Sensors and Materials, Vol. 37, No. 9 (2025) 3869–3879
MYU Tokyo

S & M 4158

EnSta-Fi: Ensemble Stacking-based Human Activity Recognition by Leveraging Channel State Information Amplitude in Wi-Fi Sensing

Dwi Joko Suroso,* Aisyah A. Susanto, Liyas F.R. Tarigan, and Nazrul Effendy

Department of Nuclear Engineering and Engineering Physics, Universitas Gadjah Mada,
Jl. Grafika No. 2, Yogyakarta 55281, Indonesia

(Received June 2, 2025; accepted August 4, 2025)

Keywords: Wi-Fi sensing, channel state information, human activity recognition, ensemble stacking

Wi-Fi sensing-based human activity recognition (HAR) research has grown over the last decade. While conventional Wi-Fi sensing employs complex and high-cost devices, we focused on lightweight Wi-Fi sensing with ESP32 for channel state information (CSI) data collection. We aim to keep the setup minimal by considering a single antenna for relatively static activities, e.g., sitting, standing, and light walking. Thus, we leverage CSI amplitude and propose EnSta-Fi, a classification model based on ensemble stacking for Wi-Fi sensing, combining baseline machine learning models, i.e., k-nearest neighbor (kNN) and support vector machine (SVM), with the logistic regression as a final classifier. Our method includes the actual measurement setup to collect CSI, ensemble stacking model training, and evaluation. Results showed that EnSta-Fi outperforms individual kNN and SVM in terms of activity classification performance with accuracy improvements of 2.29 and 1.19%, respectively. Moreover, compared with deep learning models, e.g., bidirectional gated recurrent unit (Bi-GRU) and convolutional neural network (CNN), EnSta-Fi achieves higher accuracy and less computational time (40 and 2.5 times faster than Bi-GRU and CNN, respectively). From the results of our proposed method, we can conclude that EnSta-Fi is suitable for the real deployment of the HAR system, where straightforward setup, light weight, high accuracy, and low computational complexity are emphasized.

1. Introduction

The human activity recognition (HAR) system is helpful for many applications, e.g., health, smart building, security surveillance, human–computer interaction, and entertainment, and for object localization and tracking.[1–4] Over time, both attached or tagged physical sensors have been primarily used to record body movements and conditions. However, wearing physical sensors that generally need a long time to operate will impact users' convenience and prevent flexible movements, especially for applications without any attached devices.[5] With the growing research capabilities on the device-free monitoring of the HAR system, primarily by vision or wireless, HAR is entering a new era of sensor deployment.

*Corresponding author: e-mail: dwi.jokosuroso@ugm.ac.id
https://doi.org/10.18494/SAM5740

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

Vision-based HAR primarily employs cameras for sensing. There are also RGB cameras and infrared projectors and detectors for motion and sensing input device, i.e., Microsoft Kinect.(6) However, this technology requires additional setup and limited environment coverage. Moreover, the main issues of using a camera as a sensor for HAR are its privacy obstruction nature, the fact that it cannot work in different rooms blocked by walls, and the need for specific lighting conditions for optimal recognition results.(6) Another device-free approach is implementing radio frequency (RF)-based technology, e.g., Wi-Fi, Bluetooth, radio frequency identification (RFID), radar, or ultrawide band (UWB). In this regard, Wi-Fi is the most common and widely used and has different features, size, cost, and capabilities.

Wi-Fi sensing is a term used for detecting environment changes by utilizing Wi-Fi signals, including the changes in its amplitude, phase, or temporal properties.(7) These changes can be represented by the channel state information (CSI), which shows the changes in environment–signal interaction due to signal propagation and multipath effects.(8) The benefits of using Wi-Fi sensing are as follows: it is a device-free method with high flexibility, it works well under both line-of-sight (LoS) and non-LoS (through the wall) conditions and under every lighting condition, and it has no issue with privacy obstruction.(9,10)

Wi-Fi sensing research has been attracting researchers' attention over the last decade. Moreover, the advancement of model development on machine learning (ML) or deep learning (DL) has supported the research on Wi-Fi sensing. Some beneficial applications of Wi-Fi sensing are fine-grained HAR, respiration monitoring, microgesture, heartbeat monitoring, sleep insight, intrusion detection, fall detection, indoor localization, energy efficiency in smart buildings, and security and safety in a car.(11,12)

Most of the published papers use standard Wi-Fi cards, e.g., network interface card (NIC) and AX cards taken or used directly from a personal computer, making the built system bulky and expensive. Most findings are helpful for the HAR system and can be the benchmark of Wi-Fi sensing. However, low-cost and lightweight Wi-Fi sensing is needed in some implementations that need fast and straightforward setup and calibration. As far as the authors are concerned, few works focus on lightweight and high-flexibility systems. Hernandez and Bulut proposed the ESP32-based HAR system with CSI as the central part.(13,14) CSI shows suitable signal parameters for Wi-Fi sensing as it comprises the signal's amplitude and phase for each subcarrier at the time the signal is transmitted. CSI is based on the orthogonal frequency division multiplexing (OFDM) technique used in the Wi-Fi standard to spread signals within the set subcarriers in a certain bandwidth.(15)

In this study, we aim to build a lightweight HAR system by utilizing ESP32 and CSI tools for CSI extraction in a single-antenna system setup. As a system limitation by nature, the CSI amplitude can only be used as the HAR parameter. To enhance the amplitude CSI, we need several preprocessing steps, including digital filtering, i.e., using Butterworth filters. For activity classification, we propose the ensemble stacking model EnSta-Fi formed from baseline ML models, e.g., k-nearest neighbor (kNN) and support vector machine (SVM), and a logistic regression as the final classifier. Ensemble models are robust yet light in computational perspectives.(16,17) Our considered activities are relatively static activities, e.g., sitting (transition between standing and sitting), standing (transition between sitting and standing), and slow walking.

The idea behind ensemble stacking is to achieve high accuracy without sacrificing the computational complexity in training and prediction times. Ensemble stacking has proven to be a highly effective method that is more stable, adaptable, and less prone to overfitting than traditional machine learning techniques.(16,18)

As our concern, the research gap that EnSta-Fi fills is that the leveraging solely on CSI amplitude is applied for the first time in the HAR system. Moreover, the implementation of ensemble stacking by combining ML models is still also limited in the HAR system. Thus, we highlight the contributions of EnSta-Fi in the Wi-Fi sensing-based HAR system as follows:

1. It is the first to leverage CSI amplitude in the single-antenna HAR system.
2. EnSta-Fi is built on the basis of SVM, kNN, and logistic regression.

To verify our proposal, we then compared EnSta-Fi with both individual k-NN and SVM and with DL models, e.g., bidirectional gated recurrent unit (Bi-GRU) and convolutional neural network (CNN). The comparison with DL enables us to see the practical approach on computational time (data training) and part of its accuracy.

This paper is organized as follows. First, we will introduce the problem statement and EnSta-Fi summary, followed by the system model and methods by discussing the Wi-Fi sensing, CSI, and classification models. Then, we will show and discuss the results of this study and finally provide conclusions.

2. Data, Materials, and Methods

The Wi-Fi sensing technique is depicted in Fig. 1, which shows two main approaches to Wi-Fi sensing: learning- and modeling-based.(16) We focused on shallow learning on the learning-based model to accommodate our approach to the low-cost, highly flexible, and lightweight HAR system.

2.1 CSI

CSI describes the condition of a wireless channel when a signal is sent from a transmitter (TX) to a receiver (RX). CSI allows the receiver to know how the signal changes as it passes

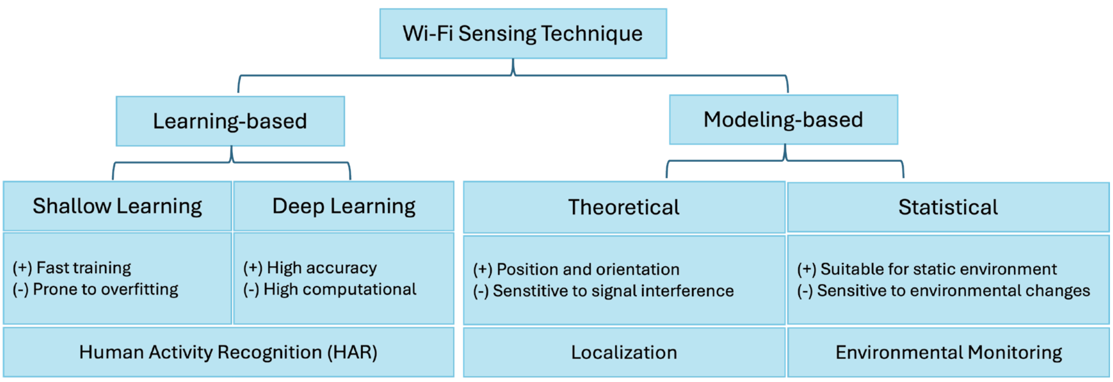

Fig. 1. (Color online) Wi-Fi sensing technique.

through various propagation paths or multipaths in a channel. CSI can be represented as the $\mathbb{H}$ matrix on the linear equation as shown in Eq. (1), where y is the vector for the received signal, x is the input vector from the pilot symbol at the transmitter, and the Gaussian noise vector is symbolized as η.[16]

$$y = \mathbb{H}x + \eta \tag{1}$$

CSI can then be seen as the channel frequency response of each subcarrier frequency $\boldsymbol{h}_i$:[17]

$$\boldsymbol{h}_i = A_i e^{j\phi_i}. \tag{2}$$

Here, $\boldsymbol{h}_i$ consists of both the real and imaginary parts, $\mathcal{R}(\boldsymbol{h}_i)$ and $\mathcal{I}(\boldsymbol{h}_i)$, respectively. A_i and ϕ_i represent the amplitude and phase shifting in the i-th subcarrier, respectively, which can be estimated as[7,19]

$$A_i = \sqrt{\mathcal{I}(\boldsymbol{h}_i)^2 + \mathcal{R}(\boldsymbol{h}_i)^2}, \tag{3}$$

$$\phi_i = a\tan 2\left(\mathcal{I}(\boldsymbol{h}_i), \mathcal{R}(\boldsymbol{h}_i)\right). \tag{4}$$

Application in an indoor environment yields variations, which affect the interaction between TX and RX. The multipath signal and propagation between TX and RX are complex in nature. This multipath physical signal can be defined as[7]

$$\boldsymbol{h}_i = \sum_{m=1}^{N} A_m e^{\frac{-2\pi f_i d_m}{c} + j\phi_m}. \tag{5}$$

In Eq. (5), the amplitude, phase, and distance of the mth path are denoted by A_m, ϕ_m, and d_m, respectively. The variable f_i represents the subcarrier frequencies, while c is the speed of light. On the basis of this perspective, the path can be categorized into two types: static and dynamic.

The static path (Ω_s) refers to the path where the signal reflects off stationary walls or immobile objects. Conversely, the dynamic path (Ω_d) represents the path influenced by the motion of a human body or moving objects. The first expression on the right-hand side of Eq. (6) is the static component ($\boldsymbol{h}_{static}$), whereas the next is the dynamic component ($\boldsymbol{h}_{dynamic}$).[7]

$$\boldsymbol{h}_i = \sum_{m \in \Omega_s} A_m e^{\frac{-2\pi f_i d_m}{c} + j\phi_m} + \sum_{n \in \Omega_d} A_n e^{\frac{-2\pi f_i d_n}{c} + j\phi_n} \tag{6}$$

2.2 HAR system configuration

Our proposal uses ESP32 as the main component in the data acquisition process, which acts as a sensing element in this system. The initial configuration of ESP32 involves designating the access point (AP) as TX and the station device (STA) as RX. Our CSI system begins by configuring ESP32 as AP and STA using the C and C++ programming languages and the ESP-IDF framework with Visual Studio Code (VS Code) as the development environment. ESP32 is then connected to Raspberry Pi via a serial connection. We developed a web-server-based system for data collection. The system runs on Raspberry Pi via a Docker container, including a web server, a database, and a web interface. After the system is initialized, the web interface can be opened via the IP address and port of Raspberry Pi.

2.3 Measurement campaign

CSI data collection was carried out in a classroom in our department. The measurement area was 5 × 5 m^2. Figure 2 shows CSI data acquisition; Raspberry Pi and ESP32 (STA) were placed in the front left corner of the room at approximately 2.5 m from the object. All devices were placed at a height of 0.8 m from the floor, and facing the center of the measurement area, where the object was located, in our proposal, one adult person (co-author) performed specific activities, namely, sitting, standing (both transition of standing-sitting or vice versa), and walking. In this research, each measurement was conducted for 20 s, and each activity was performed 10 times with certain activity procedure. The objectives were to detect Wi-Fi signal changes when the object was doing an activity. The parameter considered in this measurement was the change in amplitude resulting from the activity. Figure 3 shows the actual measurement campaign.

2.4 EnSta-Fi: ensemble stacking-based Wi-Fi sensing

Ensemble stacking classification is a machine learning method that combines multiple base models to improve prediction accuracy. In this approach, the outputs of the base models are used

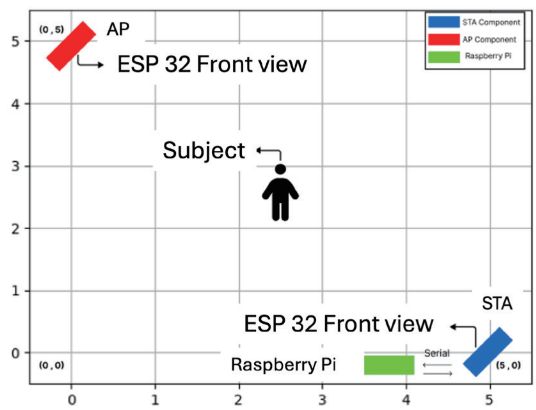

Fig. 2. (Color online) CSI data acquisition.

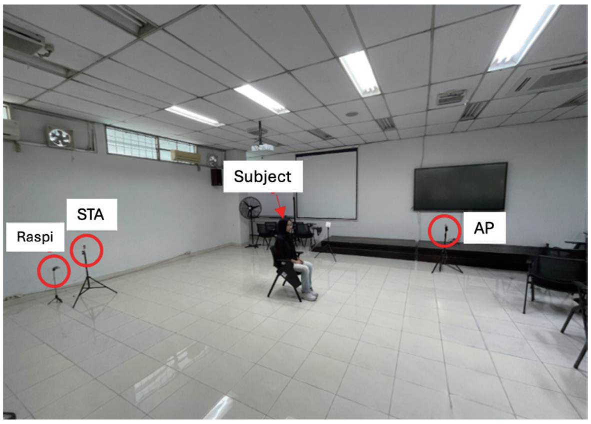

Fig. 3. (Color online) Actual measurement setup.

as inputs to an advanced model, called a metamodel, which is responsible for producing the final prediction.[16,18] The reason behind stacking baseline ML models, e.g., kNN and SVM with the meta-learner, i.e., logistic regression, is to combine their strong aspects. For instance, kNN is superior in local decision boundaries but struggles in high-dimensional data, and SVM is strong in high-dimensional data but sensitive to kernel type. Thus, stacking allows combining multiple diverse models and achieves higher performance prediction from the stacking process. Furthermore, stacking takes advantage of the base models' strengths, reducing the risk of overfitting and improving prediction accuracy.

In our approach, the collected data go through a preprocessing stage that begins with amplitude extraction and continues with filtering using a Butterworth filter for denoising CSI amplitude data. The preprocessed data are then trained using two baseline ML models, kNN and SVM. The output of these models is classified using logistic regression as a meta-learner or metamodel that takes predictions from Level-0 models to make final decisions regarding recognized activities. Figure 4 shows our ensemble stacking model architecture diagram.

The stacking process involves the following steps: (1) base models are trained with the original data, (2) predictions from the base models are made on a separate validation set, (3) metamodels are trained using predictions from the base models as input, and (4) final classification results are obtained on the basis of the metamodel. Thus, ensemble stacking is advantageous in its ability to combine different types of model. The parameter tuning to achieve a high-performance model was conducted in each baseline model. The hyperparameter tuning was performed by using GridSearchCV, resulting in the following setup: for KNN, the number of neigbors was equal to 6, the weight was made uniform, and the metrics distance was Euclidean, while for SVM, the regularized parameter C was 100, scale gamma was selected, and the kernel type was RBF. The model was built in the Google Colaboratory platform with the runtime type of Python 3 with the selected hardware accelerator CPU. This setting approached the lightweight implementation only on CPU instead of GPU.

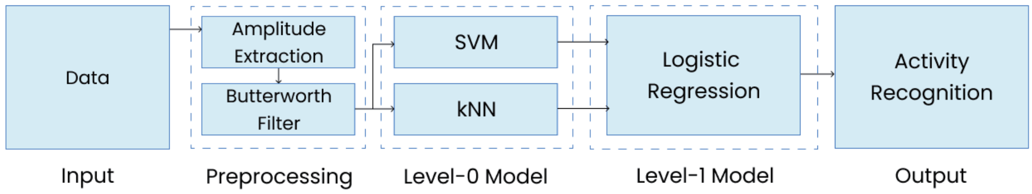

Fig. 4. (Color online) EnSta-Fi architecture.

3. Results and Discussion

3.1 CSI preprocessing data

We applied the Butterworth filter to remove noise and outliers in the raw CSI data. The cutoff parameter of the Butterworth filter is determined by varying values from 0.1 to 0.4 to determine the best value for denoising CSI data. This result shows that a cutoff frequency of 0.4 effectively compensates for noise reduction and preserves important features of CSI data without the significant loss of important information. Figure 5 shows CSI data after applying the Butterworth filter and showing CSI on five selected subcarriers for a faster observation of CSI's pattern changes.(19) The filtered CSI shows that the walking CSI signal is the most distinguishing characteristic with peaks and bottoms clearly showing signal changes. While sitting and standing have similar characteristics, the static CSI is represented by the empty-room CSI.

3.2 Performance results of EnSta-Fi

The following results were used to evaluate the performance of two individual models, SVM and kNN, in classifying various activities based on CSI data. Precision, recall, and F1-score metrics were used for performance evaluation. In addition, we compared the ensemble stacking model with the individual SVM and KNN models. Furthermore, the GRU and CNN models were also included in the comparison to position the proposed method in terms of performance with computational complexity. Table 1 shows our proposed ensemble stacking with the individual baseline MLs and DL.

Table 1 shows that in terms of all performance metrics, our proposed ensemble stacking is superior, especially in terms of accuracy and training time. The precision is slightly lower than that of GRU, and the prediction time is only 0.19 longer than that of CNN. kNN and SVM give overall good performance and computational complexity results with very short training and prediction times. Our proposed ensemble stacking improves these baseline MLs, especially in almost all performance metrics. This comparison is depicted in Fig. 6.

The red bar in Fig. 6 represents our ensemble stacking; with these results, our proposal stands between high performance and low computational time. The results indicate that the ensemble stacking approach outperforms both SVM and kNN in each metric, achieving an accuracy of

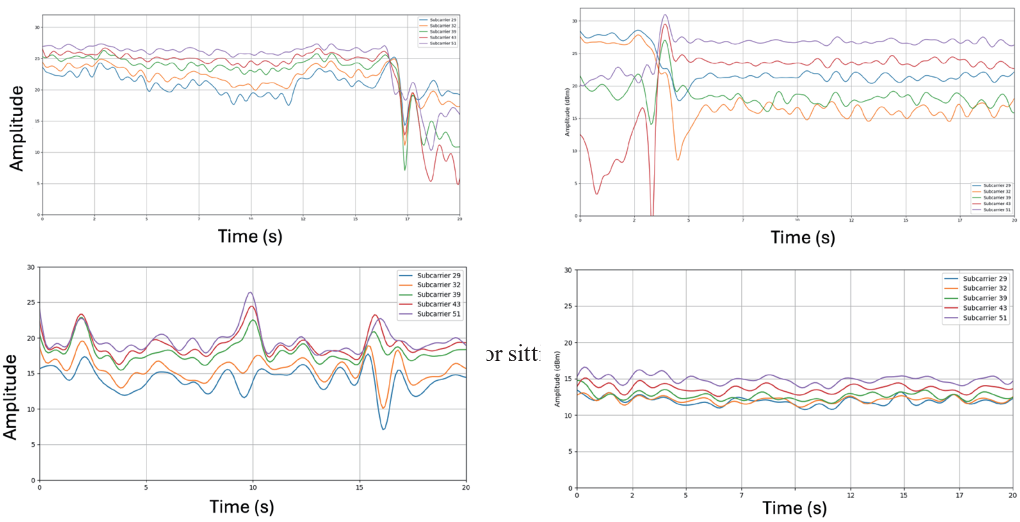

Fig. 5. (Color online) Filtered CSI. (cw) CSI for sitting, standing, empty room, and walking.

Table 1
Comparison of performance metrics between ensemble stacking and other models.

Metrics	kNN	SVM	Ensemble stacking	GRU	CNN
Accuracy	90.32%	91.40%	92.59%	92.50%	85.00%
Precision	90.53%	91.45%	92.68%	95.00%	90.83%
Recall	90.32%	91.40%	92.59%	92.50%	85.00%
F1-score	90.18%	91.38%	92.59%	92.00%	84.33%
Training time	0.0015 s	0.4239 s	3.2 s	130 s	8 s
Prediction time	0.1848 s	0.308 s	0.39 s	4 s	0.20 s

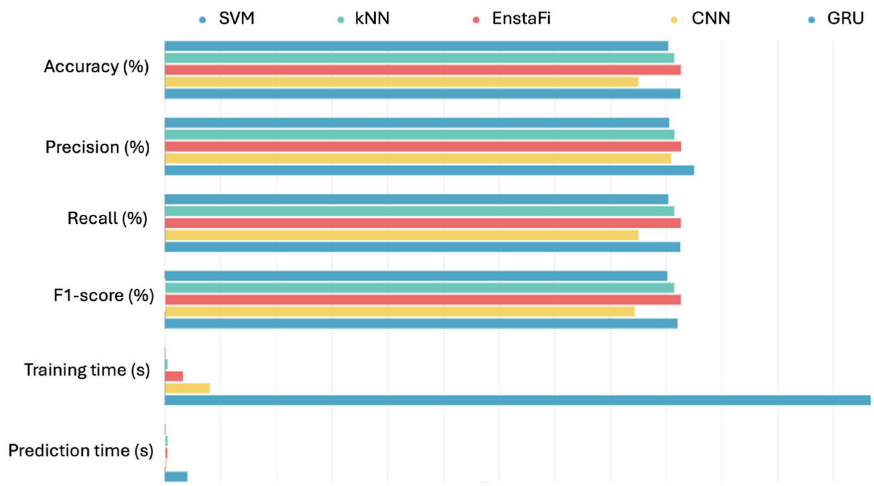

Fig. 6. (Color online) Comparison of performance metrics between EnSta-Fi and other models.

Table 2
Performance metric for each activity.

Metric vs activity	Accuracy	Precision	Recall	F1-score
Sitting	90%	83%	73%	77%
Standing	91%	74%	83%	78%
Walking	99%	99%	99%	88%
Empty room	99%	96%	99%	98%

92.59%, compared with SVM and kNN with 91.40 and 90.32%, respectively. The ensemble stacking model performs better than the individual SVM and kNN models. The model with ensemble stacking has higher accuracy, precision, recall, and F1-score than both models, reflecting its ability to combine the strengths of several models.

Compared with deep learning models, GRU and CNN require significantly longer training times. The GRU model takes 138 s to train, whereas the CNN model requires 8 s. This comparison shows that ensemble stacking achieves an accuracy similar to those of deep learning models, requiring much less training time. This is the result of the proposed model EnstaFi that can generally be trained in parallel (bagging) and by built-in early stopping so that the training time is much less than those of DL models. Thus, ensemble stacking may offer a more efficient solution for applications with limited time and computational resources than deep learning methods. Overall, our proposed model successfully compensates for the weaknesses of each model, making the ensemble method well suited for handling diverse data patterns and achieving balanced predictions that lead to improved performance. Our approach considers coarse-grained and relatively static activities: sitting, standing, and walking. CSI data show a relative change in walk activity, as seen in Fig. 5. We validate this by observing the performance metric for each activity, as shown in Table 2.

Table 2 shows that sitting and standing have similar patterns, making it difficult for the model to differentiate between these two activities; thus, their performance tends to be lower. On the other hand, walking showed the best performance apart from the empty room. Walking created more dynamic and clear movement patterns, providing signals more easily distinguished by the model, resulting in high classification accuracy. An empty room reflects a state of no activity, which is usually more manageable for the model to recognize because there is no variation (static) or no movement, thus having higher performance.

4. Conclusions

We presented the proposal for using a single-antenna setup for the HAR system leveraging its CSI amplitude by proposing EnSta-Fi, the ensemble stacking model, which incorporates high performance, short computational time, and low complexity. The two baseline ML models, kNN and SVM, are the backbone of the proposed model with the final classifier, i.e., logistic regression. We considered relatively static activities for Ensta-Fi evaluation, such as sitting, standing, and walking. The validation and position of our proposal are compared with those of the individual kNN and SVM and the DL-based GRU and CNN primarily for their accuracy performance and for their training and prediction time. Our EnSta-Fi yields higher accuracy results than kNN, SVM, and DL-based GRU and CNN. Moreover, our proposal has a shorter training time of only 3.2 s than GRU and CNN with training times of 180 and 8 s, respectively. From here, we can conclude that EnSta-Fi has comparable performance to the DL-based models and has a significantly short computational time. Thus, we can have a simple, flexible, and lightweight system by leveraging amplitude CSI from ESP32 in a single-antenna system setup with the ensemble stacking model for straightforward HAR system deployment.

Acknowledgments

This work is supported by Hibah Kompetensi Doktor UGM 2024. The authors would like to thank Dr. Nopphon Keeratirovanan and Professor Jun-ichi Takada of the Institute of Science Tokyo (previously Tokyo Institute of Technology) for the fruitful discussion and the visit to the Takada Lab for the research on HAR.

References

1 H. Raeis, M. Kazemi, and S. Shirmohammadi: IEEE Instrum. Meas. Mag. **24** (2021) 46.
2 Y. Zhang, G. Wang, H. Liu, W. Gong, and F. Gao: IEEE Internet Things J. **11** (2024) 29335. https://doi.org/10.1109/JIOT.2024.3397708
3 Z. Hussain, Q. Z. Sheng, and W.E. Zhang: J. Network Comput. Appl. **167** (2020) 1. https://doi.org/10.1016/j.jnca.2020.102738
4 H. Yan, Y. Zhang, Y. Wang, and K. Xu: IEEE Sens J. **20** (2020) 296. https://doi.org/10.1109/JSEN.2019.2938245
5 H. Taherdoost: Comput. Mater. Continua **81** (2024) 79. https://doi.org/10.32604/cmc.2024.054378
6 W. Zeng: IEEE MultiMedia **19** (2012) 4. https://doi.org/10.1109/MMUL.2012.24
7 J. A. Armenta-Garcia, F. F. Gonzalez-Navarro, and J. Caro-Gutierrez: Comput. Commun. **224** (2024) 254. https://doi.org/10.1016/j.comcom.2024.06.011
8 Y. Ma, G. Zhou, and S. Wang: ACM Comput. Surv. **52** (2019) 1. https://doi.org/10.1145/3310194
9 Z. Wang, K. Jiang, Y. Hou, W. Dou, C. Zhang, Z. Huang, and Y. Guo: IEEE Access **7** (2019) 155986. https://doi.org/10.1109/ACCESS.2019.2949123
10 Y. Chu, K. Cumanan, S. K. Sankarpandi, S. Smith, and O. A. Dobre: IEEE Access **11** (2023) 83763. https://doi.org/10.1109/ACCESS.2023.3300726
11 H. Gu, J. Yang, G. Gui, and H. Gacanin: IEEE Trans. Veh. Technol. **72** (2023) 16905. https://doi.org/10.1109/tvt.2023.3289631
12 J. Yang, X. Chen, H. Zou, C. X. Lu, D. Wang, S. Sun, and L. Xie: Patterns **4** (2023) 1. https://doi.org/10.1016/j.patter.2023.100703
13 S. M. Hernandez and E. Bulut: IEEE Commun. Surv. Tutorials **25** (2023) 46. https://doi.org/10.1109/COMST.2022.3209144
14 S. M. Hernandez and E. Bulut: 21st IEEE Int. Symp. World of Wireless, Mobile and Multimedia Networks, WoWMoM 2020. Institute of Electrical and Electronics Engineers Inc. (2020) pp. 277–286. https://doi.org/10.1109/WoWMoM49955.2020.00056
15 F. Demrozi, C. Turetta, A. Masrur, M. Schmidhammer, C. Gentner, S. Chakraborty, G. Pravadelli, and P. Kindt: IEEE Sens. J. **XX** (2023) 1. https://doi.org/10.36227/techrxiv.22284421.v1
16 K. Kim and J. Jeong: Proc. 26th Int. Conf. Circuits, Systems, Communications and Computers (CSCC, Institute of Electrical and Electronics Engineers Inc.) (2022) pp. 341–346. https://doi.org/10.1109/CSCC55931.2022.00066
17 D. J. Suroso and F. Y. M. Adiyatma: IEEE Access **12** (2024) 166381. https://doi.org/10.1109/ACCESS.2024.3493889
18 A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef: Decis. Anal. J. **7** (2023) 1. https://doi.org/10.1016/j.dajour.2023.100242
19 Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang: Int. Joint Conf. Pervasive and Ubiquitous Computing (2019) 1–26. https://doi.org/10.1145/3351279

About the Authors

Dwi Joko Suroso received his B. Eng. degree in engineering physics from Universitas Gadjah Mada (UGM), Indonesia, in 2010 and his M. Eng. and D. Eng. degrees from King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2012 and 2023, respectively. Since 2024, he has been an assistant professor at UGM. His research interests are in wireless sensing, indoor localization, robotics, and sensor development. (dwi.jokosuroso@ugm.ac.id)

Aisyah Attini Susanto received her B. Eng. degree in engineering physics from UGM in 2025. She is a research assistant in Wi-Fi sensing research with emphasis on machine learning model development.
(aisyah.a@mail.ugm.ac.id)

Liyas Ferdy Ray Tarigan received his B. Eng. degree in engineering physics from UGM in 2024. He is currently a hardware engineer at LG Electronics Indonesia. (liyas.f.r@mail.ugm.ac.id)

Nazrul Effendy is a professor at the Department of Nuclear Engineering and Engineering Physics at UGM. He is a member of the Indonesian Association of Pattern Recognition, the Indonesian Society for Soft Computing, and the Indonesian Artificial Intelligence Society. (nazrul@ugm.ac.id)