

Evaluating BiLSTM and CNN+GRU Approaches for Human Activity Recognition Using WiFi CSI Data

Almustapha A. Wakili, Babajide J. Asaju, and Woosub Jung
 CIS Department, Towson University
 Towson, Maryland, USA
 Email: {awakili, basaju, woosubjung}@towson.edu

Abstract—This paper compares the performance of BiLSTM and CNN+GRU deep learning models for Human Activity Recognition (HAR) on two WiFi-based Channel State Information (CSI) datasets: UT-HAR and NTU-Fi HAR. The findings indicate that the CNN+GRU model has a higher accuracy on the UT-HAR dataset (95.20%) thanks to its ability to extract spatial features. In contrast, the BiLSTM model performs better on the high-resolution NTU-Fi HAR dataset (92.05%) by extracting long-term temporal dependencies more effectively. The findings strongly emphasize the critical role of dataset characteristics and preprocessing techniques in model performance improvement. We also show the real-world applicability of such models in applications like healthcare and intelligent home systems, highlighting their potential for unobtrusive activity recognition.

Index Terms—Human Activity Recognition; Deep Learning; WiFi Channel State Information (CSI); Smart Cities.

I. INTRODUCTION

Human Activity Recognition (HAR) has become a critical area of research due to its vast applications in all areas of smart cities and healthcare, including security surveillance, smart home monitoring, and lifestyle management. For instance, HAR systems are instrumental in elderly care, enabling early detection of falls or health anomalies, and in smart homes, they enhance automation by understanding user behavior [1].

Traditionally, HAR relied on wearable sensors, such as accelerometers and gyroscopes, which, although effective, introduce limitations such as discomfort during long-term use and potential user resistance due to their intrusive nature [2]. WiFi-based systems, leveraging existing wireless infrastructure, have emerged as a promising alternative, enabling seamless, non-intrusive, and real-time activity recognition [3].

Deep learning has revolutionized HAR in recent years by providing models capable of processing complex, high-dimensional data streams from environmental sensors. Deep learning has been pivotal in advancing HAR, offering robust solutions capable of processing complex and high-dimensional data streams from environmental sensors [3], [4]. Bidirectional Long Short-Term Memory (BiLSTM) networks excel at capturing temporal dependencies in activity sequences, while Convolutional Neural Networks combined with Gated Recurrent Units (CNN+GRU) leverage the strengths of spatial and temporal feature extraction [5]. These models have shown promise in recognizing and classifying human activities, espe-

cially when combined with advanced preprocessing techniques that enhance WiFi Channel State Information (CSI) signals.

This study aims to comprehensively evaluate the BiLSTM and CNN+GRU models in HAR by utilizing two distinct WiFi-based CSI datasets: UT-HAR [6] and NTU-Fi HAR [7]. By focusing on the effect of data pre-processing techniques, model architectures, and dataset characteristics, this study investigates how to improve HAR model performance regarding the accuracy and computational efficiency to make this model applicable for real-time in practical applications such as smart homes and healthcare environments.

The rest of the paper is structured as follows: Section II discusses related work, Section III elaborates on the methodologies, Section IV presents experimental results, Section V discusses findings, and Section VI concludes with future directions.

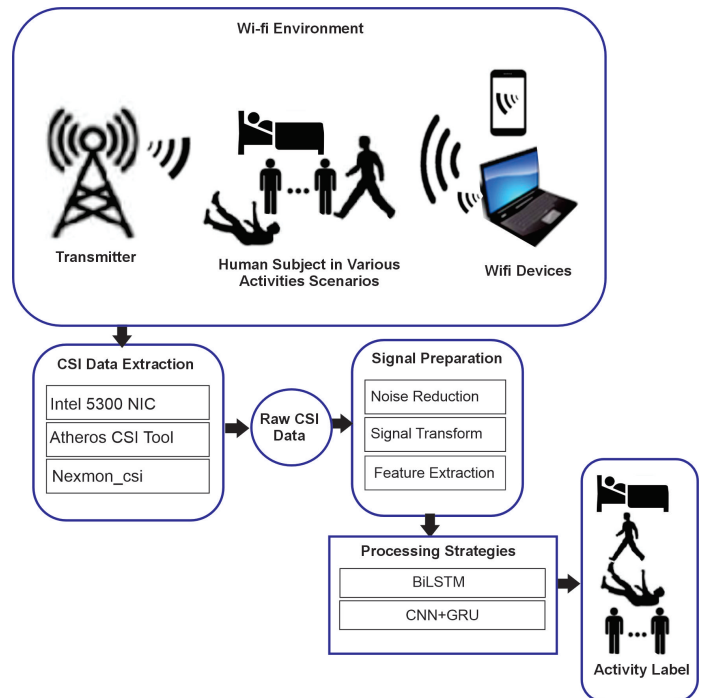


Fig. 1: WiFi-based Human Activity Recognition Pipeline.

The presented framework (Figure 1) outlines the complete

process of WiFi-based Human Activity Recognition (HAR). It begins with extracting Channel State Information (CSI) data using tools like Intel 5300 NIC, Atheros CSI Tool, and Nexmon CSI in a WiFi-enabled environment. The raw CSI data undergoes signal preparation, which includes noise reduction, signal transformation, and feature extraction, to improve its quality and usability. Finally, the processed signals are fed into BiLSTM and CNN+GRU models, which classify human activities into specific categories. This modular pipeline highlights the system's adaptability to different scenarios and potential applications in healthcare, smart homes, and other real-world settings.

II. RELATED WORKS

Human Activity Recognition (HAR) has gained widespread attention in recent years, driven by its transformative applications in healthcare, smart homes, and security systems. Traditional approaches primarily relied on wearable devices to monitor activities. While effective, these devices often raise concerns about user discomfort and practicality in daily life [8]. This limitation has motivated researchers to explore non-invasive solutions, particularly those leveraging WiFi signals for activity recognition.

A. WiFi-Based Human Activity Recognition

WiFi-based HAR offers a compelling alternative to sensor-dependent methods by utilizing Channel State Information (CSI) to capture the impact of human motion on wireless signals. Unlike wearable sensors, WiFi-based systems allow for seamless, device-free monitoring, which is especially appealing for applications in smart environments [3]. Device-free sensing approaches, which rely on ambient signals like WiFi, have proven particularly effective in scenarios where privacy is a concern. Compared to device-based sensing using wearable trackers, these methods eliminate the need for physical sensors on users, making them more suitable for healthcare and smart environments [8].

Studies like Yang et al. [7] have demonstrated the potential of WiFi-based HAR to detect activities with high precision, thanks to the fine-grained subcarrier-level information provided by CSI. Similarly, Ahmad et al. [3] explored the integration of WiFi signals with deep learning techniques to achieve precise activity classification, highlighting the non-invasive nature of RF-based sensing. However, the reliability of WiFi-based systems is often challenged by environmental noise, signal interference, and the complexity of real-world settings, such as non-line-of-sight (NLoS) scenarios.

Lightweight CNN architectures, as proposed by Zhang et al. [9], have shown promise in addressing computational constraints while maintaining accuracy. Data augmentation techniques have also been employed to improve generalization across diverse environments. Advanced preprocessing pipelines incorporating denoising, phase correction, and data segmentation are fundamental to ensuring reliable model performance [10]. These preprocessing techniques enhance the

clarity of CSI data and mitigate the impact of environmental noise, a critical step in developing robust HAR systems.

B. Deep Learning Approaches in HAR

The rise of deep learning has significantly advanced the capabilities of HAR systems, providing tools to analyze complex data patterns and uncover subtle features within activity datasets. Among the most prominent models are Bidirectional Long Short-Term Memory (BiLSTM) networks, which excel at capturing temporal dependencies. For example, Li et al. [11] demonstrated the effectiveness of BiLSTM in recognizing sequential activity patterns by leveraging its ability to process data in both forward and backward temporal directions. This bidirectional nature allows the model to capture more nuanced transitions between activities.

Hybrid architectures, such as those combining Convolutional Neural Networks (CNNs) with Gated Recurrent Units (GRUs), have also emerged as powerful alternatives. By integrating spatial feature extraction through CNNs and temporal sequence modeling through GRUs, these models offer a balanced approach to processing activity data. Studies like Kim et al. [12] and Shahverdi et al. [8] highlight the efficiency of CNN+GRU architectures, particularly in datasets with lower temporal resolution, where spatial patterns play a more critical role. However, the trade-offs between computational efficiency and temporal modeling remain an ongoing challenge [13].

Emerging technologies, including transformer-based models, offer promising solutions to these challenges. These architectures, known for their ability to model long-range dependencies, are beginning to be adapted for HAR tasks [2]. Additionally, multimodal approaches integrating visual and non-visual modalities have been explored to enhance activity recognition in complex settings, as demonstrated by Bouchabou et al. [14].

C. Preprocessing for CSI Data

The importance of preprocessing cannot be overstated in WiFi-based HAR, as raw CSI data is often noisy and susceptible to interference. Preprocessing techniques such as denoising, normalization, and sliding window segmentation significantly influence model accuracy [15]. Fourier and wavelet transforms are widely used to extract frequency-domain features that reveal periodic patterns associated with human activities [7]. High-pass filtering is another common technique employed to eliminate low-frequency noise, enhancing signal quality.

Recent work by Yousefi et al. [15] introduced Body-coordinate Velocity Profiles (BVP) to reduce environmental dependencies by focusing on motion-specific characteristics of CSI. However, challenges persist even with these advanced methods. While effective, techniques like sliding window segmentation often overlook fine-grained temporal variations, which can be critical for distinguishing between similar activities. Advanced architectures, such as attention mechanisms and transformers, may provide a pathway to overcoming these limitations by improving temporal feature extraction [2].

D. Current Challenges and Emerging Trends

Despite significant advancements, HAR research faces limited dataset diversity, scalability, and real-time performance. Current datasets often lack the variability needed to generalize models across different environments, including diverse layouts, overlapping activities, and dynamic noise sources [3]. Lightweight models, such as those explored by Zhang et al. [9], and large-scale annotated datasets have been identified as critical for advancing the field.

As highlighted by Kumar et al. [2], transformer architectures and attention mechanisms offer promising solutions by prioritizing relevant features and discarding irrelevant noise. These innovations could significantly enhance the interpretability and robustness of HAR systems. Future research should also address trade-offs between accuracy and computational cost, particularly for resource-constrained devices [13].

Interdisciplinary collaborations are increasingly considered vital to advancing HAR research. Integrating domain knowledge from healthcare, ergonomics, and urban design could lead to more adaptable and user-centric solutions, ensuring that HAR technologies meet the demands of practical, real-world applications. This review establishes the foundation for our study, focusing on leveraging deep learning techniques, particularly BiLSTM and CNN+GRU models, to enhance WiFi-based HAR systems for diverse applications.

III. METHODOLOGY

This study evaluates the performance of BiLSTM and CNN+GRU models on two WiFi-based CSI datasets: NTU_Fi_HAR and UT_HAR. The methodology includes dataset preprocessing, model implementation, training strategies, and computational efficiency analysis.

A. Dataset Overview

1) *NTU_Fi_HAR Dataset*: The NTU_Fi_HAR dataset, collected using the Atheros CSI Tool, features high-resolution CSI data with 114 subcarriers per antenna pair and captures a variety of human activities in different scenarios [7]. This dataset includes amplitude and phase information, enabling a comprehensive analysis of the CSI data's spatial and temporal features.

Table I shows the distribution of training and validation instances across the six activity classes included in the dataset.

Activity	Training Instances	Validation Instances
Clean	156	44
Fall	156	44
Run	156	44
Walk	156	44
Jump	156	44
Circle	156	44

TABLE I: Distribution of activities and the number of instances in the NTU_Fi_HAR dataset.

2) *UT_HAR Dataset*: The UT_HAR dataset, collected using the Intel 5300 NIC, is designed for HAR tasks and features continuously recorded CSI data. Unlike NTU_Fi_HAR, this dataset lacks explicit activity segmentation. Therefore, preprocessing steps, such as applying a sliding window technique, are required to segment the data into meaningful samples [6].

Table II presents the distribution of training, validation, and testing instances for the six activity classes included in the dataset.

Class	Training Instances	Validation Instances	Testing Instances
Class 0	3977	496	500
Class 1	3977	496	500
Class 2	3977	496	500
Class 3	3977	496	500
Class 4	3977	496	500
Class 5	3977	496	500

TABLE II: Distribution of classes and the number of instances in the UT_HAR dataset.

B. Preprocessing

Effective preprocessing of raw CSI data is crucial for achieving high model performance. This study employs the following preprocessing steps tailored to the characteristics of the NTU_Fi_HAR and UT_HAR datasets:

- **Noise Reduction**: High-pass filters are applied to eliminate low-frequency noise, enhancing signal clarity for activity recognition.
- **Normalization**: CSI amplitude and phase data are scaled uniformly, ensuring sample consistency and reducing data biases.
- **Feature Extraction**: Fourier and wavelet transforms capture essential frequency and time-domain features, critical for spatial and temporal analysis.

Table III summarizes the preprocessing pipeline, highlighting the unique adaptations for each dataset.

Preprocessing Step	NTU_Fi_HAR	UT_HAR
Noise Reduction	High-pass filtering	High-pass filtering
Normalization	Amplitude and phase scaling	Amplitude scaling
Feature Extraction	Doppler profiles, Fourier transforms	Sliding window segmentation

TABLE III: Comparison of preprocessing steps for NTU_Fi_HAR and UT_HAR datasets.

The workflow illustrated in Figure 2 provides a detailed overview of the experimental methodology used in this study. The process starts with the NTU-Fi HAR and UT-HAR datasets, each undergoing specific preprocessing steps such as noise reduction, feature extraction, and segmentation. Preprocessed data is loaded into separate pipelines for training and validation using BiLSTM and CNN+GRU models. These models predict human activities based on the processed data, and their performance metrics are computed for evaluation. Finally, the performances of both models are compared to understand their relative strengths and suitability for different datasets and scenarios.

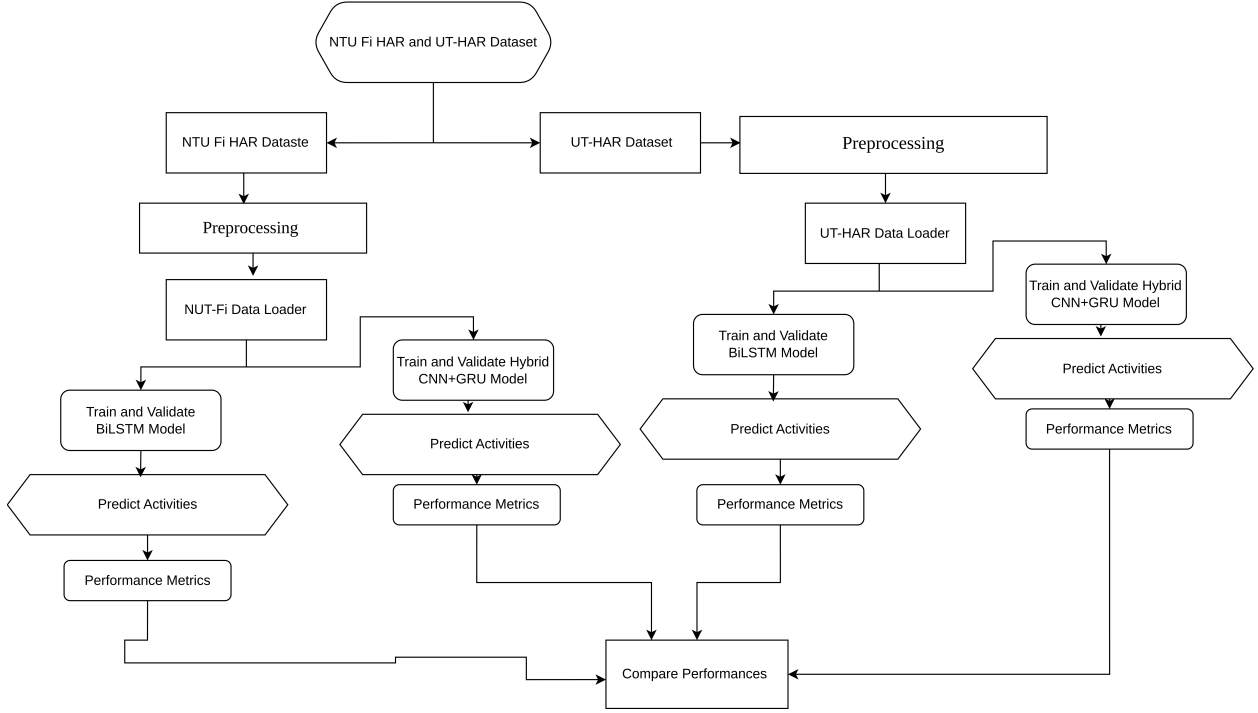


Fig. 2: Workflow for Training and Evaluating BiLSTM and CNN+GRU Models on NTU-Fi HAR and UT-HAR Datasets.

C. Model Implementation

1) *BiLSTM Model*: BiLSTM is designed to capture long-term dependencies. The model includes three bidirectional LSTM layers, each with 128 units, followed by a dense layer with softmax activation.

2) *CNN+GRU Model*: The CNN+GRU model combines convolutional layers for spatial feature extraction with GRU layers for temporal sequence analysis. This hybrid architecture enhances the model's ability to handle the spatiotemporal characteristics of CSI data. Figure 3 shows the CNN+GRU architecture.

D. Training and Computational Efficiency

Both models are trained using the Adam optimizer with a learning rate 0.001 and early stopping to prevent overfitting. Table IV compares the inference time and memory usage, highlighting the computational advantages of CNN+GRU.

Model	Inference Time (milliseconds)	Memory Usage (MB)
BiLSTM	120	250
CNN+GRU	90	150

TABLE IV: Comparison of Computational Efficiency. This table shows the inference time and memory usage for the BiLSTM and CNN+GRU models.

IV. RESULTS AND EVALUATION

This section presents the experimental results and provides a comparative analysis of the performance of the CNN+GRU

and BiLSTM models on the NTU-Fi HAR and UT-HAR datasets. The results highlight distinct trends in model performance, reflecting the influence of dataset characteristics and architectural strengths. The CNN+GRU model consistently demonstrated superior performance on the UT-HAR dataset, primarily due to its advanced spatial feature extraction capabilities. Conversely, the BiLSTM model excelled on the NTU-Fi HAR dataset, effectively leveraging its ability to capture long-term temporal dependencies from high-resolution CSI data.

A. Performance on NTU_Fi_HAR Dataset

The NTU_Fi_HAR dataset, characterized by its high-resolution CSI data, provided a robust platform for evaluating both temporal and spatial feature extraction capabilities. Table V summarizes the overall performance metrics for CNN+GRU and BiLSTM models. The CNN+GRU model achieved the highest overall accuracy, precision, recall, and F1-score, highlighting its versatility. However, the BiLSTM model also performed strongly, particularly in activities that require intricate temporal analysis.

Model	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
CNN+GRU	99.24	99.24	99.24	99.24
BiLSTM	92.05	92.83	92.05	92.03

TABLE V: Overall performance metrics for CNN+GRU and BiLSTM models on the NTU_Fi_HAR dataset.

CNN+GRU MODEL

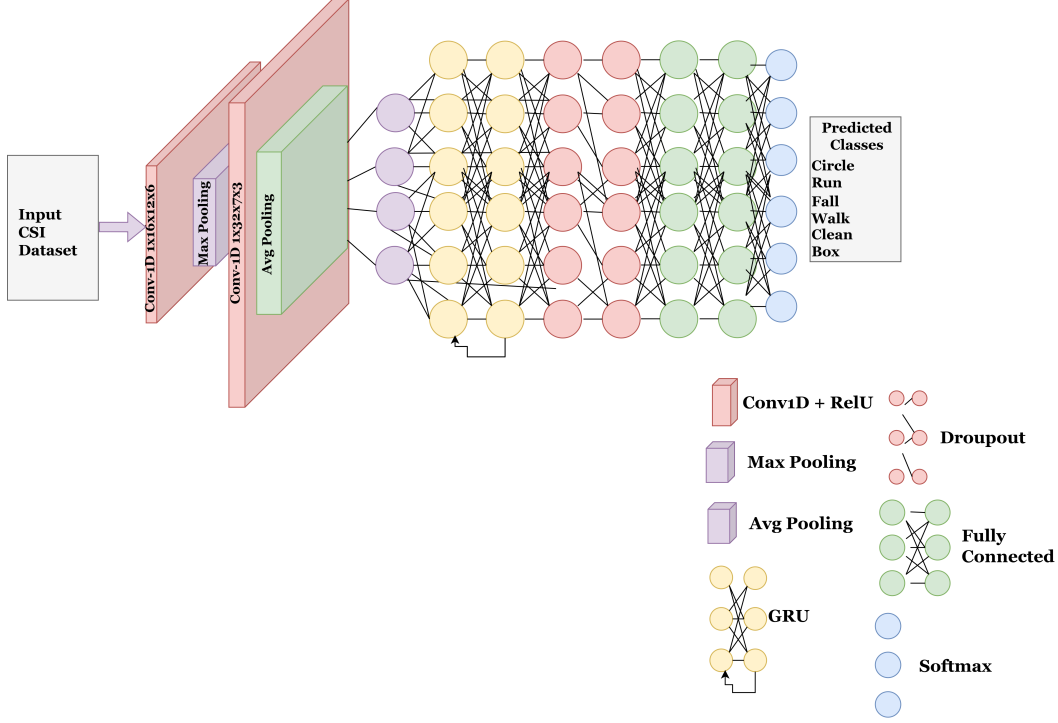


Fig. 3: CNN+GRU Model Architecture. This hybrid architecture integrates convolutional layers with GRU layers to enhance spatiotemporal analysis.

Table VI provides a detailed class-wise performance comparison. The CNN+GRU model consistently outperformed BiLSTM across all classes, with notable precision and recall for activities such as "Run" and "Walk." The BiLSTM model, while performing adequately overall, exhibited minor misclassifications in more complex activities like "Run" and "Box," possibly due to the intricate temporal patterns involved.

Class	CNN+GRU			BiLSTM		
	Precision	Recall	F1	Precision	Recall	F1
Clean	98	100	99	100	93	96
Fall	100	100	100	100	100	100
Run	100	98	99	82	91	86
Walk	100	98	99	86	100	93
Box	98	100	99	89	77	83
Circle	100	100	100	98	91	94

TABLE VI: Class-wise performance metrics (%) for CNN+GRU and BiLSTM models on the NTU-Fi_HAR dataset.

Figure 4 illustrates the confusion matrix for the CNN+GRU model on the NTU-Fi HAR dataset. The matrix reflects near-perfect classification accuracy across all activity classes, with only minor misclassifications observed in activities such as "Box." Similarly, Figure 5 shows the confusion matrix for the BiLSTM model, which, while performing well overall, struggled slightly in distinguishing between similar activities like "Run" and "Box," likely due to overlapping temporal

patterns.

B. Performance on UT_HAR Dataset

The UT-HAR dataset, characterized by lower resolution and limited temporal details, presented unique challenges for both models. Table VII summarizes the overall performance, with the CNN+GRU model outperforming BiLSTM in accuracy, precision, recall, and F1-score. The CNN+GRU model's superior ability to extract spatial features was instrumental in achieving higher accuracy, particularly given the reduced temporal resolution of the UT-HAR dataset.

Model	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
CNN+GRU	95.20	94.32	93.70	93.72
BiLSTM	66.80	64.35	59.29	59.26

TABLE VII: Overall performance metrics for CNN+GRU and BiLSTM models on the UT-HAR dataset.

Class-wise performance metrics are detailed in Table VIII. While the CNN+GRU model performed consistently across all classes, the BiLSTM model encountered significant challenges, particularly in activities such as "Walk" and "Box." These results underline the limitations of the BiLSTM model in handling datasets with constrained temporal information.

The confusion matrix for the CNN+GRU model, presented in Figure 6, highlights its robustness in classifying most activities accurately, with only minor misclassifications observed

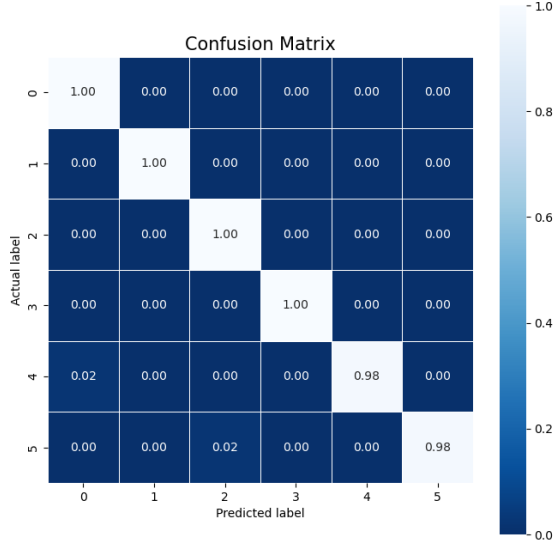


Fig. 4: Confusion matrix for CNN+GRU model on NTU-Fi HAR dataset, demonstrating high classification accuracy across all activities. Minor misclassifications are evident but do not significantly impact performance.

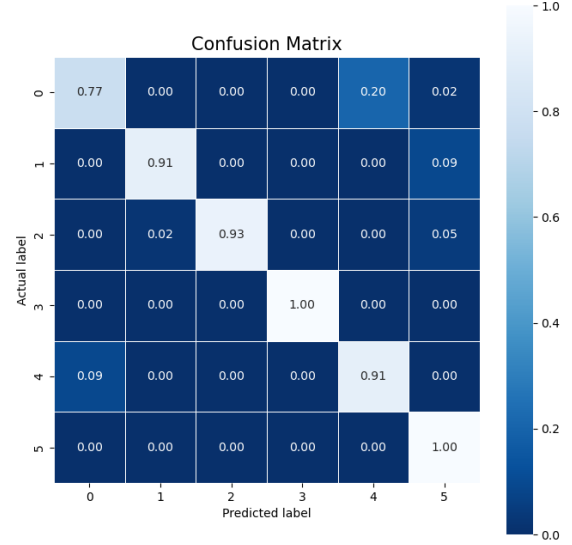


Fig. 5: Confusion matrix for BiLSTM model on NTU-Fi HAR dataset, showcasing strong classification performance with minor misclassifications in complex activities.

Class	CNN+GRU			BiLSTM		
	Precision	Recall	F1	Precision	Recall	F1
Clean	96	96	96	73	85	78
Fall	91	100	95	65	52	58
Run	98	99	98	79	70	74
Walk	87	85	86	38	28	32
Box	100	91	95	55	56	55
Circle	96	98	97	64	82	72

TABLE VIII: Class-wise performance metrics (%) for CNN+GRU and BiLSTM models on the UT-HAR dataset.

in activities such as "Clean" and "Fall." Figure 7 illustrates the confusion matrix for the BiLSTM model, which reveals substantial misclassifications in activities like "Walk" and "Box," underscoring its limited suitability for datasets with restricted temporal granularity.

C. Summary

In summary, the CNN+GRU model outperformed BiLSTM across both datasets, particularly excelling on the UT-HAR dataset due to its advanced spatial feature extraction capabilities. While the BiLSTM model demonstrated strengths in capturing temporal dependencies on the NTU-Fi HAR dataset, its overall performance remained inferior to the CNN+GRU model, particularly in scenarios requiring efficient handling of spatial features.

D. Discussion

The experimental results underscore the critical importance of aligning model architectures with dataset characteristics to achieve optimal Human Activity Recognition (HAR) performance. The CNN+GRU model consistently outperformed the BiLSTM model on the UT-HAR dataset, highlighting its superior ability to extract spatial features from segmented and lower-resolution CSI data. Conversely, the BiLSTM model excelled on the NTU-Fi HAR dataset due to its ability to capture long-term temporal dependencies inherent in high-resolution CSI data.

The observed performance differences between the NTU_Fi_HAR and UT_HAR datasets stem from several critical factors, including the resolution and quality of the collected CSI data, the applied preprocessing techniques, and the inherent properties of the datasets. These findings provide valuable insights into the interplay between dataset characteristics and model architecture in HAR applications.

1) Key Contributors to Performance:

a) *CSI Collection Tools.*: The quality and resolution of the CSI data play a pivotal role in model performance. The NTU_Fi_HAR dataset, collected using the Atheros CSI Tool, offers high-resolution data with 114 subcarriers per antenna pair. This high granularity enables models to capture subtle patterns associated with human activities, benefiting architectures like BiLSTM that excel in temporal feature extraction. In contrast, the UT_HAR dataset, collected using the Intel 5300 NIC, provides lower-resolution data with fewer subcarriers, which limits the ability of temporal models like BiLSTM

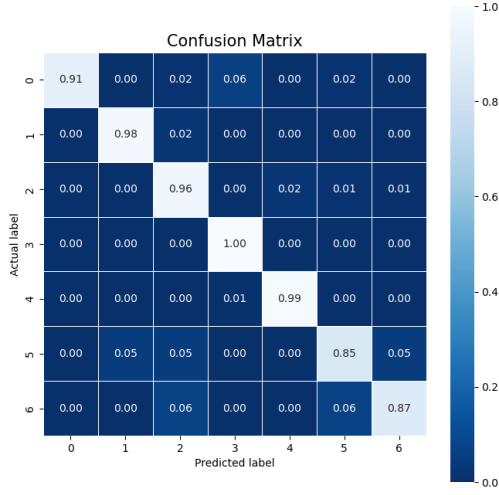


Fig. 6: Confusion matrix for CNN+GRU model on UT-HAR dataset, indicating strong spatial feature extraction capabilities with minimal misclassifications.

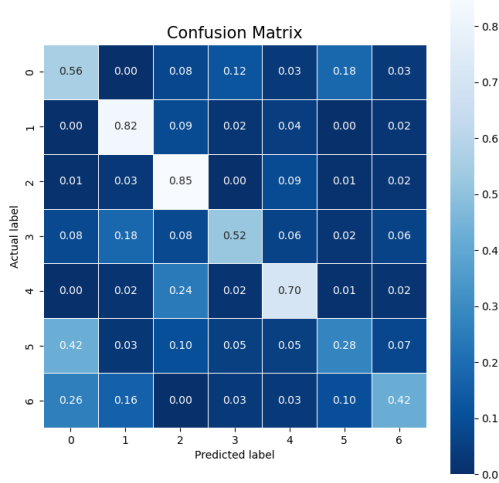


Fig. 7: Confusion matrix for BiLSTM model on UT-HAR dataset, showing substantial misclassifications in activities with less distinct temporal dependencies.

to achieve optimal performance. This limitation, however, is better managed by the CNN+GRU model, which effectively extracts spatial features even from lower-resolution data.

b) Preprocessing Techniques.: The preprocessing pipeline significantly influences model performance by enhancing the quality and interpretability of raw CSI data. The NTU_Fi_HAR dataset benefited from comprehensive

preprocessing, including:

- **Noise Reduction:** High-pass filters removed low-frequency noise, improving signal clarity.
- **Normalization:** Scaling of amplitude and phase data ensured consistency across samples.
- **Feature Extraction:** Fourier and wavelet transform captured critical frequency and time-domain features.
- **Processed Doppler Representation:** Advanced techniques, such as Body-coordinate Velocity Profiles (BVP), reduced environmental dependencies while emphasizing human motion.

In contrast, preprocessing for the UT_HAR dataset was more basic, primarily involving sliding window segmentation and statistical feature extraction. While effective, this limited preprocessing pipeline constrained the BiLSTM model's ability to utilize the temporal aspects of the data fully.

c) Dataset Characteristics.: The NTU_Fi_HAR dataset's high resolution aligns well with BiLSTM's ability to model detailed temporal patterns, resulting in superior performance for complex, time-sensitive activities. On the other hand, the lower resolution of the UT_HAR dataset posed challenges for BiLSTM, while the CNN+GRU model adapted well by leveraging its combined spatial and temporal feature extraction capabilities.

d) Temporal and Spatial Feature Extraction.: The CNN+GRU model demonstrated its effectiveness in scenarios where spatial patterns dominate, as seen in the UT-HAR dataset. Its convolutional layers efficiently captured spatial features, while the GRU layers modeled temporal dynamics. In contrast, the BiLSTM model excelled in capturing temporal dependencies in the NTU-Fi HAR dataset, where the high resolution provided sufficient temporal granularity. These results underscore the importance of matching model architecture to the specific demands of the dataset.

2) Implications of the Study: The findings highlight the potential of hybrid deep learning architectures, such as CNN+GRU, for real-time HAR applications, particularly in resource-constrained environments. High-resolution datasets like NTU_Fi_HAR benefit from advanced temporal modeling techniques, while lower-resolution datasets like UT_HAR require robust spatial feature extraction. These results emphasize the importance of high-quality data collection tools and preprocessing pipelines in enhancing model performance.

V. LIMITATIONS AND FUTURE DIRECTIONS

Despite the promising results, this study has certain limitations that present opportunities for future research:

- **Dataset Diversity:** The datasets utilized in this study primarily represent controlled environments. Extending the evaluation to more diverse and dynamic scenarios, such as outdoor settings or crowded environments, would improve the generalizability of the findings.
- **Computational Constraints:** While the BiLSTM model demonstrated strong performance on high-resolution data, its higher inference times and memory usage limit its

applicability in resource-constrained devices. Future work could focus on optimizing BiLSTM for such environments.

- **Temporal Variability:** Sliding window segmentation, while effective, may overlook finer temporal variations. Advanced temporal modeling techniques, such as attention mechanisms or transformers, could enhance temporal feature extraction and improve model accuracy.

Future research should explore integrating attention mechanisms and transformer architectures to improve temporal and spatial feature extraction. Additionally, testing these models on large-scale, real-world datasets would validate their scalability and robustness for practical applications.

VI. CONCLUSION

This study compares the BiLSTM and CNN+GRU architectures on UT-HAR and NTU-Fi HAR datasets, emphasizing how dataset characteristics and preprocessing workflows shape the performance of deep learning-based HAR models. The CNN+GRU model excels in scenarios where spatial feature extraction is paramount (UT-HAR), while the BiLSTM model demonstrates superior performance with high-resolution temporal data (NTU-Fi HAR). These findings underscore the importance of aligning model choice with data characteristics in real-world deployments, such as smart homes and healthcare monitoring systems. This research provides practical insights into choosing and tailoring advanced deep learning models for robust, scalable, and efficient HAR applications.

REFERENCES

- [1] D. Thakur and A. Pal, "Human activity recognition: Trends and challenges," *Activity Recognition and Prediction for Smart IoT Environments*, pp. 161–182, 2024.
- [2] P. Kumar, S. Chauhan, and L. K. Awasthi, "Human activity recognition (har) using deep learning: Review, methodologies, progress and future research directions," *Archives of Computational Methods in Engineering*, vol. 31, no. 1, pp. 179–219, 2024.
- [3] I. Ahmad, A. Ullah, and W. Choi, "Wifi-based human sensing with deep learning: Recent advances, challenges, and opportunities," *IEEE Open Journal of the Communications Society*, 2024.
- [4] Y. Ge, A. Taha, S. A. Shah, K. Dashtipour, S. Zhu, J. Cooper, Q. H. Abbasi, and M. A. Imran, "Contactless wifi sensing and monitoring for future healthcare-emerging trends, challenges, and opportunities," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 171–191, 2022.
- [5] S. K. Challa, A. Kumar, and V. B. Semwal, "A multibranch cnn-bilstm model for human activity recognition using wearable sensor data," *The Visual Computer*, vol. 38, no. 12, pp. 4095–4109, 2022.
- [6] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your csi: A channel state information extraction platform for modern wi-fi chipsets," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, pp. 21–28, 2019.
- [7] J. Yang, X. Chen, H. Zou, C. X. Lu, D. Wang, S. Sun, and L. Xie, "Sensefi: A library and benchmark on deep-learning-empowered wifi human sensing," *Patterns*, vol. 4, no. 3, 2023.
- [8] H. Shahverdi, P. F. Moshiri, M. Nabati, R. Asvadi, and S. A. Ghorashi, "A csi-based human activity recognition using canny edge detector," in *Human Activity and Behavior Analysis*, pp. 67–82, CRC Press, 2024.
- [9] H. El Zein, F. Mourad-Chehade, and H. Amoud, "Csi-based human activity recognition via lightweight cnn model and data augmentation," *IEEE Sensors Journal*, 2024.
- [10] M. M. Islam, S. Nooruddin, and F. Karray, "Multimodal human activity recognition for smart healthcare applications," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 196–203, 2022.
- [11] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, "Bi-lstm network for multimodal continuous human activity recognition and fall detection," *IEEE Sensors Journal*, vol. 20, no. 3, pp. 1191–1201, 2019.
- [12] Y.-W. Kim, K.-L. Joa, H.-Y. Jeong, and S. Lee, "Wearable imu-based human activity recognition algorithm for clinical balance assessment using 1d-cnn and gru ensemble model," *Sensors*, vol. 21, no. 22, p. 7628, 2021.
- [13] Y. Liu, S. Li, J. Yu, A. Dong, L. Zhang, C. Zhang, and Y. Cao, "Wifi sensing for drastic activity recognition with cnn-bilstm architecture," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 706–711, 2022.
- [14] D. Bouchabou, S. M. Nguyen, C. Lohr, B. LeDuc, and I. Kanellos, "A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning," *Sensors*, vol. 21, no. 18, p. 6037, 2021.
- [15] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.