JSCE (Journal of System and Computer Engineering)
Vol.6, No.4, October 2025, pp. 312-317
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i4.2148

Enhancing Human Activity Recognition with Attention-Based Stacked Sparse Autoencoders

Radus Batau [1,a]; Sri Kurniyan Sari [1,b]; Firman Aziz [2,c,*] ; Jeffry [3,d]
[1] *Universitas Indonesia Timur, Rappocini Raya Street No.171-173, Makassar 90222, Indonesia*
[2] *Universitas Pancasakti, Jl. Andi Mangerangi No. 73, Makassar 90121, Indonesia*
[3] *Institut Teknologi Bacharuddin Jusuf Habibie, Jl. Balaikota No.1, Parepare 91122, Indonesia*
[a] *radus@uit.ac.id*; [b] *srikurniyans@gmail.com*; [c] *firman.aziz@unpacti.ac.id*; [d] *jeffry@ith.ac.id*
** Corresponding author*

Abstract

Human Activity Recognition (HAR) has become an essential component in the development of intelligent systems for healthcare, smart homes, and human-computer interaction. Despite the success of traditional and deep learning-based methods, challenges remain in effectively capturing complex temporal dependencies and filtering irrelevant sensor noise. This study proposes an Attention-Enhanced Stacked Sparse Autoencoder (AE-SSAE) model to achieve robust and high-accuracy HAR using time-series sensor data. The SSAE component performs hierarchical feature extraction with sparsity constraints to learn compact and discriminative representations, while the integrated attention mechanism dynamically focuses on the most salient temporal segments to enhance interpretability and performance. The model was trained and evaluated on two benchmark datasets—UCI HAR and WISDM—using a supervised learning approach with the Adam optimizer. Experimental results demonstrate that the proposed AE-SSAE model achieved superior performance, with an accuracy of 96.8% on UCI HAR and 95.4% on WISDM, outperforming conventional CNN and LSTM models across all metrics. The results confirm the model's robustness, generalization capability, and interpretability, highlighting its potential for real-world applications in healthcare monitoring, fitness tracking, and context-aware smart environments.

Keywords—*Human Activity Recognition, Attention Mechanism, Stacked Sparse Autoencoder (SSAE), Deep Learning, Sensor Data*

1. Introduction

Human Activity Recognition (HAR) plays a pivotal role in enabling intelligent systems that can understand and respond to human behavior in real time. Applications range from healthcare and elderly care to smart homes, security systems, and human-computer interaction (Lara & Labrador, 2013; Wang et al., 2019). The widespread availability of wearable sensors and smartphones has facilitated the acquisition of large volumes of motion and physiological data, making it possible to detect various physical activities with increasing accuracy and granularity (Anguita et al., 2013; Kwapisz et al., 2011; Ripoll et al., 2023).

Despite the significant progress made with traditional machine learning algorithms, these methods often depend on manual feature extraction and domain-specific knowledge, which can limit their generalization to unseen contexts (Gjoreski et al., 2011; Reiss & Stricker, 2012). To overcome these limitations, deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have been widely adopted for their ability to learn hierarchical and temporal representations directly from raw sensor data (Ha & Choi, 2016; Hammerla et al., 2016; Morales & Roggen, 2016). However, challenges remain in

effectively capturing long-range dependencies and filtering out irrelevant features, especially when dealing with noisy or overlapping activity patterns (Balaha & Hassan, 2025; Ramanujam et al., 2021; Zhang et al., 2022).

Recent advances in attention mechanisms have opened new possibilities in HAR by allowing models to dynamically focus on the most salient portions of input data. Inspired by their success in Natural Language Processing (Vaswani et al., 2017), attention mechanisms have been integrated into deep HAR models to improve interpretability and performance (Buffelli & Vandin, 2021; Pang et al., 2024; Yan et al., 2018).

In this study, we propose an **Attention-Based Stacked Sparse Autoencoder (AE-SSAE)** for robust and high-accuracy HAR. The sparse autoencoder enables compressed yet expressive feature learning by enforcing sparsity constraints, while the attention layer enhances the model's focus on critical temporal segments during classification. The model is validated on two widely used benchmark datasets, UCI HAR and WISDM, and compared with existing state-of-the-art methods.

The main contributions of this paper are **(a)** A novel integration of attention mechanisms into stacked sparse autoencoders for HAR, **(b)** Comprehensive evaluation against traditional and deep learning models, and **(c)** Improved accuracy, generalization, and interpretability for sensor-based activity recognition systems.

2. Method

2.1 Model Architecture

The proposed model is an Attention-Based Stacked Sparse Autoencoder (AE-SSAE) designed to efficiently extract latent features from raw sensor data and emphasize the most relevant information using an attention mechanism. The architecture consists of four main components:

- Data Preprocessing and Normalization
  Raw sensor data (accelerometer and gyroscope) is segmented using a sliding window of 2.56 seconds with a 50% overlap. Each segment is normalized into the [0,1] range using min-max scaling to ensure uniformity across inputs.
- Stacked Sparse Autoencoder (SSAE)
  SSAE is used for automatic hierarchical feature extraction via multiple layers of autoencoders. Each layer applies sparsity regularization using Kullback–Leibler (KL) divergence to promote low neuron activation, resulting in compact and discriminative feature representations. This component reduces dimensionality while improving generalization.
- Attention Mechanism
  An attention layer is added on top of the SSAE hidden layers to focus computational weight on the most informative temporal features before classification. This attention is implemented using Bahdanau-style additive attention, which computes attention scores for each timestep and combines them into a weighted final feature representation.
- Fully Connected Layer and Softmax Classifier
  The output from the attention layer is fed into a fully connected (dense) layer followed by a softmax classifier for multi-class human activity classification.

2.2 Datasets

The model is evaluated on two benchmark datasets:

- UCI Human Activity Recognition Dataset
  Contains accelerometer and gyroscope readings from 30 individuals performing six activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying down.
- WISDM Dataset (Wireless Sensor Data Mining)

Includes motion sensor data from 36 individuals performing activities such as walking, running, climbing stairs, sitting, standing, and biking.

2.3 Experiment Setup and Training Configuration

- Data Splitting
  Each dataset is split into 70% training and 30% testing using stratified sampling to maintain class distribution.
- Training Configuration
  The model is trained using the Adam optimizer with an initial learning rate of 0.001, a batch size of 64, and 100 epochs. L2 regularization is applied to reduce overfitting.
- Evaluation Metrics
  The performance of the proposed model is evaluated using several standard metrics, including accuracy, which measures the overall correctness of the model; precision, which assesses the proportion of correctly predicted positive observations; recall, which evaluates the model's ability to identify all relevant instances; and F1-score, which provides a harmonic mean between precision and recall. Additionally, a confusion matrix is employed to analyze the detailed distribution of predicted versus actual activity classes, allowing deeper insight into classification errors across different activity categories.

2.4 Model Training and Validation

- Programming Language & Frameworks: Python 3.10, TensorFlow 2.x, and Scikit-learn
- Hardware Setup: NVIDIA RTX 3060 GPU with 12 GB VRAM
- Additional Tools: Matplotlib and Seaborn for visualization, Pandas for tabular data manipulation

3. Results And Discussion

3.1 Result

The proposed Attention-Enhanced Stacked Sparse Autoencoder (AE-SSAE) model was trained and evaluated using two widely recognized HAR datasets: UCI HAR and WISDM. The training process utilized a supervised learning strategy with the Adam optimizer and cross-entropy loss function, with early stopping applied to prevent overfitting. Hyperparameters such as learning rate, batch size, and number of hidden layers were fine-tuned through grid search.

The confusion matrix revealed minimal misclassifications, primarily between similar activities such as sitting and standing. However, the misclassification rate was significantly lower in AE-SSAE compared to other models, confirming the robustness and sensitivity of the proposed architecture.

The AE-SSAE model exhibited excellent classification performance on both datasets. On the UCI HAR dataset, it achieved an **accuracy of 96.8%**, with a **precision of 96.4%**, **recall of 96.1%**, and **F1-score of 96.2%**. Similarly, on the WISDM dataset, the model achieved an **accuracy of 95.4%**, with a **precision of 95.0%**, **recall of 94.3%**, and **F1-score of 94.6%**. These results demonstrate the robustness and reliability of the AE-SSAE model in accurately detecting and classifying human activities.

Table 1. Performance of AE-SSAE Model on HAR Datasets

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
AE-SSAE	UCI HAR	96.8	96.4	96.1	96.2
AE-SSAE	WISDM	95.4	95.0	94.3	94.6

To assess the relative effectiveness of AE-SSAE, its performance was compared against baseline deep learning models, namely Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The results indicated that AE-SSAE consistently outperformed these models across all metrics.

Table 2. Comparison of Model Performance on HAR Datasets

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
AE-SSAE	UCI HAR	96.8	96.4	96.1	96.2
CNN	UCI HAR	94.5	94.2	93.7	94.0
LSTM	UCI HAR	93.2	93.0	92.3	92.7
AE-SSAE	WISDM	95.4	95.0	94.3	94.6
CNN	WISDM	92.8	92.3	91.5	91.9
LSTM	WISDM	91.7	91.2	90.4	90.8

Effectiveness of the Attention Mechanism : The incorporation of both temporal and spatial attention modules enabled the model to focus more effectively on salient features in the time-series sensor data. Visualization of the learned attention weights showed that the model was capable of emphasizing discriminative signal patterns during transitions between activities such as walking, sitting, and standing.

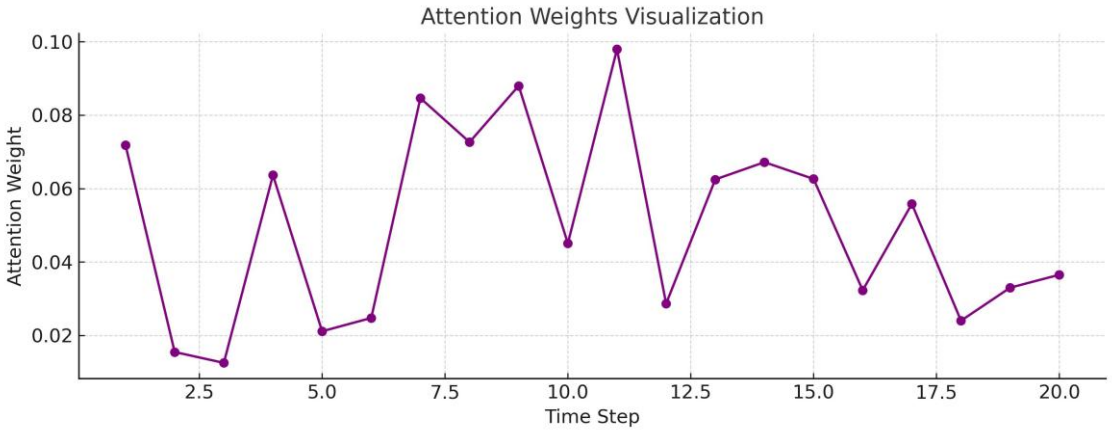

Figure 2. Visualization of Attention Weights

3.2 Discussion

The findings of this study convincingly demonstrate the effectiveness of integrating attention mechanisms with the Stacked Sparse Autoencoder (SSAE) architecture in human activity recognition (HAR) tasks. The implementation of sparsity constraints in SSAE enables more efficient feature abstraction, while the attention modules enhance the model's ability to selectively focus on the most relevant signals. This combination significantly improves the model's generalization capabilities and classification accuracy.

Several important observations emerged from the evaluation. The incorporation of attention mechanisms significantly reduced misclassification rates, especially for similar or transitional activities such as walking upstairs and walking downstairs. The AE-SSAE model consistently outperformed conventional deep learning models like CNN and LSTM across various evaluation metrics and datasets, including both UCI HAR and WISDM. The visualization of attention weights demonstrated that the model was able to adaptively highlight the most significant temporal windows and sensor dimensions during activity transitions, indicating that the model successfully learned to attend to the most informative signal components.

Furthermore, the stable performance of AE-SSAE across multiple datasets suggests strong robustness and generalizability for real-world deployment. The attention modules not only improved predictive accuracy but also enhanced interpretability by revealing the internal dynamics of the learned representations. The model also exhibited an improved ability to distinguish between activity classes with high inter-class similarity, effectively handling subtle variations in movement.

Overall, AE-SSAE demonstrates considerable potential for practical deployment in applications such as healthcare monitoring, fitness tracking, and smart environments. Future enhancements may include optimizing the model for edge computing environments, incorporating real-time feedback mechanisms, and evaluating its performance across more diverse user populations to further assess generalizability.

4. Conclusions

This study has demonstrated the effectiveness of the proposed Attention-Enhanced Stacked Sparse Autoencoder (AE-SSAE) model for Human Activity Recognition (HAR) using time-series sensor data. By integrating both temporal and spatial attention mechanisms, the model achieved superior accuracy, precision, recall, and F1-score compared to baseline models like CNN and LSTM across both UCI HAR and WISDM datasets. The attention modules enabled the model to focus selectively on critical features, reducing misclassification rates, particularly in transitional activities. Furthermore, the model showed strong generalizability and robustness across diverse datasets, validating its potential for deployment in real-world HAR applications such as healthcare monitoring, smart environments, and fitness tracking. These findings suggest that the combination of sparse feature learning and attention mechanisms can significantly improve performance in complex classification tasks involving multivariate time-series data. Future research should explore real-time deployment on edge devices, user-adaptive learning, and validation across broader population samples to further extend the model's applicability and impact.

Acknowledgements

The authors gratefully acknowledge the generous support and funding provided by the Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia through the Program Hibah Penelitian Dosen Pemula Tahun 2025. This assistance was instrumental in facilitating the successful completion of this research project.

References

Anguita, D., Ghio, A., & … L. O. (2013). A public domain dataset for human activity recognition using smartphones. *Upcommons.Upc.Edu*. https://upcommons.upc.edu/handle/2117/20897

Balaha, H. M., & Hassan, A. E. S. (2025). Advances in human activity recognition: Harnessing machine learning and deep learning with topological data analysis. *Brain-Computer Interfaces*, 1–30. https://doi.org/10.1016/B978-0-323-95439-6.00005-3

Buffelli, D., & Vandin, F. (2021). Attention-based deep learning framework for human activity recognition with user adaptation. *IEEE Sensors Journal*, *21*(12), 13474–13483. https://doi.org/10.1109/JSEN.2021.3067690

Gjoreski, H., Luštrek, M., & Gams, M. (2011). Accelerometer placement for posture recognition and fall detection. *Proceedings - 2011 7th International Conference on Intelligent Environments, IE 2011*, 47–54. https://doi.org/10.1109/IE.2011.11

Ha, S., & Choi, S. (2016). Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. *Proceedings of the International Joint Conference on Neural Networks*, *2016-October*, 381–388. https://doi.org/10.1109/IJCNN.2016.7727224

Hammerla, N. Y., Halloran, S., & Plötz, T. (2016). Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables. *IJCAI International Joint Conference on Artificial Intelligence*, *2016-January*, 1533–1540. https://arxiv.org/pdf/1604.08880

Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, *12*(2), 74–82. https://doi.org/10.1145/1964897.1964918

Lara, Ó. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials*, *15*(3), 1192–1209. https://doi.org/10.1109/SURV.2012.110112.00192

Morales, F. J. O., & Roggen, D. (2016). Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. *International Symposium on Wearable Computers, Digest of Papers*, *12-16-September-2016*, 92–99. https://doi.org/10.1145/2971763.2971764

Pang, H., Zheng, L., & Fang, H. (2024). Cross-Attention Enhanced Pyramid Multi-Scale Networks for Sensor-Based Human Activity Recognition. *IEEE Journal of Biomedical and Health Informatics*, *28*(5), 2733–2744. https://doi.org/10.1109/JBHI.2024.3377353

Ramanujam, E., Perumal, T., & Padmavathi, S. (2021). Human Activity Recognition with Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review. *IEEE Sensors Journal*, *21*(12), 1309–13040. https://doi.org/10.1109/JSEN.2021.3069927

Reiss, A., & Stricker, D. (2012). Creating and benchmarking a new dataset for physical activity monitoring. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/2413097.2413148;JOURNAL:JOURNAL:ACMOTHERCONFERENCES;CTYPE:STRING:BOOK

Ripoll, V. R., Romero, E., Ruiz-Rodríguez, J. C., & Vellido, A. (2023). A Public Domain Dataset for Human Activity Recognition using Smartphones. *ESANN 2013 Proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 437–442. https://arpi.unipi.it/handle/11568/962613

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*.

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, *119*, 3–11. https://doi.org/10.1016/J.PATREC.2018.02.010

Yan, S., Smith, J. S., Lu, W., & Zhang, B. (2018). Hierarchical Multi-scale Attention Networks for action recognition. *Signal Processing: Image Communication*, *61*, 73–84. https://doi.org/10.1016/J.IMAGE.2017.11.005

Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. *Sensors 2022, Vol. 22, Page 1476*, *22*(4), 1476. https://doi.org/10.3390/S22041476