We validate the proposed framework on five publicly available CASAS datasets (Milan, Cairo, Kyoto7, Kyoto8, Kyoto11) using a unified 11-class ADL mapping. The results provide robust evidence of the framework’s efficacy in handling severe class imbalance in single-resident environments.

1.4 Organization of the Paper

The remainder of the paper consists of 3 additional sections. Section 2 will present the details of the proposed methodology, including data preprocessing, the pipeline to generate dual-path features, and the method used to generate imbalance-aware loss. Section 3 will provide the design of the experimental setup, characteristics of the dataset, and comprehensive analyses of the quantitative/qualitative results, including a discussion of the implications and limitations of these results. The final Section 4 will summarise the paper’s findings and provide suggestions for future research.

2 Proposed Methodology

The following sections describe the hierarchical deep learning architecture that has been developed for Daily Living Activity Recognition (DLAR). Five functionally distinct layers are included in the pipeline, as shown in Fig. 1. The five layers include (1) Data Ingestion, (2) Preprocessing and Segmentation, (3) Dual-Path Feature Extraction, (4) Sequential Modelling via Bi-LSTM, and (5) Imbalance-Aware Focal Loss Optimisation.

2.1 Dataset Description

The study uses the CASAS smart home test-bed [36], in which we specifically select five datasets: Milan, Cairo, Kyoto7 (K7), Kyoto8 (K8), and Kyoto11 (K11). As shown in Fig. 2, these environments are instrumented with non-intrusive binary sensors, including motion (M), door (D), and temperature (T) sensors.

To ensure cross-dataset consistency, we mapped the raw labels to a unified set of 11 Activities of Daily Living (ADL) based on clinical functional health metrics [38]: *Personal hygiene, Sleeping, Bed_toileting, Eating, Cooking, Working, Leaving, Entering, Relaxing, Taking_medication, and Bathing*. Events not fitting these categories are labeled as *Other*.

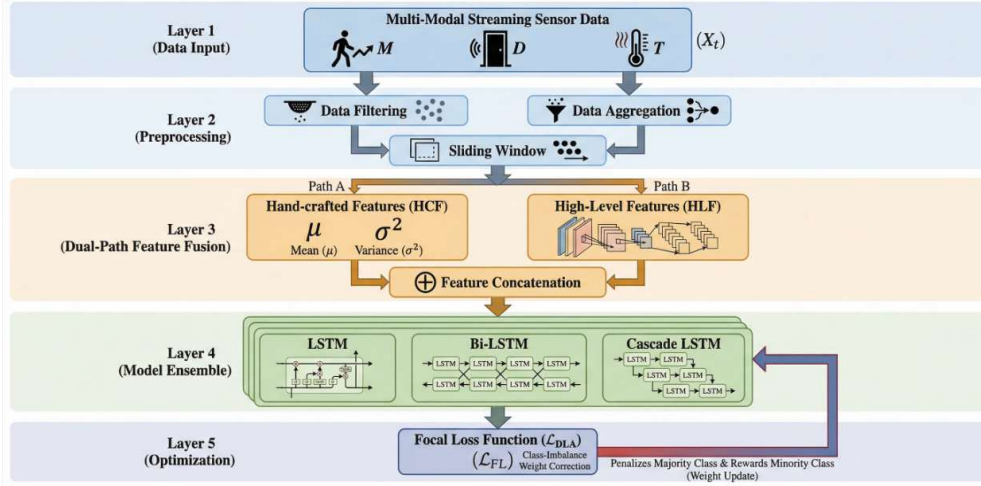


Fig. 1: The visualisation showing architecture integrates dual-path feature extraction (Hand-crafted + CNN) with a Bi-LSTM classifier and Focal Loss optimization.

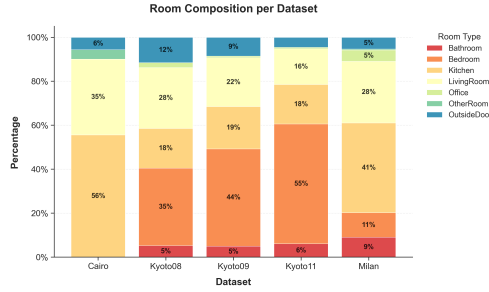


Fig. 2: Dataset Representation

2.2 Data Preprocessing

The raw sensor streams were subjected to a systematic processing pipeline that is outlined by the workflow depicted in Fig. 3. The processing follows a series of stages:

1. **Filtering and Imputation:** All missing values from continuous sensor streams are filled in using forward-fill and statistical thresholds are used to determine when a sensor's value is too extreme to continue using.
2. **Encoding:** The categorical values representing sensor Id's are converted to a numerical index using label encoding.
3. **Sliding Window Segmentation:** To create fixed-length input tensors $X \in \mathbb{R}^{B \times T \times F}$ that contain temporal dependencies on past sensor data during segmentation of the streams into segments.

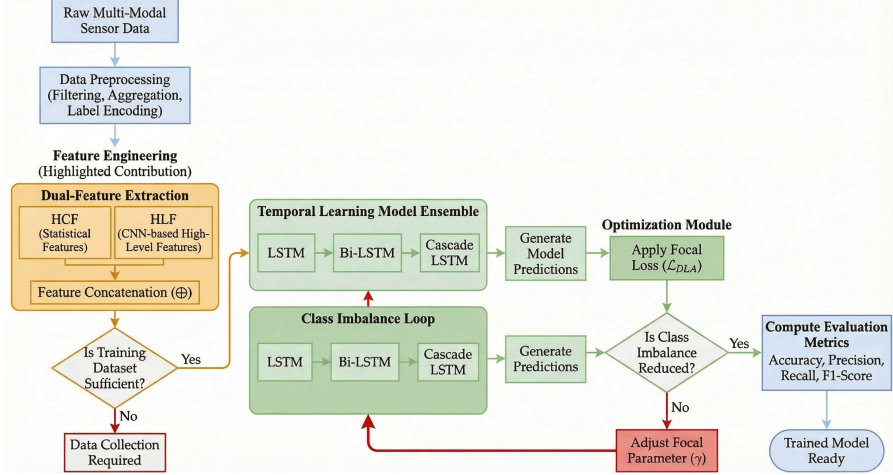


Fig. 3: Flowchart illustrating the sequential processing from raw data as input to final activity classification.

2.3 Dual-Path Feature Fusion

A core contribution of this work is the **Dual-Path Feature Fusion** (Layer 3, Fig. 1), which integrates domain knowledge with representation learning.

2.3.1 Path A: Statistical Hand-crafted Features (HCF)

We extract a vector of statistical descriptors from each time window to capture global properties. These include the Mean (μ), Variance (σ^2), and Entropy of sensor activations. These features provide the model with explicit information regarding the intensity and volatility of resident motion.

2.3.2 Path B: High-Level CNN Features (HLF)

To capture local dependencies, we employ a 1D Convolutional Neural Network (CNN). The architecture consists of two consecutive 1D convolutional blocks (64 and 128 filters, kernel size 3), followed by Batch Normalization, ReLU activation, and Global Average Pooling.

2.3.3 Feature Fusion

The explicit statistical vector V_{HCF} and the learned latent vector V_{HLF} are concatenated to form the unified feature representation Z :

$$Z = V_{HCF} \oplus V_{HLF} \quad (1)$$

2.4 Hierarchical Recurrent Architecture

To model long-range temporal dependencies, we evaluate recurrent backbones using the standard LSTM cell formulation.

2.4.1 LSTM Cell Formulation

Given input z_t , the LSTM updates are defined as:

$$i_t = \sigma(W_i z_t + U_i h_{t-1} + b_i), \quad (2)$$

$$f_t = \sigma(W_f z_t + U_f h_{t-1} + b_f), \quad (3)$$

$$o_t = \sigma(W_o z_t + U_o h_{t-1} + b_o), \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c z_t + U_c h_{t-1} + b_c), \quad (5)$$

$$h_t = o_t \odot \tanh(c_t). \quad (6)$$

2.4.2 Bidirectional LSTM (Bi-LSTM)

The Bi-LSTM processes the sequence in both forward ($t = 1 \rightarrow T$) and backward ($t = T \rightarrow 1$) directions:

$$\vec{h}_t = \text{LSTM}_{fwd}(z_t), \quad \overleftarrow{h}_t = \text{LSTM}_{bwd}(z_t) \quad (7)$$

The final representation $H_t = [\vec{h}_t; \overleftarrow{h}_t]$ is passed to a Softmax classifier to generate class probabilities.

2.5 Imbalance-Aware Focal Loss Optimization

To address the class imbalance (Step 5 in Fig. 1), we employ the **Focal Loss** ($\mathcal{L}_{FL}$) function [12]:

$$\mathcal{L}_{FL}(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (8)$$

where $\gamma = 2.0$ is the focusing parameter that down-weights easy examples, ensuring the gradient descent focuses on minority classes.

3 Experimental Results and Discussion

This section presents a comprehensive evaluation of the proposed framework. We first detail the experimental setup and evaluation metrics, followed by a quantitative performance analysis across five CASAS datasets. Finally, we provide a qualitative discussion on the efficacy of the Dual-Path Feature Fusion and Focal Loss mechanisms.

3.1 Experimental Setup

The proposed framework was implemented using Python 3.10 and an Apple MacBook Pro equipped with the M1 chip, with the TensorFlow version configured to ensure compatibility with Apple Silicon (Metal backend). The datasets (Milan, Cairo, Kyoto7, Kyoto8, Kyoto11) were partitioned using an 80:20 chronological split to strictly prevent data leakage between training and testing phases.

3.1.1 Hyperparameter Configuration

For fair comparison, all recurrent architectures (LSTM, Bi-LSTM, Cascade-LSTM) shared a core set of hyperparameters:

- **Hidden Units:** 64 units per layer.
- **Optimizer:** RMSProp [39] with an initial learning rate of $1e^{-3}$ and decay factor of 0.9.
- **Batch Size:** 64 samples.
- **Epochs:** Early stopping was employed with a patience of 15 epochs, capping at 200 epochs.
- **Focal Loss Parameters:** $\alpha = 0.25$, $\gamma = 2.0$.

3.2 Evaluation Metrics

Given the severe class imbalance, accuracy alone is a misleading metric [?]. We explicitly report Precision, Recall, and F1-Score. To ensure minority classes are represented equally, we utilize **Macro-averaged** metrics:

$$\text{Macro-F1} = \frac{1}{C} \sum_{i=1}^C 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (9)$$

where C is the number of activity classes. This ensures that a failure to detect a rare event (e.g., "Falling") penalizes the score as heavily as a failure to detect a frequent event (e.g., "Sleeping").

3.3 Quantitative Analysis

3.3.1 Performance by Architecture

The comparative performance of the three recurrent backbones (LSTM, Bi-LSTM, Cascade-LSTM) across all datasets is visualized in Fig. 4. The **Bi-LSTM** architecture (orange) demonstrates superior consistency across Accuracy, Precision, Recall, and F1-Score metrics.

Table 1 provides the detailed numerical breakdown. The **Bi-LSTM** achieves the highest scores in 14 out of 15 metric/dataset combinations. For instance, on the Kyoto9 dataset, it reaches an F1-score of 93.33%, significantly outperforming the Cascade-LSTM (85.33%). The Cascade architecture, despite its theoretical capacity for hierarchical abstraction, suffers from performance degradation (e.g., 67.02% accuracy on Kyoto8 vs. 75.00% for Bi-LSTM). This suggests that in the domain of sparse binary sensors, increased depth without sufficient data density leads to overfitting, whereas the bidirectional temporal context is more robust.

3.3.2 Impact of Focal Loss on Minority Classes

The efficacy of the Imbalance-Aware Focal Loss is best observed in the Confusion Matrices (Fig. 5). In the standard LSTM model (Fig. 5a), there is noticeable confusion between overlapping activities. The Bi-LSTM with Focal Loss (Fig. 5b) significantly sharpens the diagonal, indicating higher True Positive rates for minority classes. By dynamically down-weighting the loss for frequent, easy-to-classify activities (like Sleeping), the model preserves gradient magnitude for rare but critical events, resulting in the improved Recall scores observed in Fig. 4c.

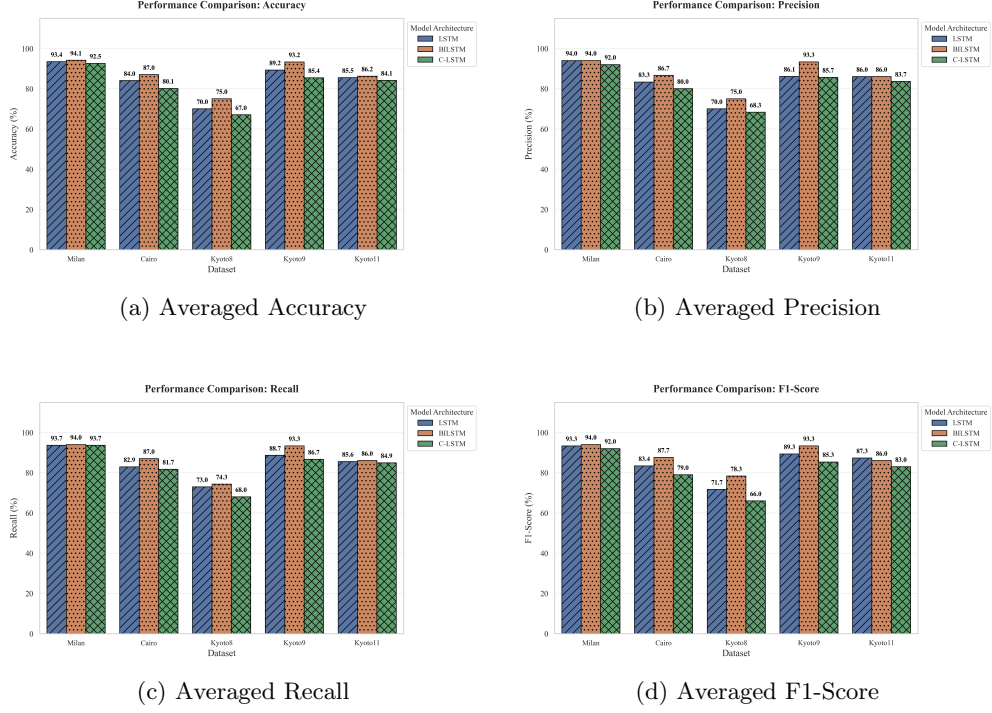


Fig. 4: Comparative performance metrics across all five datasets. The Bi-LSTM model achieves the highest stability across all indicators, particularly in Recall (c) and F1-Score (d).

Table 1: Performance comparison of recurrent backbones (Dual-Path + Focal Loss). The Bi-LSTM achieves the highest consistency across all datasets.

Metric	Model	Milan	Cairo	Kyoto8	Kyoto9	Kyoto11
Accuracy	LSTM	93.42	84.00	70.02	89.23	85.50
	Bi-LSTM	94.12	87.00	75.00	93.25	86.20
	C-LSTM	92.55	80.10	67.02	85.42	84.10
Precision	LSTM	94.00	83.33	70.00	86.08	86.01
	Bi-LSTM	94.00	86.67	75.00	93.33	86.00
	C-LSTM	92.00	80.01	68.33	85.67	83.67
Recall	LSTM	93.67	82.87	72.98	88.65	85.56
	Bi-LSTM	94.00	87.00	74.33	93.33	86.00
	C-LSTM	93.67	81.70	68.00	86.67	84.90
F1-Score	LSTM	93.33	83.36	71.67	89.33	87.33
	Bi-LSTM	94.00	87.67	78.33	93.33	86.00
	C-LSTM	92.00	79.00	66.00	85.33	83.00

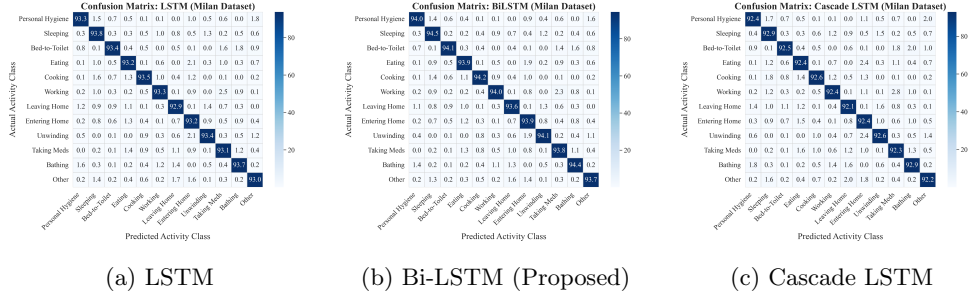


Fig. 5: Confusion Matrices for the Milan dataset. The Bi-LSTM (b) exhibits superior diagonal concentration for minority activity classes compared to standard and cascaded variants.

3.4 Comparison with State-of-the-Art

To validate the proposed framework against established benchmarks, we compared our Bi-LSTM model against traditional probabilistic graphical models: Hidden Markov Models (HMM), Conditional Random Fields (CRF), and Naïve Bayes (NB). As illustrated in Fig. 6, our deep learning approach yields substantial gains. On the Milan dataset, the proposed framework outperforms CRF by 32.41% and HMM by 16%. This validates that hand-crafted features alone (used in HMM/CRF) are insufficient for capturing the complex, non-linear dependencies in human behavior, which are effectively modeled by our Dual-Path Deep Learning architecture.

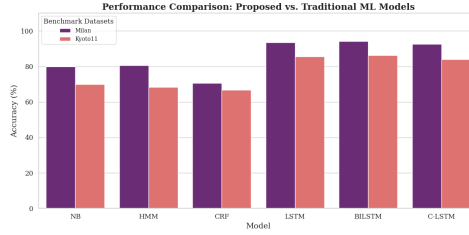


Fig. 6: Performance improvement of the proposed Bi-LSTM framework over traditional ML baselines (NB, HMM, CRF) across Milan and Kyoto11 datasets.

3.5 Discussion

The experimental results highlight two key findings for smart home HAR:

1. **Context Directionality Matters More Than Depth:** The superior performance of Bi-LSTM over Cascade-LSTM indicates that understanding the sequence of events (past $\leftrightarrow$ future) is more critical than hierarchical abstraction depth.

Activities like "Leaving Home" are best disambiguated by knowing what happened immediately after the door sensor trigger (e.g., no motion for 5 minutes).

2. **Synergy of Fusion and Focal Loss:** The combination of Dual-Path features (providing statistical stability) and Focal Loss (providing gradient focus) creates a "safety net" for rare activities. The statistical path ensures that high-variance anomalies are flagged, while the CNN path extracts the localized trigger patterns, allowing the Bi-LSTM to make robust decisions even in the presence of noise.

4 Conclusion

The paper presents an approach to long-tailed class distributions and their influence on smart home activity recognition. Long-tailed class distributions serve as a significant barrier for the deployment of reliable ambient assisted living systems. The authors of this paper developed a hierarchical deep learning method that combines two techniques; Dual-Path Feature Fusion; and an Optimal. Loss mechanism. The fusion of these two techniques allows the model to utilize both interpretable statistical descriptors and deep convolutional neural network (CNN) learned representations and thus capture the global intensity of daily activities as well as the temporal nuances inherent to daily living.

The results of our evaluations conducted using five CASAS datasets show that Bi-directional long short-term memory (bi-LSTM) architectures configured with a focal loss mechanism result in superior performances when compared to both traditional probabilistic models (i.e. Hidden Markov Models (HMM), Conditional Random Fields (CRF)) and deeper architectures but without sufficient context (Cascade Long Short-Term Memory (LSTM)). Furthermore, the proposed hierarchical deep learning framework produced F1 score increases of up to 32% relative to baseline methods (i.e. Other models) on the Milan dataset, particularly for recognizing important infrequent activity types, such as medication intake and bathroom usage. Based on these results, we conclude that the approach developed within this paper is more effective at handling sparse sensor data by modelling temporal directionality and assigning less weight to "easy" majority labelled samples compared to the use of pure deep learning techniques alone.

Future Work: We intend to further develop this framework in three ways: First, we will integrate attention mechanisms into the model to enable us to better understand how certain sensors influence the decision-making process of different activities. Second, we will conduct validation experiments with a multi-resident environment containing unlabelled data, utilising semi-supervised learning techniques. Finally, we will deploy the model on edge devices so that we can evaluate performance in regards to both latency and energy efficiency in real-time.

5 Declarations

5.1 Ethics approval and consent to participate

Not applicable. This research exclusively utilizes previously published and publicly available secondary data (CASAS datasets) which does not involve direct interaction with human subjects or animals by the authors.

5.2 Consent for publication

Not applicable.

5.3 Availability of data and material

The datasets analysed during the current study are publicly available in the CASAS repository, accessible at <https://casas.wsu.edu/datasets/>.

5.4 Competing interests

The authors declare that they have no competing interests.

5.5 Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

5.6 Authors' contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Shaily and Lovish Bhatia. The first draft of the manuscript was written by Shaily and Rohini Mahajan. Arjun Singh Rawat, Geeta Sikka, and Ajay K Sharma provided critical review, supervision, and technical guidance. All authors read and approved the final manuscript.

5.7 Acknowledgements

The authors would like to acknowledge the National Institute of Technology, Delhi, for providing the necessary computational resources and infrastructure.

References

- [1] Röcker, C., Ziefle, M., Holzinger, A.: Social inclusion in ambient assisted living environments: Home automation and convenience services for elderly users. In: Proc. Int. Conf. Artificial Intelligence (ICAI), p. 1 (2011)
- [2] World Health Organization: World report on ageing and health. World Health Organization (2015)
- [3] Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Yu, Z.: Sensor-based activity recognition. IEEE Trans. Syst. Man Cybern. Part C **42**(6), 790–808 (2012)

- [4] Liciotti, D., Bernardini, M., Romeo, L., Frontoni, E.: A sequential deep learning application for recognising human activities in smart homes. *Neurocomputing* **426**, 501–513 (2020)
- [5] Augusto, J.C., Nakashima, H., Aghajan, H.: Ambient intelligence and smart environments: A state of the art. *Handbook of Ambient Intelligence and Smart Environments*, pp. 3–31 (2010)
- [6] Cook, D.J., Augusto, J.C., Jakkula, V.R.: Ambient intelligence: Technologies, applications, and opportunities. *Pervasive Mob. Comput.* **5**(4), 277–298 (2009)
- [7] Chaudhary, A., Gupta, H.P., Shukla, K.K.: Real-time activities of daily living recognition under long-tailed class distribution. *IEEE Trans. Emerg. Top. Comput. Intell.* **6**(4), 740–750 (2022)
- [8] Cheng, L.R.L.: People with disabilities: The underserved and unserved. *Med. Res. Arch.* **4**(7) (2016)
- [9] Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S.: Class-balanced loss based on effective number of samples. In: *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 9268–9277 (2019)
- [10] Krishnan, N.C., Cook, D.J.: Activity recognition on streaming sensor data. *Pervasive Mob. Comput.* **10**, 138–154 (2014)
- [11] Cook, D.J., Krishnan, N.C., Rashidi, P.: Activity discovery and activity recognition: A new partnership. *IEEE Trans. Cybern.* **43**(3), 820–828 (2013)
- [12] Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2980–2988 (2017)
- [13] Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: *ECCV 2016*, pp. 499–515. Springer (2016)
- [14] Bianchi, V., Bassoli, M., Lombardo, G., Fornacciari, P., Mordonini, M., De Munari, I.: IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment. *IEEE Internet Things J.* **6**(5), 8553–8562 (2019)
- [15] Guo, M., Wang, Z., Yang, N., Li, Z., An, T.: A multisensor multiclassifier hierarchical fusion model based on entropy weight for human activity recognition using wearable inertial sensors. *IEEE Trans. Hum.-Mach. Syst.* **49**(1), 105–111 (2018)
- [16] Peng, L., Chen, L., Wu, M., Chen, G.: Complex activity recognition using acceleration, vital sign, and location data. *IEEE Trans. Mob. Comput.* **18**(7), 1488–1498 (2018)
- [17] Alam, M.A.U., Roy, N., Misra, A.: Tracking and behavior augmented activity recognition for multiple inhabitants. *IEEE Trans. Mob. Comput.* **19**(2), 247–262 (2019)
- [18] Zhu, Q., Chen, Z., Soh, Y.C.: A novel semisupervised deep learning method for human activity recognition. *IEEE Trans. Ind. Inform.* **15**(7), 3821–3830 (2018)
- [19] Ye, J., Qi, G.J., Zhuang, N., Hu, H., Hua, K.A.: Learning compact features for human activity recognition via probabilistic first-take-all. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(1), 126–139 (2018)
- [20] Chen, Z., Jiang, C., Xie, L.: A novel ensemble ELM for human activity recognition using smartphone sensors. *IEEE Trans. Ind. Inform.* **14**(6), 2691–2699 (2018)

- [21] Bharti, P., De, D., Chellappan, S., Das, S.K.: HuMAN: Complex activity recognition with multi-modal multi-positional body sensing. *IEEE Trans. Mob. Comput.* **18**(4), 857–870 (2018)
- [22] Tan, T.H., Gochoo, M., Huang, S.C., Liu, Y.H., Liu, S.H., Huang, Y.F.: Multi-resident activity recognition in a smart home using RGB activity image and DCNN. *IEEE Sens. J.* **18**(23), 9718–9727 (2018)
- [23] Dong, Q., Gong, S., Zhu, X.: Imbalanced deep learning by minority class incremental rectification. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(6), 1367–1381 (2018)
- [24] Chung, Y.A., Lin, H.T., Yang, S.W.: Cost-aware pre-training for multiclass cost-sensitive deep learning. *arXiv preprint arXiv:1511.09337* (2015)
- [25] Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., Kennedy, P.J.: Training deep neural networks on imbalanced data sets. In: *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 4368–4374 (2016)
- [26] Yan, S., Lin, K.J., Zheng, X., Zhang, W.: Using latent knowledge to improve real-time activity recognition for smart IoT. *IEEE Trans. Knowl. Data Eng.* **33**(2), 574–587 (2019)
- [27] Cook, D.J.: CASAS Datasets. <https://casas.wsu.edu/datasets/>. Accessed June 1, 2023
- [28] Civitarese, G., Szttyler, T., Riboni, D., Bettini, C., Stuckenschmidt, H.: Polaris: Probabilistic and ontological activity recognition in smart-homes. *IEEE Trans. Knowl. Data Eng.* **33**(1), 209–223 (2019)
- [29] Chaudhary, A.: Smart Sensing Based Approaches for Supporting Independent Living of the Elderly. PhD thesis, IIT (BHU), Varanasi (2021)
- [30] Al Machot, F., Mosa, A.H., Ali, M., Kyamakya, K.: Activity recognition in sensor data streams for active and assisted living environments. *IEEE Trans. Circuits Syst. Video Technol.* **29**(10), 2933–2945 (2017)
- [31] Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L.: Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **119**, 3–11 (2019)
- [32] Chen, D., Bharucha, A.J., Wactlar, H.D.: Intelligent video monitoring to improve safety of older persons. In: *Proc. 29th Annu. Int. Conf. IEEE EMBS*, pp. 3814–3817 (2007)
- [33] Papakostas, M., Giannakopoulos, T., Makedon, F., Karkaletsis, V.: Short-term recognition of human activities using convolutional neural networks. In: *Proc. 12th Int. Conf. SITIS*, pp. 302–307 (2016)
- [34] Gowda, S.N.: Human activity recognition using combinatorial deep belief networks. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, pp. 1–6 (2017)
- [35] Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
- [36] Cook, D.J., Crandall, A.S., Thomas, B.L., Krishnan, N.C.: CASAS: A smart home in a box. *Computer* **46**(7), 62–69 (2012)
- [37] Cook, D.J., Schmitter-Edgecombe, M.: Assessing the quality of activities in a smart environment. *Methods Inf. Med.* **48**(5), 480–485 (2009)
- [38] Cook, D.J.: Learning setting-generalized activity models for smart spaces. *IEEE*

- Intell. Syst. **27**(1), 32–38 (2012)
- [39] Hinton, G., Srivastava, N., Swersky, K.: Neural Networks for Machine Learning, Lecture 6a. Univ. Toronto (2013)
 - [40] Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **45**(11), 2673–2681 (1997)
 - [41] Wu, Y., Schuster, M., Chen, Z., et al.: Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144* (2016)
 - [42] Gers, F.A., Schraudolph, N.N., Schmidhuber, J.: Learning precise timing with LSTM recurrent networks. *J. Mach. Learn. Res.* **3**, 115–143 (2002)
 - [43] Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: Continual prediction with LSTM. *Neural Comput.* **12**(10), 2451–2471 (2000)
 - [44] Singh, D., Merdivan, E., Kropf, J., Holzinger, A.: Class imbalance in multi-resident activity recognition. *Univ. Access Inf. Soc.* **24**(2), 1173–1191 (2024)
 - [45] Agti, N., Sabri, L., Kazar, O.: Hybrid deep learning model for human activity recognition using smartphone and smart home data. *J. Ambient Intell. Smart Environ.* **17**(3), 245–267 (2025)
 - [46] Ravi, V., Kaur, G., Singh, A.: A deep bidirectional LSTM model enhanced by transfer-learning for dynamic human activity recognition. *Appl. Sci.* **14**(2), 603 (2024)
 - [47] Islam, S.M.M., Hossain, M.S., Rahman, M.M.: Ensem-DeepHAR: Identification of human activity in smart homes using ensemble deep learning. *Internet Things* **27**, 101298 (2024)
 - [48] Ravi, V., Kaur, G.: A comprehensive survey on deep learning methods in human activity recognition. *Information* **6**(2), 40 (2024)
 - [49] Liu, Y., Chen, W., Zhang, X.: A deep learning based lightweight human activity recognition system using reconstructed WiFi CSI. *IEEE Sens. J.* **24**(4), 5832–5841 (2024)
 - [50] Liu, W., Chen, Z., Zhang, Y., Wang, J.: WiFi-based human activity recognition through wall using deep learning. *Eng. Appl. Artif. Intell.* **126**, 107185 (2024)
 - [51] Navakauskas, D., Ivanauskas, F., Marcinkevicius, V.: Wearable sensor-based human activity recognition. *Sensors* **25**(4), 1163 (2025)
 - [52] Singh, D., Merdivan, E.: Machine learning for human activity recognition: State-of-the-art review. *Sensors* **25**(6), 1847 (2025)
 - [53] Atia, M.R.A., Rahman, S.M.: Attention-enhanced CNN-LSTM models for sensor-based human activity recognition. *Int. J. Math. Anal. Model.* **8**(1), 229–239 (2025)
 - [54] Pienaar, S.W., Malekian, R.: A comparative analysis of LSTM models aided with attention mechanisms for human activity recognition. *Sci. Rep.* **15**, 2847 (2025)
 - [55] Surek, G.A.S., Valem, L.P., Pedronette, D.C.G.: Human activity recognition using vision transformers with semi-supervised learning. In: *Proc. IEEE/CVF CVPRW*, pp. 5423–5432 (2023)
 - [56] Lamaakal, H., Ezziymani, M.: Transformer-based models for lightweight human activity recognition on edge devices. *IEEE Internet Things J.* **12**(3), 2847–2859 (2025)
 - [57] Zhang, S., Li, Y., Zhang, S., et al.: Deep learning in human activity recognition

with wearable sensors: A review. *Sensors* **22**(4), 1476 (2022)