2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

DeepSense++: Robust HAR with Missing Data

Suryakangeyan Kandasamy Gowdaman[1] and Sayma Akther[2]
San Jose State University, California, USA
suryakangeyan.kandasamygowdaman@sjsu.edu[1], sayma.akther@sjsu.edu[2]

Abstract— **Human Activity Recognition (HAR) using wearable sensors is increasingly applied in healthcare, sports, and intelligent environments. Performance is however hindered in the majority of the cases by absent sensor values, class imbalance, and inter-subject variability. We present a robust HAR pipeline that utilizes Principal Component Analysis (PCA) for reducing dimensions and Generative Adversarial Networks (GANs) for realistic imputation of absent values and minority-class oversampling. This is integrated into an improved DeepSense architecture with convolutional and recurrent layers for spatial–temporal feature learning. Comparisons on the OPPORTUNITY dataset, in terms of K-Fold, Leave-One-Session-Out (LOSEO), and Leave-One-Subject-Out (LOSO) schemes, demonstrate improved accuracy (+3.7%) and F1 score (+2.9%) over baseline DeepSense. The results highlight the applicability of hybrid imputation-augmentation pipelines in bringing HAR to practical, noisy sensing scenarios.**

Keywords—Wearable sensing, human activity recognition (HAR), missing data recovery, principal component analysis (PCA), generative adversarial networks (GANs), data augmentation, imputation techniques, DeepSense architecture, deep learning models.

I. INTRODUCTION

Wearable sensor–based Human Activity Recognition (HAR) supports applications in healthcare, sports analytics, smart environments, and human–computer interaction [1], [2]. These systems analyze continuous multivariate time-series signals from embedded devices such as accelerometers, gyroscopes, and magnetometers to classify activities ranging from simple postures to complex motions [3], [4].

While deep learning has advanced HAR performance, practical deployment faces three persistent challenges: (1) **Limited and imbalanced labeled data**—rare but critical activities (e.g., falls) are underrepresented; (2) **Incomplete inputs**—missing values arise from sensor dropout, misplacement, or transmission errors; (3) **Variability across users and contexts**—motion signatures differ significantly across individuals and environments. Collecting large, diverse, high-quality datasets requires extensive annotation and long-term deployments, and public datasets often suffer from noisy labels, restricted activity coverage, or missing synchronized ground truth.

Data collection settings also impact model generalization. Scripted recordings provide reproducibility but limited variability, semi-scripted setups offer moderate diversity, and wild settings capture natural behaviors but introduce label noise and synchronization issues.

A. Proposed Approach

We address incompleteness and imbalance by integrating Principal Component Analysis (PCA)–based imputation with Generative Adversarial Network (GAN)–based augmentation into a unified HAR pipeline. PCA exploits the low-rank structure of human activity patterns to restore missing segments, while GANs synthesize realistic activity-specific sequences to enrich minority classes and improve robustness [5]–[7]. Conditional GAN variants further allow controlled generation aligned with target activity labels [8]–[10]. We formalize the proposed pipeline as:

$$X_{\text{final}} = X_{\text{real}} \,\cup\, \mathcal{G}_\theta\big(\mathrm{PCA}(X_{\text{real}})\big), \tag{1}$$

where:

- X_{real}: Original multimodal sensor data (incomplete, imbalanced).
- $\mathrm{PCA}(\cdot)$: Low-rank imputation to restore missing segments.
- $\mathcal{G}_\theta(\cdot)$: Conditional GAN generating activity-specific synthetic data.
- X_{final}: Augmented dataset combining real and synthetic samples.

This unified PCA–GAN strategy jointly recovers missing sensor streams, balances class distributions, and enhances model robustness for real-world HAR deployments.

II. RELATED WORK

Human Activity Recognition (HAR) using wearable sensors has evolved from traditional feature-engineered methods to deep generative and self-supervised learning approaches. Early benchmarks relied on classifiers such as Support Vector Machines (SVMs), Decision Trees, and k-Nearest Neighbors (k-NN) [5], which required handcrafted features and domain heuristics. While effective on small, clean datasets, these models performed poorly with high-dimensional, noisy, or incomplete time-series data.

Deep learning architectures—particularly convolutional and recurrent networks—have since become the standard for HAR. Temporal Convolutional Networks (TCNs) [7] capture long-range dependencies in activity sequences, and Ignatov [6] demonstrated a CNN-based model for real-time accelerometer-based recognition. However, such models often assume fully observed signals and are vulnerable to performance degradation under sensor dropout or severe class imbalance.

*This work was completed in partial fulfillment of the requirements for the M.S. degree of [1]Suryakangeyan Kandasamy Gowdaman in Computer Science at San José State University. [2]Sayma Akther is the corresponding author.

Generative modeling has been proposed to address these limitations. ActivityGAN [9] uses adversarial learning to reconstruct temporal dynamics, while SensorGAN [8] synthesizes fine-grained sensor streams for augmentation. Despite promising gains, these methods can suffer from training instability and reduced fidelity. Our approach extends this direction by combining GAN-based imputation with PCA-based compression to produce more structured and realistic reconstructions of missing segments.

Adversarial learning has also been applied to domain adaptation and invariant feature extraction. Bai *et al.* [10] introduced a multi-view adversarial network to generalize across sensor placements, while XHAR [11] and Wang *et al.* [12] applied adversarial transfer learning to reduce domain shift. These strategies improve cross-context generalization but do not directly address incomplete data—a core focus of our work.

Self-supervised learning (SSL) methods have recently shown promise for HAR. Haresamudram *et al.* [13] applied contrastive predictive coding for downstream classification, and Song *et al.* [14] explored contrastive pretraining for ambient sensing. Su *et al.* [15] proposed disentangled behavior patterns to enhance transferability and interpretability. While SSL reduces the dependence on large labeled datasets, performance declines when missing data is prevalent.

Finally, advanced augmentation techniques have emerged. Li *et al.* [16] introduced a statistical diffusion model for high-fidelity signal generation, and Um *et al.* [17] developed AutoAugHar, an automated reinforcement learning–driven augmentation framework. Our work complements these by unifying temporal augmentation and missing-data imputation in a single GAN–PCA module, integrated with a DeepSense [18] based classifier to jointly address incompleteness, imbalance, and temporal variability.

III. Benchmark Datasets and Processing Workflow

A. Primary Benchmark: OPPORTUNITY [19]

The OPPORTUNITY dataset, captured in a sensor-rich environment, includes synchronized data streams from body-worn IMUs, ambient sensors, and object-mounted devices. Four subjects (S_1–S_4) performed daily activities, including locomotion (`Stand`, `Walk`, `Sit`, `Lie`) and object-use tasks. Data were recorded at $f_s = 30$ Hz across > 200 channels, yielding over 10^6 labeled samples with realistic noise and dropout.

B. Data Preparation Workflow

The raw multichannel sequence $X \in \mathbb{R}^{T \times D}$ undergoes the transformation:

$$\hat{X} = \mathcal{S}\big(\mathcal{N}(\mathcal{I}(X), M)\big), \tag{2}$$

where:

- $\mathcal{I}(\cdot)$: initial mean imputation for missing entries;
- M: binary mask (0 = missing, 1 = observed);
- $\mathcal{N}(\cdot)$: z-score normalization per channel;

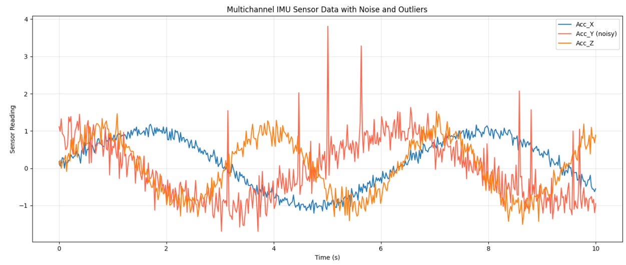

Fig. 1: Depicts the three accelerometer axes - Acc X, Acc Y, and Acc Z, highlighting that Acc Y shows increased noise and outlier spikes, indicating sensor disturbances or data corruption.

- $\mathcal{S}(\cdot)$: segmentation into overlapping 1s windows (30 frames, 50% overlap).

Before transformation, column names are standardized across sessions and absent channels are appended as zero-valued placeholders.

C. Loading and Cleaning Raw Sensor Data

Raw sensor streams are first validated for structural consistency, with rows padded or trimmed to 250 fields to ensure fixed input dimensions for deep learning frameworks. Features (columns 1–242) and activity labels (column 243) are separated, and rows with missing or invalid labels are removed. A binary mask M records missing entries, guiding advanced imputation methods such as GAN-based augmentation [9]. Summary statistics, including missing-data patterns, class balance, and sensor-specific gaps, are computed to inform preprocessing choices. Cleaned data, masks, and metadata are compiled for efficient model training. Figure 1 shows accelerometer channels (Acc_X, Acc_Y, Acc_Z) from the Opportunity dataset, where Acc_Y exhibits noise spikes, likely from sensor disturbances. Z-score normalization mitigates such anomalies by scaling extreme values, preserving motion continuity without explicit spike removal.

D. Missing Value Handling and Feature Normalization

In the Opportunity dataset, missing values—caused by sensor faults, interference, or recording errors—can distort feature learning and degrade prediction accuracy. Missingness is often sensor-specific, with wrist and ankle sensors showing higher dropout rates, sometimes in multi-second contiguous gaps. We first apply column-wise mean imputation, replacing missing entries with the mean of observed values for each channel ($x_{t,i} \leftarrow \mu_i$ if missing), ensuring no **NaN** values persist for subsequent processing. This preserves data continuity for segmentation and scaling.

To standardize heterogeneous sensor readings, we apply Z-score normalization:

$$\hat{x}_{t,i} = \frac{x_{t,i} - \mu_i}{\sigma_i}$$

where μ_i and σ_i are the mean and standard deviation of feature i. This transformation aligns all channels to zero mean and unit variance, improving convergence, stabilizing

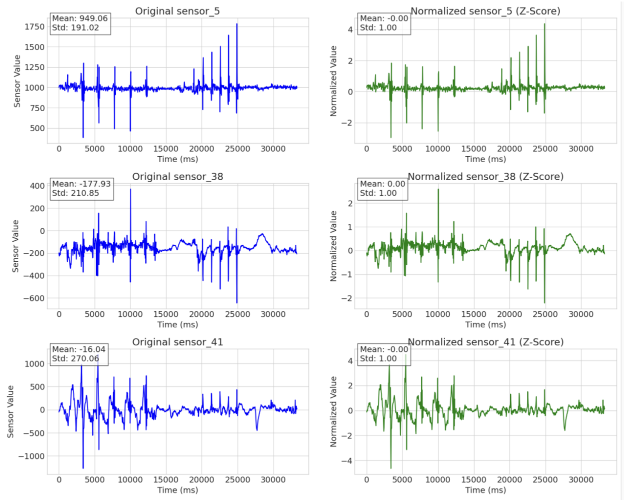

Fig. 2: Comparison of raw and Z-score normalized signals from three sensors: Accelerometer HIP accX (Sensor 5), IMU BACK accX (Sensor 38), and IMU BACK gyroX (Sensor 41) - demonstrating how normalization aligns scale and variance across heterogeneous sensor channels.

gradients, and reducing bias toward high-magnitude features. Figure 2 illustrates how normalization equalizes scales across accelerometer and gyroscope channels, enhancing cross-sensor comparability and model robustness.

E. Sliding Window Segmentation

In HAR, sensor data streams are segmented into overlapping windows to capture temporal patterns like walking or posture transitions. We use 1-second windows (30 time steps) with 50% overlap, which balances activity representation and computational cost. Empirical tests and prior work on the Opportunity dataset show this size preserves full activity cycles while avoiding label ambiguity from overly long windows. Overlaps boost training samples, maintain temporal context, and ensure smooth label transitions, providing a solid foundation for augmentation methods.

IV. Methodology

The proposed Human Activity Recognition (HAR) framework processes wearable sensor data through three stages: (i) data acquisition and preprocessing, (ii) missing data imputation with augmentation, and (iii) DeepSense-based classification. The pipeline (Figure 3) addresses challenges of class imbalance, noise, subject variability, and sensor dropouts.

A. Data Preprocessing Pipeline Specific Implementation Details

Raw multivariate time-series data $X_{\text{raw}} \in \mathbb{R}^{T\times D}$, where T is the number of timesteps and D the sensor channels, undergoes structural validation and Z-score normalization:

$$X_{\text{norm}}^{(d)} = \frac{X_{\text{clean}}^{(d)} - \mu_d}{\sigma_d}$$

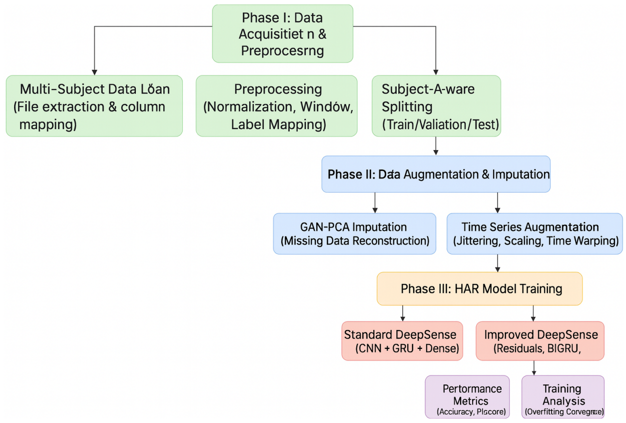

Fig. 3: Pipeline: data loading and preprocessing; PCA/GAN for completion and synthesis; DeepSense classifier; multi-protocol evaluation.

Sliding windows of size $w = 30$ and stride $s = 15$ segment the stream:

$$\{X_{\text{win}}^{(i)}\}_{i=1}^{N}, \quad X_{\text{win}}^{(i)} \in \mathbb{R}^{w\times D}$$

Each window receives a majority-vote label from its constituent samples.

B. Imputation and Augmentation (Pipeline-Specific Methods Extending Section III)

To mitigate missing data and class imbalance, a two-step imputation and augmentation strategy is used:

PCA Completion: Principal Component Analysis reduces high-dimensional sensor space while imputing low-rank structure:

$$X_{\text{PCA}} = (X - \bar{X})V_{[:,1:k]}$$

Retaining k components preserves $\geq 95\%$ variance, suppresses noise, and improves efficiency.

GAN-based Imputation and Synthesis: A conditional GAN fills missing values using a binary mask M and generates class-balanced synthetic windows $\hat{X}_{\text{aug}}$:

$$\hat{X} = G(X, M, z), \quad z \sim \mathcal{N}(0, I)$$

The generator G employs bi-directional GRUs for temporal inference, while the discriminator D uses CNN layers to distinguish real and imputed data. The objective combines adversarial and reconstruction losses:

$$\mathcal{L} = \mathcal{L}_{\text{GAN}} + \lambda\mathcal{L}_{\text{recon}}$$

with $\lambda = 10$.

Augmentation techniques include jittering, scaling, time warping, and permutation, applied only to training data.

C. DeepSense Architecture

DeepSense integrates convolutional and recurrent blocks to learn spatial-temporal dependencies:

- **Convolutional Feature Extractor:** Three 1D CNN layers (64, 128, 256 filters; kernel sizes 3, 5, 7) with batch normalization and ReLU.

- **Temporal Modeling:** Two bi-directional GRUs (128 units each, dropout 0.4) model sequential dependencies.
- **Classifier:** Dense layer (128 units) followed by softmax for C activity classes.

D. Training Protocol

Models are trained with categorical cross-entropy loss and Adam optimizer ($\eta = 10^{-3}$) with adaptive learning rate scheduling. Early stopping (patience=10) prevents overfitting. Class weights address imbalance, and L_2 regularization ($\lambda = 0.001$) is applied to dense and convolutional layers.

E. Evaluation

Performance is measured via Accuracy, Precision, Recall, and F1-score under k-fold, Leave-One-Subject-Out (LOSO), and Leave-One-Session-Out (LOSEO) validation schemes. The final training set:

$$X_{\text{train}} = \{X_{\text{win}}\} \cup \{\hat{X}_{\text{aug}}\}, \quad y_{\text{train}} = \{y\} \cup \{\hat{y}\}$$

provides a balanced, noise-tolerant dataset for robust HAR classification.

V. Results

This section presents the performance evaluation of our Human Activity Recognition (HAR) framework using three complementary validation strategies: *K-Fold Cross-Validation*, *Leave-One-Session-Out (LOSEO)*, and *Leave-One-Subject-Out (LOSO)*. Each method provides a distinct lens on model generalization under varying data partitioning schemes. Model performance is evaluated on a four-class (stand, walk, sit, lie) locomotion recognition task derived from the OPPORTUNITY dataset. These classes were chosen as they have good representation across subjects and sessions to ensure reliable training and meaningful cross-validation. More granular labels about object interactions were avoided on purpose, because such labels naturally suffer from high sparsity and extreme imbalance, which would confound the assessment of robustness to missing data rather than reflect the effective performance of the proposed DeepSense++ framework.

A. K-Fold Cross-Validation

We applied a stratified 5-fold cross-validation to ensure balanced class representation in each split. In each iteration, four folds were used for training and one for testing, cycling through all combinations.

For the baseline configuration (DeepSense + mean imputation), the model achieved an average accuracy of **88.75%** ($\pm$0.74%) and an F1-score of **88.28%** ($\pm$0.99%). Precision and recall averaged **88.90%** ($\pm$1.02%) and **87.70%** ($\pm$0.85%) respectively, as illustrated in Figure 4, indicating balanced detection across classes despite sensor noise and imbalance.

Replacing mean imputation with a GAN-based imputer followed by PCA dimensionality reduction improved performance, yielding an accuracy of **92.48%** ($\pm$0.49%) and F1-score of **90.86%** ($\pm$0.35%), , as illustrated in Figure 5. The reduced variance across folds highlights the model's consistency.

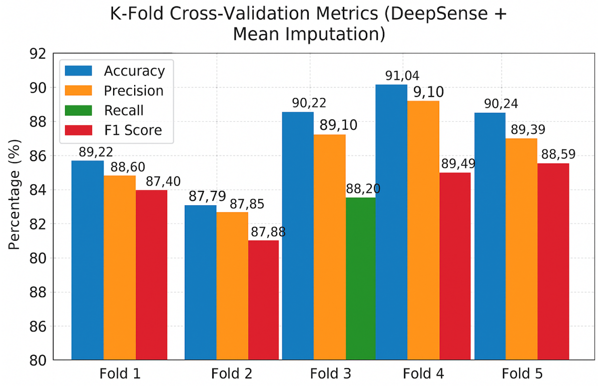

Fig. 4: DeepSense with mean imputation: 5-fold cross-validation results on the OPPORTUNITY dataset.

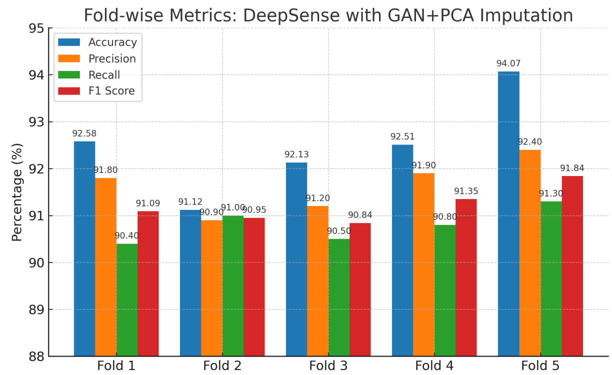

Fig. 5: DeepSense with GAN + PCA imputation: 5-fold cross-validation results.

Observations:

- Accuracy improved by **+3.71%** and F1-score by **+2.92%** over the baseline.
- Lower variability reflects enhanced stability.
- High per-fold accuracy (>92%) indicates robustness to fold-specific data shifts.

B. Leave-One-Session-Out (LOSEO)

LOSEO-CV evaluates temporal generalization by holding out one entire recording session for testing while training on the rest.

Accuracy ranged from 82.9% to 86.7%, with a low standard deviation ($\sim$1.22%), as demonstrated in Figure 6, suggesting stable temporal generalization. Sessions from Subject 2 consistently performed better, potentially due to clearer and more representative activity patterns. The mean accuracy of **84.91%** was about 3.8% lower than the K-Fold average, indicating that session-to-session variability remains a challenge.

C. Leave-One-Subject-Out (LOSO)

LOSO-CV provides the most stringent generalization test, simulating deployment on unseen users.

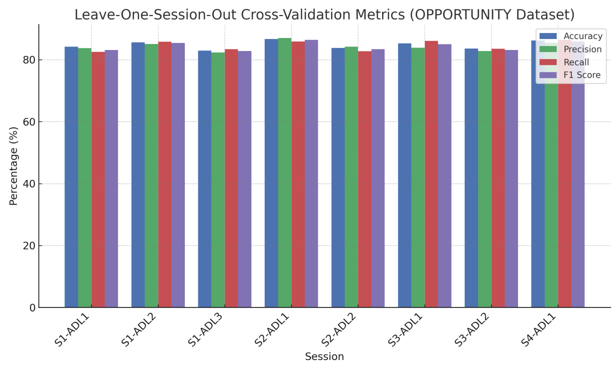

Fig. 6: LOSEO-CV results for DeepSense on the OPPORTUNITY dataset.

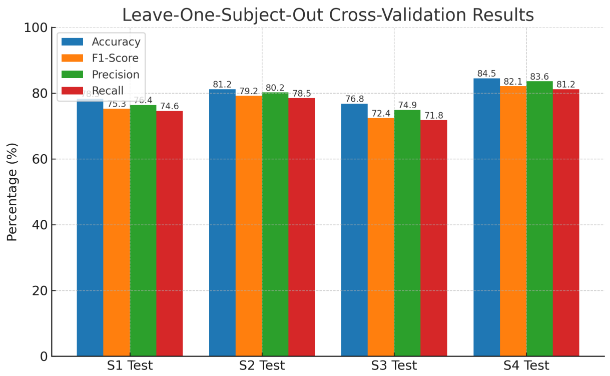

Fig. 7: LOSO-CV results: subject-wise performance variation.

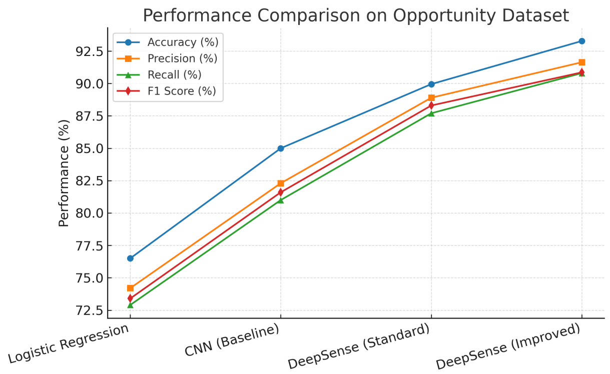

Fig. 8: Comparative performance across baselines and Improved DeepSense.

Performance varied by up to 7.7% across subjects, with S4 achieving the highest accuracy (84.5%) and S3 the lowest (76.8%), as shown in Figure 7. The average LOSO accuracy (**80.2%**) was notably lower than K-Fold and LOSEO results, highlighting subject-specific movement styles as the most difficult generalization barrier. Static activities generalized better (avg. F1: 0.842) than dynamic activities (avg. F1: 0.704).

D. Comparative Performance

The Improved DeepSense model outperformed all baselines across all metrics. It achieved **93.28%** accuracy, **91.64%** precision, **90.80%** recall, and **90.86%** F1-score (Figure 8). All baselines and the proposed model were evaluated using 5-fold stratified cross-validation for standard evaluation. The combination of GAN-based imputation and PCA effectively addressed missing and noisy sensor data, while the hybrid CNN-GRU architecture captured both spatial and temporal dependencies essential for structured HAR datasets like OPPORTUNITY.

ll baselines and the proposed model were evaluated using identical protocols:

5-fold stratified cross-validation for standard evaluation

E. Context with Recent HAR Methods

Recent general HAR methods, not limited to the OPPORTUNITY dataset published after 2022 show competitive performance. *Augmented Adversarial Learning* leads with 90.0% accuracy, followed by *Statistical Diffusion for HAR* at 87.0%, and both *Leveraging SSL* and *DNN Benchmarking* at 86.0%. The lowest, *Disentangled Behavior Patterns*, achieved 85.0%. Methods incorporating data augmentation and generative strategies tend to outperform purely self-supervised or standard deep networks.

VI. Conclusion and Future Directions

This work presented a comprehensive deep learning pipeline to address practical challenges in sensor-based Human Activity Recognition (HAR), with a focus on handling missing data, class imbalance, and generalization. Extending the DeepSense architecture, we proposed a hybrid GAN-PCA imputation technique to restore missing sensor streams and improve representation quality. The model was rigorously evaluated on the OPPORTUNITY dataset using subject-aware 5-fold cross-validation, Leave-One-Session-Out (LOSEO), and Leave-One-Subject-Out (LOSO) protocols.

a) Key findings include:: The Improved DeepSense model—combining PCA-based dimensionality reduction with GAN-based imputation—achieved the highest overall performance, with **93.28%** accuracy and **90.86%** F1 score. These results highlight the synergy of advanced imputation, dimensionality reduction, and temporal modeling for multi-sensor HAR.

Across K-Fold, LOSEO, and LOSO evaluations, performance decreased progressively, illustrating the increased difficulty of generalizing across temporal segments and unseen users. Nevertheless, the Improved DeepSense maintained robust performance, demonstrating strong generalization capability.

Logistic regression and CNN baselines, though computationally efficient, fell short in capturing complex activity patterns compared to DeepSense-based models. The integration of time-series augmentation, dimensionality reduction, and adaptive training further enhanced model resilience and learning efficiency.

Overall, the results confirm that a carefully designed hybrid imputation and augmentation pipeline, coupled with a sequence-aware deep learning architecture, can significantly enhance the practicality and dependability of HAR systems in real-world deployments. These findings provide valuable insights for designing generalizable, high-performance activity recognition models.

A. Future Work

Future research can enhance the HAR pipeline by exploring self-supervised pretraining (e.g., contrastive learning, masked reconstruction) to learn robust representations from large unlabeled datasets, improving performance with limited labeled data or unseen subjects. Adaptive imputation methods, such as temporal attention-based techniques, could better handle long-range missing data or sensor dropout.

Incorporating multi-modal fusion (e.g., ambient sensors, location tags, wearable video) may provide richer context and improve classification of ambiguous activities, requiring architectural modifications for data synchronization. Although we adopt standard z-score normalization to remain consistent with prior work on the OPPORTUNITY dataset, including the original DeepSense and similar HAR pipelines, we do note that this can be sensitive to outliers. A robust z-score based on the median and MAD represents a promising alternative for datasets with stronger sensor noise, and integrating such robust normalization strategies may further improve generalization in future extensions of DeepSense++.

Finally, expanding evaluations to diverse datasets, conducting cross-dataset generalization studies, and investigating privacy-preserving approaches like federated learning will further assess and improve the framework's robustness, portability, and practical relevance.

References

[1] S. Akther, N. Saleheen, S. A. Samiei, V. Shetty, E. Ertin, and S. Kumar, "moral: An mhealth model for inferring oral hygiene behaviors in-the-wild using wrist-worn inertial sensors," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, pp. 1–25, 2019.

[2] S. Akther, N. Saleheen, M. Saha, V. Shetty, and S. Kumar, "mteeth: Identifying brushing teeth surfaces using wrist-worn inertial sensors," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 5, no. 2, pp. 1–25, 2021.

[3] N. R. Agumamidi and S. Akther, "Gesture recognition dynamics: Unveiling video patterns with deep learning," in *2024 International Conference on Data Science and Network Security (ICDSNS)*, pp. 1–7, IEEE, 2024.

[4] N. Saleheen, A. A. Ali, S. M. Hossain, H. Sarker, S. Chatterjee, B. Marlin, E. Ertin, M. Al'Absi, and S. Kumar, "puffmarker: a multi-sensor approach for pinpointing the timing of first lapse in smoking cessation," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 999–1010, 2015.

[5] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Beyond the state of the art for human activity recognition using wearable sensors," *Computers*, vol. 7, no. 1, p. 13, 2018.

[6] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Applied Soft Computing*, vol. 62, pp. 915–922, 2018.

[7] N. Nair and et al., "Human activity recognition using temporal convolutional networks," *Proceedings of the International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence (iWOAR)*, 2018.

[8] L. Zhang, J. Yin, Y. Wang, X. Zhang, and C. Zhang, "Sensorgan: Synthesizing fine-grained sensor data for human activity recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 6928–6935, 2020.

[9] C. Ma, J. Zhao, J. Cao, P. S. Yu, and Y. Xie, "Activitygan: Generative adversarial network for human activity recognition," in *2019 IEEE International Conference on Data Mining (ICDM)*, pp. 839–848, IEEE, 2019.

[10] L. Bai, L. Yao, X. Wang, S. S. Kanhere, B. Guo, and Z. Yu, "Adversarial multi-view networks for activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1–22, 2020.

[11] X. Li *et al.*, "Xhar: Cross-domain human activity recognition via adversarial learning," in *2020 17th IEEE International Conference on Sensing, Communication and Networking (SECON)*, IEEE, 2020.

[12] S. Wang, Y. Chen, W. Ma, H. Li, and J. Yang, "Cross-dataset activity recognition via adaptive spatial-temporal transfer learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 615–622, 2020.

[13] H. Haresamudram, I. Essa, and T. Pl"otz, "Contrastive predictive coding for human activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–26, 2021.

[14] X. Song, L. Cui, W. Li, W. Jiang, and X. Wang, "Leveraging self-supervised learning for human activity recognition with ambient sensors," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8311–8322, 2022.

[15] J. Su, Z. Wen, T. Lin, and Y. Guan, "Learning disentangled behaviour patterns for wearable-based human activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 1, 2022.

[16] T. Li, W. Wu, J. Wang, and Y. Zhu, "Unsupervised statistical feature-guided diffusion model for sensor-based human activity recognition," in *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 5273–5282, ACM, 2023.

[17] T. Um, J. Kim, S. Lee, S. Moon, and D. Park, "Autoaughar: Human activity recognition using sensor data with automatic augmentation policy search," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, ACM, 2023.

[18] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "Deepsense: A unified deep learning framework for time-series mobile sensing data processing," in *Proceedings of the 26th international conference on world wide web*, pp. 351–360, 2017.

[19] C. A. N.-D. L.-V. C. R. Roggen, Daniel and H. Sagha, "OPPORTUNITY Activity Recognition." UCI Machine Learning Repository, 2010. DOI: https://doi.org/10.24432/C5M027.