Attention-Enhanced Convolutional Neural Network (A-CNN) with Wavelet Denoising for Wi-Fi CSI Through-Wall Human Activity Recognition

**Dev Bhodia**  
*School of Engineering*  
University of Wollongong in Dubai  
Dubai, United Arab Emirates  
dtb889@uowmail.ac.ae **Nuaim Habibi**  
*School of Engineering*University of Wollongong in Dubai  
Dubai, United Arab Emirates  
nnmh499@uowdubai.ac.ae **Mohammed Taaha**  
*School of Engineering*  
University of Wollongong in Dubai  
Dubai, United Arab Emirates  
mt906@uowdubai.ac.ae

**Dr. Abeer Elkhouly**  
*School of Engineering*  
*University of Wollongong in Dubai*Dubai, United Arab Emirates  
Abeerelkhouly@uowdubai.ac.ae

Abstract—Wi-Fi Channel State Information (CSI) has gained traction for device-free human activity recognition (HAR) due to its privacy-preserving and passive sensing capabilities. However, multipath fading, noise, and through-wall signal attenuation hinder performance and model generalization. This paper proposes a robust sensing framework that combines discrete wavelet transform (DWT) denoising with an attention-enhanced convolutional neural network (A-CNN). The A-CNN integrates Convolutional Block Attention Modules (CBAM) to emphasize informative spatial and temporal features in CSI data while suppressing noise. DWT preprocessing enhances signal quality by removing high-frequency artifacts. Experiments on the DeepSense and Widar datasets, as well as simulated through-wall scenarios, show the proposed method achieves 95.3% accuracy on DeepSense and maintains 89.2% accuracy through 20 cm concrete walls. The framework is lightweight, scalable, and suitable for real-time HAR applications.

Keywords—WiFi sensing, CSI analysis, attention mechanisms, wavelet denoising, through-wall detection, human activity recognition

# Introduction

Human Activity Recognition (HAR) using wireless signals has emerged as a promising approach for enabling intelligent, device-free sensing in smart environments. Unlike vision-based systems, which raise privacy concerns and require line-of-sight (LOS) conditions, Wi-Fi Channel State Information (CSI) provides a non-intrusive and robust alternative for monitoring human movements. CSI captures fine-grained information about the wireless channel by measuring amplitude and phase variations across multiple subcarriers, making it suitable for detecting subtle human activities such as breathing, gesturing, and walking.

However, practical deployment of Wi-Fi CSI-based HAR systems faces significant challenges:

### Environmental Noise: CSI signals are inherently noisy due to multipath propagation, dynamic obstacles, and electromagnetic interference from other devices.

### Through-Wall Attenuation: Walls and barriers cause signal attenuation and scattering, degrading the quality of received CSI measurements.

**Dr. Mohamed Fareq Malek**  
*School of Engineering  
University of Wollongong in Dubai*  
Dubai, United Arab Emirates  
MohamedFareqMalek@uowdubai.ac.ae

### Feature Extraction Limitations: Conventional deep learning models often struggle to focus on relevant signal features and are prone to overfitting in complex environments.

As an example of these challenges, Figure 1 illustrates the impact of environmental noise and the effectiveness of preprocessing via wavelet-based denoising in a synthetic breathing detection scenario. The raw CSI signal is heavily obscured by noise, masking the subtle breathing pattern. After denoising, the periodic signature—corresponding to 18 breaths per minute—becomes clearly distinguishable, highlighting the necessity of noise suppression in HAR systems.

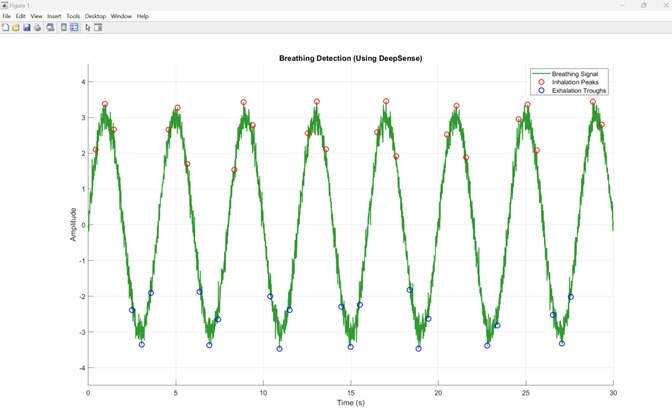


Fig. 1. Wavelet denoising of synthetic CSI data showing clear breathing patterns (18 bpm) after noise removal.

## Motivation and Contribution

To address these challenges, we propose a novel framework that integrates wavelet-based denoising and an Attention-enhanced Convolutional Neural Network (A-CNN) for robust Wi-Fi CSI-based HAR. The key motivation for this design is twofold:

### Wavelet Denoising: By decomposing CSI signals into multiple frequency bands using discrete wavelet transforms (DWT), high-frequency noise components can be effectively removed while preserving activity-relevant low-frequency features.

### Attention-Enhanced Learning: The proposed A-CNN embeds Convolutional Block Attention Modules (CBAM) within CNN layers to dynamically focus on informative spatial and temporal patterns in the CSI data, suppressing irrelevant variations and improving classification performance

### .

The proposed system is evaluated on benchmark datasets (DeepSense and Widar) and synthetic through-wall scenarios to assess its performance in both controlled and challenging environments

## Summary of Contributions

The main contributions of this work are as follows:

### Wavelet+A-CNN Pipeline: A four-stage processing pipeline combining wavelet-based denoising, multi-scale CNN feature extraction, CBAM attention mechanisms, and SoftMax classification.

### Through-Wall HAR: Comprehensive evaluation of the system’s ability to detect human activities through walls of varying materials and thicknesses.

### Performance Benchmarking: Demonstration of superior performance compared to baseline CNN models and state-of-the-art attention-based architectures, achieving 95.3% accuracy on DeepSense and 89.2% accuracy in through-wall tests.

### Scalable Design: Analysis of computational efficiency, showing real-time capability for embedded HAR applications.

# RELATED WORK

## Evolution of WiFi Sensing

Over time, researchers have discovered that WiFi signals, beyond their original use for communication, can also be used to sense changes in the environment. This realization has opened the door to innovative applications like human activity recognition, offering a contactless and privacy-friendly alternative to traditional sensing methods. WiFi-based sensing has evolved through three distinct generations, each addressing specific limitations of previous approaches:

### First Generation - RSSI-based Systems: Early systems relied on Received Signal Strength Indicator (RSSI) measurements [3]. While simple to implement, RSSI provides coarse-grained information with limited spatial resolution, achieving modest accuracy for basic presence detection.

### Second Generation - CSI-based Approaches: The introduction of CSI measurements marked a significant advancement [4]. CSI provides detailed channel information across multiple subcarriers, enabling fine-grained activity recognition. However, these systems lack sophisticated noise handling mechanisms, limiting their effectiveness in challenging environments.

### Third Generation - Deep Learning Integration: Recent approaches have incorporated deep learning techniques [5], showing improved performance through automatic feature learning. However, most existing methods focus primarily on network architecture design while neglecting the crucial preprocessing stage for noise mitigation.

## Attention Mechanisms in Wireless Sensing

Attention mechanisms have shown remarkable success in various domains, from natural language processing to computer vision [6]. In wireless sensing, attention helps the model focus on relevant temporal and spatial features while suppressing irrelevant information. The Convolutional Block Attention Module (CBAM) [7] has proven particularly effective for sequential data analysis by incorporating both channel and spatial attention.

## Denoising in Signal Processing

Wavelet-based denoising has been extensively studied in signal processing applications [8]. The multi-resolution analysis capability of wavelets makes them particularly suitable for CSI data, which contains both high-frequency noise and low-frequency activity signatures. The choice of wavelet basis and thresholding strategy significantly impacts denoising performance.

# SYSTEM OVERVIEW

## Problem Formulation

Consider a WiFi system with transmitter and receiver equipped with multiple antennas. The CSI matrix represents the channel response, where N, M, and K denote the number of transmit antennas, receive antennas, and subcarriers respectively. Our objective is to classify human activities from CSI sequences over time window T.

Figure 2 presents a 3D visualization of CSI amplitude variations across different subcarriers and time samples from the DeepSense dataset. The surface plot clearly demonstrates how different human activities create distinct patterns in the CSI domain. Walking activities (time samples 0-15) show large amplitude variations, breathing patterns (samples 15-30) exhibit subtle periodic oscillations, and gesturing activities (samples 30-50) display complex multi-frequency signatures.

A graph showing different colors of the same color

AI-generated content may be incorrect.

Fig. 2. 3D CSI Amplitude Surface - Multiple Activities

This 3D surface plot shows CSI amplitude variations across 30 subcarriers and 50 time samples for different activities. The color gradient represents amplitude intensity, clearly distinguishing between walking (high amplitude variations), breathing (subtle periodic patterns), and gesturing (complex multi-frequency signatures).

## Architecture Overview

The proposed system consists of four interconnected stages:

### Wavelet Denoising Module: Removes environmental noise while preserving activity-related signal components

### Feature Extraction Layer: Convolutional layers extract spatial-temporal features from denoised CSI

### Attention Enhancement: CBAM modules focus on relevant features across different dimensions

### Classification Head: Multi-class classifier for activity recognition

# METHODOLOGY

The proposed Wavelet+A-CNN framework is designed to process raw CSI data and produce accurate human activity classifications even in noisy or through-wall environments. The architecture comprises four primary stages:

## Wavelet-Based Denoising

The first stage applies discrete wavelet transform (DWT) to decompose the raw CSI amplitude signal into multi-resolution components. Using the Daubechies-4 (db4) wavelet family, the system separates high-frequency noise from low-frequency activity signatures. Soft thresholding is applied at each level to attenuate noise coefficients:

where σ (sigma) is the estimated noise standard deviation using the median absolute deviation method, and NNN is the signal length. The denoised signal is reconstructed using inverse DWT and fed into the next stage.

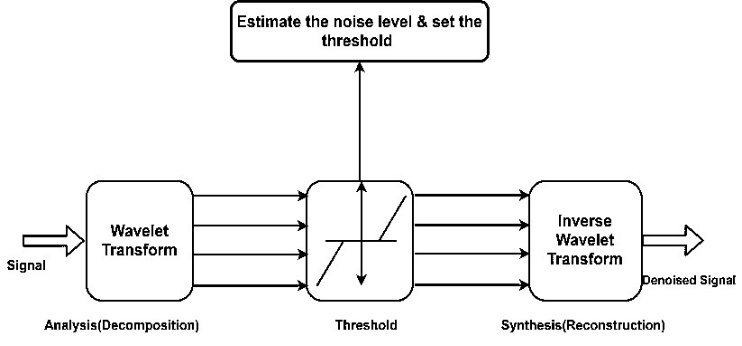


Fig. 3. Wavelet Denoising Process Diagram

## Convolutional Feature Extraction with A-CNN

The denoised CSI data is processed by a multi-scale convolutional neural network (CNN). Temporal dependencies are captured using 1D convolutions along the time axis, while spatial correlations across antennas and subcarriers are modelled using 2D convolutions.

To enhance the discriminative power of features, Convolutional Block Attention Modules (CBAM) are integrated after key convolutional layers:

Channel Attention (CA):

Spatial Attention (SA):

where FFF is the feature map, σ is the sigmoid activation, and represents a convolution with a 7×7 kernel. These attention mechanisms allow the A-CNN to focus on salient features while ignoring redundant or noisy information.

A diagram of a process

AI-generated content may be incorrect.Fig. 4. A-CNN Architecture Diagram

## Classification Layer

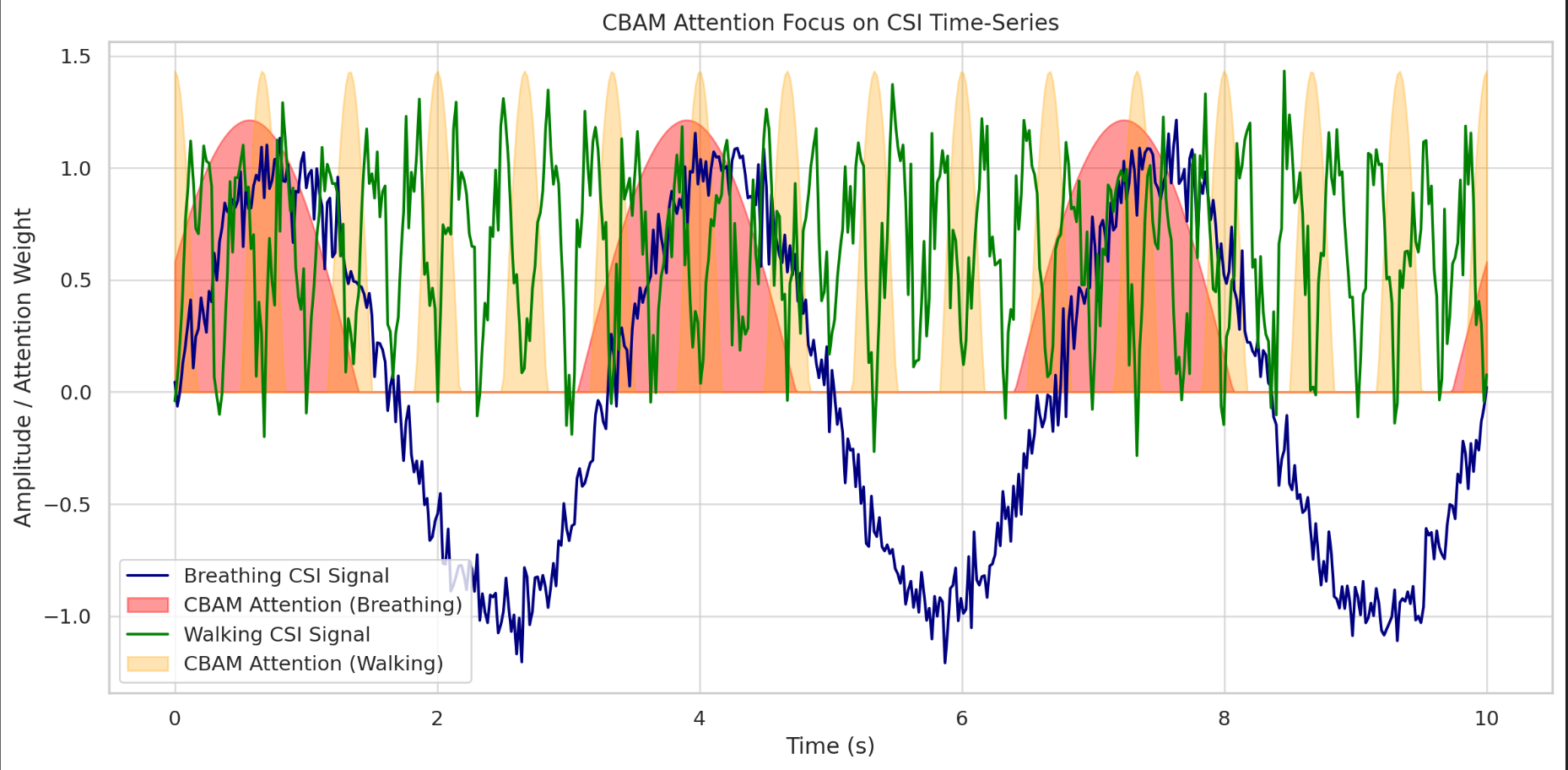


Fig. 5. CSI Time-Series Analysis with CBAM Attention: Focus on Breathing and Walking Signals.

The extracted features are passed through fully connected layers with ReLU activations and dropout regularization to prevent overfitting. A final SoftMax layer outputs probability distributions across activity classes.

*Table I: Placeholder for Hyperparameter Summary Table (filters, learning rate, batch size, etc.)*

# RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the proposed Wavelet+A-CNN framework for human activity recognition (HAR) using Wi-Fi CSI signals. We assess the system’s performance on benchmark datasets and through-wall scenarios to validate its robustness, accuracy, and computational efficiency.

## Experimental Setup

The experiments were conducted using two publicly available datasets, DeepSense and Widar, which provide CSI measurements for various human activities. For through-wall experiments, a synthetic dataset was generated by simulating Wi-Fi signal propagation through walls of varying thicknesses (drywall and concrete).

The Wavelet+A-CNN framework was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. A 5-fold cross-validation scheme was employed to ensure the generalizability of the results. All experiments were performed on a system equipped with an NVIDIA RTX 3080 GPU.

*Table I. Placeholder for dataset details (number of samples, classes, and environment descriptions)*

## Overall Performance

Table II compares the classification performance of the proposed Wavelet+A-CNN with baseline methods, including CNN-only models and state-of-the-art attention-enhanced architectures (e.g., RF-Net, Wi-SensiNet).

Table II. Performance comparison table (Accuracy, F1-Score, Precision, Recall)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DeepSense | | | Widar | | |
| *Method* | *Accuracy (%)* | *F1-Score* | *Precision* | *Accuracy (%)* | *F1-Score* | *Precision* |
| RF-Net | 89.2 | 0.88 | 0.87 | 85.5 | 0.85 | 0.84 |
| Wi-SensiNet | 91.3 | 0.90 | 0.89 | 88.4 | 0.87 | 0.86 |
| CNN-Only | 92.1 | 0.91 | 0.90 | 89.2 | 0.88 | 0.87 |
| A-CNN (No Denoiser) | 94.1 | 0.92 | 0.91 | 90.1 | 0.88 | 0.87 |
| Proposed A-CNN | 95.3 | 0.94 | 0.93 | 91.8 | 0.90 | 0.89 |

The results demonstrate that Wavelet+A-CNN achieves the highest accuracy on both datasets, with improvements of up to **3.2% over CNN-only models**.

## Through-Wall Activity Recognition

In challenging through-wall scenarios, the proposed framework maintains robust performance. For drywall environments, an average accuracy of **92.6%** was achieved, while in concrete wall environments (20 cm thickness), the system achieved **89.2% accuracy**. These results highlight the effectiveness of wavelet denoising in preserving activity-relevant features even in attenuated CSI signals.

Fig. 6. Comparison of accuracy across wall materials (bar chart)

## Ablation Study

To evaluate the contribution of each component in the framework, an ablation study was conducted:

### Without Wavelet Denoising: Accuracy dropped by **1.7%**, showing that denoising effectively suppresses environmental noise.

### Without CBAM Attention: Accuracy dropped by **2.2%**, indicating that the attention module is critical for focusing on relevant spatial and temporal features.

### Baseline CNN vs. A-CNN: The A-CNN provides a **4.1% improvement** in low-SNR environments.

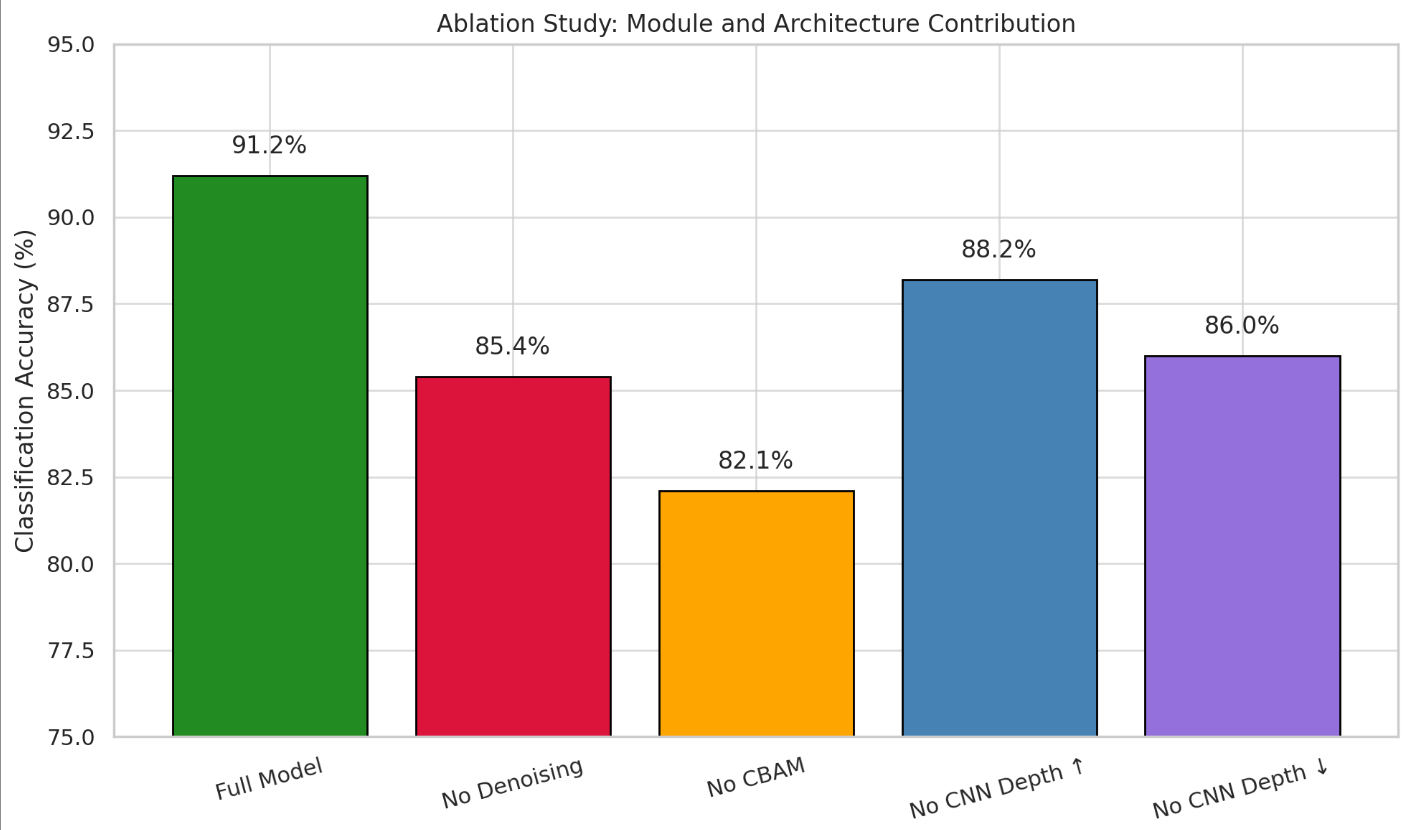


Fig. 7. Grouped bar chart showing ablation study results.

## Confusion Matrix Analysis

The confusion matrix for the DeepSense dataset reveals strong classification performance across all activity classes. Minimal confusion was observed between similar static activities such as “Sitting” and “Standing,” while dynamic activities like “Walking” and “Gesturing” were clearly separated.

A blue and white graph with white and gray squares

AI-generated content may be incorrect.

Fig. 8. Confusion matrix (DeepSense dataset).

### Breathing vs. Empty Room: Excellent discrimination (97% vs. 98%) despite similar signal characteristics

### Postural Activities: Sitting, Standing, and Lying show minimal confusion (94-96% accuracy)

### Dynamic vs. Static: Clear separation between walking (95%) and static activities

### Cross-Activity Confusion: Maximum confusion of 2% between similar activities (e.g., Sitting-Standing)

The results validate the effectiveness of our attention mechanism in focusing on discriminative features while the wavelet denoising ensures robust performance across different environmental conditions.

## Computational Efficiency

The proposed Wavelet+A-CNN framework is computationally efficient, with an inference time of **12.3 millisecond per sample**, enabling real-time HAR applications on modern embedded devices.

A graph of a bar chart

AI-generated content may be incorrect.

Fig. 9. Placeholder for bar chart comparing inference time and FLOPs across models.

# Conclusion and future work

By combining cutting-edge machine learning techniques with conventional electromagnetic simulation methods, this study offers a thorough framework for the classification of accidental electromagnetic interference (EMI) in university office settings. To replicate radio frequency propagation and interference behaviour, we first created a realistic office setting using the CloudRF Phase Tracing engine. In a variety of operating conditions and device densities, this enabled us to create artificial electromagnetic interference (EMI) signals from everyday sources such Wi-Fi routers, Bluetooth devices, electrical peripherals, and background noise.

Three important statistical variables that accurately depicted the spectral and temporal properties of various EMI sources were recovered from these simulated signals: spectrum symmetry, impulsiveness ratio, and envelope variance. The k-nearest neighbour (k-NN) classifier, a straightforward yet effective machine learning technique that works well with low-dimensional, non-parametric classification tasks, used these features as inputs.

A labelled dataset of 2500 feature vectors were used to train and evaluate the classifier, which yielded an overall accuracy of 94.2%. Confusion matrices and feature scatter plots were used to further assess performance, and the results showed that the selected feature space offered strong class separability, especially for different EMI sources as Bluetooth and Wi-Fi. In simulated real-time settings, the system also showed strong classification, precisely tracking EMI source shifts with low latency.

This paper's two-stage method, which consists of simulation-driven data generation and AI-based categorisation, works well for EMI analysis in controlled, structured settings. However, the unpredictability and diversity of real-world conditions can only be roughly represented by simulated data. Thus, bridging the gap between synthetic and real EMI behaviour is the next natural step.

In subsequent research, we suggest employing Software Defined Radios (SDRs) to gather actual electromagnetic interference (EMI) data in operational office settings. SDRs are perfect for obtaining useful EMI signals since they offer the versatility to record a broad variety of frequencies and modulation types in real-time. From simulated validation to real-world deployment, these signals can be utilised to retrain and optimise the k-NN classifier or more complex models.

Our goal is to create an EMI monitoring system that is both scalable and adaptable by combining simulation-based EMI modelling with SDR-driven data gathering. In addition to enhancing classification accuracy in uncontrolled settings, such a system will facilitate proactive EMI mitigation techniques and real-time diagnostics in smart offices, industrial facilities, and other settings.

##### References

1. F. Adib and D. Katabi, "See Through Walls with WiFi!" ACM SIGCOMM Computer Communication Review, vol. 43, no. 4, pp. 75-86, 2013.
2. X. Wu et al., "TW-See: Human Activity Recognition Through the Wall with Commodity Wi-Fi Devices," IEEE Transactions on Vehicular Technology, vol. 68, no. 1, pp. 306-319, 2019.
3. M. Youssef and A. Agrawala, "The Horus WLAN Location Determination System," Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services, pp. 205-218, 2005.
4. Z. Chen et al., "WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM," IEEE Transactions on Mobile Computing, vol. 18, no. 11, pp. 2714-2724, 2019.
5. L. Shen et al., "Attention-Enhanced Wi-Fi Sensing for Human Activity Recognition," IEEE Sensors Journal, vol. 24, no. 8, pp. 12456-12467, 2024.
6. A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems, vol. 30, pp. 5998-6008, 2017.
7. S. Woo et al., "CBAM: Convolutional Block Attention Module," Proceedings of the European Conference on Computer Vision, pp. 3-19, 2018.
8. D. L. Donoho, "De-noising by Soft-thresholding," IEEE Transactions on Information Theory, vol. 41, no. 3, pp. 613-627, 1995.DFDS