Jurnal Rekayasa Elektrika Vol. 21, No. 2, June 2025, page 104-114
ISSN. 1412-4785; e-ISSN. 2252-620X, Accredited by Kemendikbudristek No. 105/E/KPT/2022
DOI: 10.17529/jre.v21i2.44828

Hybrid Deep Learning Models Using LSTM with Random Forest for Radio Frequency-Based Human Activity Recognition in Line-of-Sight and Non-Line-of-Sight Environments

Niko Andriano and Feddy Setio Pribadi*
Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Semarang
E11 Building, Sekaran, Gunung Pati, Kota Semarang, Jawa Tengah 50229
*e-mail: feddy.setio@mail.unnes.ac.id

Abstract—**Human activity recognition (HAR) has become an important field of study because of its wide range of applications in healthcare, security, and smart living systems. Radio frequency (RF)-based HAR offers a non-invasive and privacy-preserving alternative to traditional vision-based systems. This study proposes a hybrid deep learning model combining long short-term memory (LSTM) networks with Random Forest classifiers for RF-based HAR, aiming to improve recognition accuracy across diverse environments. The model was evaluated using channel state information (CSI) and received signal strength indicator (RSSI) features under line-of-sight (LOS) and non-line-of-sight (NLOS) conditions. synthetic minority over-sampling technique (SMOTE) was integrated to balance the dataset and K-fold cross-validation was employed to assess robustness. The dataset included data from 8 subjects performing 10 different activities. The model achieved high classification accuracy, with 99.40% in Environment 1 (LOS), 97.58% in Environment 2 (LOS), and 98.30% in Environment 3 (NLOS), demonstrating the model's adaptability and effectiveness. The results highlight the potential of the hybrid LSTM with random forest approach for scalable and reliable RF-based HAR systems that can be integrated into real-world Internet of Things (IoT) applications.**

Keywords: ***human activity recognition, hybrid deep learning, long short-term memory, radio frequency-based, random forest***

I. Introduction

Human activity recognition (HAR) has become an important field of study because of its wide range of applications in healthcare, security, and smart living environments [1]-[3]. The capability to accurately monitor and classify human activities is pivotal for developing intelligent systems that enhance the quality of life. While traditional vision-based HAR systems have demonstrated significant success in controlled environments, their performance is often constrained in scenarios involving occlusions, low-light conditions, or privacy concerns [4]. Consequently, radio frequency (RF)-based HAR has gained prominence as a non-invasive alternative, employing technologies such as channel state information (CSI), it analyzes the attributes of the signal route between the receiver and transmitter, and received signal strength indicator (RSSI), which quantifies the strength of the received signal [5], [6]. Both technologies help achieve robust and privacy-preserving activity recognition in non-visual conditions. The performance of RF-based HAR systems is substantially influenced by environmental conditions, particularly the distinction between line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios. LOS environments facilitate direct signal propagation between the transmitter and receiver and typically yield stable, predictable measurements [7], [8]. Conversely, NLOS conditions introduce complexities such as signal reflections, diffractions, and attenuations, leading to degraded signal quality and reduced algorithm performance [8]. Addressing these challenges necessitates advanced algorithmic solutions that capture RF signal's temporal and spatial dynamics in diverse conditions.

Recent advancements in machine learning (ML) and deep learning (DL) have provided innovative approaches to mitigating challenges in HAR. ML algorithms, such as random forest, are widely recognized for their simplicity and robustness in classification tasks, while DL models, particularly long short-term memory (LSTM) networks, excel at modeling temporal dependencies and extracting meaningful patterns from complex datasets [9], [10]. This study proposes a hybrid deep learning model combining LSTM and random forest for RF-based HAR in both LOS) and NLOS environments. The LSTM network captures temporal features from CSI and RSSI data. At the same time, random forest serves as a robust ensemble classifier to enhance decision-making. The hybrid framework aims to improve recognition accuracy and robustness

Received 4 Nov. 2024; Revised 9 April 2025; Accepted 11 April 2025

across diverse conditions [8]. Additionally, the model is augmented by the synthetic minority over-sampling technique (SMOTE) for balancing class distribution and K-fold cross-validation (CV) to enhance model reliability. By integrating CSI and RSSI data, the model improves its capacity to identify subtle differences in activity signals, where CSI provides detailed insights into RF channel properties, and RSSI captures overall signal strength. This hybrid approach addresses key challenges and advances the development of reliable, scalable HAR systems suitable for integration into real-world Internet of Things (IoT) ecosystems.

II. Related Works

Recent advancements in HAR using RF-based techniques have significantly improved non-invasive and privacy-preserving activity monitoring. These improvements address key limitations of vision-based methods, such as susceptibility to occlusions and privacy concerns [11]. RF-based HAR systems utilize signals like CSI and RSSI to identify and classify activities, even in complex environments [12]. The combination of ML and DL has significantly improved the robustness and accuracy of these systems [13]. In recent years, HAR using RF signals has gained considerable attention, especially in device-free environments, which aim to detect human activities without requiring individuals to wear any sensors.

Early work in the field, such as Bokhari et al., employed deep gated recurrent units (DGRU) with RF transceivers, achieving an impressive 95-99% accuracy for walking, running, and sitting [14]. Their model faced limitations when applied to more complex environments, struggling with generalization. Similarly, Khan et al. utilized LSTM networks combined with convolutional neural networks (CNNs) using RF signals, yielding 95.3% accuracy in LSTM controlled environments [11]. Despite the high performance, the study highlighted challenges in accurately classifying certain walking activities, suggesting the need for further refinement in their model.

A notable trend in recent HAR research is the exploration of hybrid models that combine different signal types and machine learning algorithms. Mohtadifar et al. took this approach by combining RF and acoustic signals, achieving a high accuracy of over 98% in controlled settings [15]. However, their model performed less effectively in noisy environments, pointing to the challenges faced by acoustic-based HAR methods. Likewise, Strohmayer & Kampel employed EfficientNetV2 (CNN) to recognize activities in both LOS and NLOS environments, reporting a high accuracy of 92.0% in NLOS conditions [16]. Their model faced challenges with generalization across diverse activity classes and environments, indicating the difficulty of achieving consistent performance in real-world applications.

In addition to hybrid models, some researchers have focused on enhancing activity recognition by addressing the specific challenges of NLOS environments. Huang & Dai introduced a packet-based RF signal analysis method to recognize human activities in indoor NLOS environments, achieving over 76.7% accuracy [17]. While their method performed well in detecting basic activities, it struggled in highly congested environments, where signal fluctuations caused by minor velocity variations could not be reliably distinguished. Similarly, Liu et al. used one-class support vector machines (OSVM) for motion detection, achieving 90.89% accuracy for strenuous movements like kicking, but their model faced limitations in distinguishing subtle activities in NLOS conditions [5].

Recent advancements in HAR also focus on fine-grained recognition, which remains challenging. Moghaddam et al. proposed a feature-level fusion of CSI and RSSI signals to improve the recognition of fine-grained human activities, achieving an impressive 94.16% accuracy using random forest and excelling in human-to-human interaction recognition. This method showed excellent precision and recall but faced challenges distinguishing similar activities, highlighting the complexity of recognizing fine-grained activities in device-free environments [6]. Meanwhile, Chen et al. demonstrated the potential of combining ultra-wide band (UWB) radio and RF signals for activity recognition, achieving a precision of 96.9% and recall of 96.5%, but their model struggled with crowded RF channels and the limitations of RF-based sensing using commodity hardware [18]. Islam et al. utilized spatio-temporal convolution with nested LSTM (STC-NLSTMNet), which combines CNN and nested LSTM (NLSTM) architectures in a CSI-based setup, achieving 94.68% accuracy on the NLOS subset of the multi-environment dataset [10]. Despite the dataset's higher activity diversity and complex environment, their research performed and focused on the LOS subset, reflecting the challenges of signal propagation issues in NLOS settings.

III. Method

A. *Preliminaries Long Short-Term Memory (LSTM) with Random Forest*

The integration of LSTM and random forest is effective for HAR tasks, as LSTM networks capture temporal dependencies in sequential sensor data, while random forest classifiers offer robustness in managing noisy data, collectively improving the performance of HAR systems [19]. LSTM networks are specialized recurrent neural networks (RNNs) that effectively capture temporal patterns and long-range dependencies in sequential data [20]. Recent research has demonstrated their suitability for HAR, mainly when applied to sensor-based datasets [21]. As illustrated in Figure 1, each layer in the LSTM architecture contains LSTM cells that utilize gated mechanisms such as the forget gate, input gate, and output gate, which work together to regulate the flow of information across time steps.

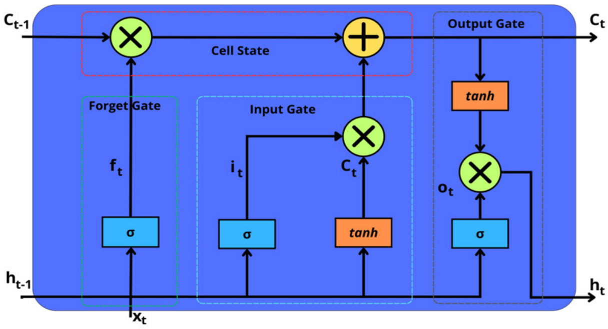

Figure 1. LSTM cell inner layer architecture

$$f_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right). \quad (1)$$

The forget gate (f_t) computed using the weight matrix (W_f) and bias term (B_f), determines which part of the information from the previous cell state ($C_{t\text{-}1}$) should be retained or discarded. It does this by applying a sigmoid activation function (σ) to a weighted combination of the previous hidden state ($h_{t\text{-}1}$) and the current input (x_t), as shown in (1). The result is a vector of values between 0 and 1, where 0 means "completely forget" and 1 means "completely retain".

$$i_t = \sigma\left(W_i \times [h_{t-1}, x_t] + b_i\right), \quad (2)$$

$$\tilde{C}_t = \text{üüüü}\left(W_c \times [h_{t-1} \; x_t] + b_c\right)$$

By performing two steps in (2) and (3), the input gate controls the addition of new information to the current cell state:

a. A sigmoid function generates a filter (i_t) using weight matrix W_i and b_i bias to decide which values to update.
b. A hyperbolic tangent (tanh(.)) function creates a vector of candidate values ($\tilde{c}_t$) using weight matrix W_c and bias (b_c) to be added to the cell state. These two outputs are multiplied and added to the scaled previous cell state to update (C_t).

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t. \quad (4)$$

The new cell state (C_t) is computed by combining the forget gate's scaled old cell state ($f_t \times C_{t\text{-}1}$) and the input gate's filtered candidate values ($i_t \times \tilde{c}_t$), as shown in (4). This process allows the LSTM to retain important information and discard irrelevant details.

$$o_t = \sigma\left(W_o \times [h_{t-1}, x_t] + b_o\right),$$
$$h_t = o_t \times \tanh(C_t) \quad (5)$$

The output gate (o_t) calculated by applying the sigmoid activation function to the weighted combination of the previous hidden state ($h_{t\text{-}1}$) and the current input (x_t), using the weight matrix (W_o) and bias (b_o), as shown in (5). This gate determines which parts of the updated cell state (C_t), after passing through the activation should be included in the output gate. The result is the hidden state (h_t)), which serves as the LSTM's output and is forwarded to the next step.

Random forest is a robust ensemble learning technique widely applied for HAR tasks because it can handle noisy datasets and classify activities with high accuracy [22]. random forest employs bagging, where several decision trees are trained independently on random subsets of the training data, promoting diversity and reducing overfitting. At each split during tree construction, features are selected from a random subset, ensuring that each tree uncovers different patterns in the data, thereby enhancing the ensemble's robustness. As shown in (6), to select splits, the algorithm minimizes Gini (T) impurity (p_i) for the proportion of samples of class (i) in the node:

$$T = 1 - \sum_{i=1}^{n} p_i^2. \quad (6)$$

After training all trees, the final prediction is made by taking the majority vote (for classification) or the average (for regression) of the outputs from all trees. This ensemble strategy reduces variance and increases prediction accuracy. Through majority voting, the final prediction is determined for prediction from the n-th tree, as follow,

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \ldots\ldots, T_n(x)\}. \quad (7)$$

The hybrid use of LSTM and random forest leverages their respective strengths, creating a system that excels in feature extraction and robust classification, as demonstrated in multiple studies [23].

B. Proposed Research Methodology

The proposed methodology develops a hybrid deep learning model for HAR using RF signals. This model leverages the strengths of LSTM networks for feature extraction and random Forest for classification. Figure 2 illustrates the workflow of the proposed hybrid RF-based HAR model, following a structured sequence to ensure accurate human activity recognition. The process begins with data acquisition, where RF signals are collected. The collected data then undergoes preprocessing to remove noise and errors, followed by data balancing using SMOTE to address class imbalances and improve model learning. Next, the hybrid deep learning model processes the refined data by combining LSTM networks for capturing temporal dependencies and a random forest classifier for robust decision-making. The model undergoes K-Fold Cross Validation to validate performance, ensuring reliability across multiple training and testing sets. Finally, the system integrates a HAR monitoring system, which processes classified data to detect and monitor human activities in real-time, enhancing applications in security, healthcare, and smart environments.

Figure 3 details the LSTM and random forest model layer architectures. These figures demonstrate the multiple layers of input, LSTM, dropout, dense, and output used in LSTM, followed by the random forest classifier layers to construct the hybrid HAR model, enabling precise and scalable activity recognition in RF-based systems. The

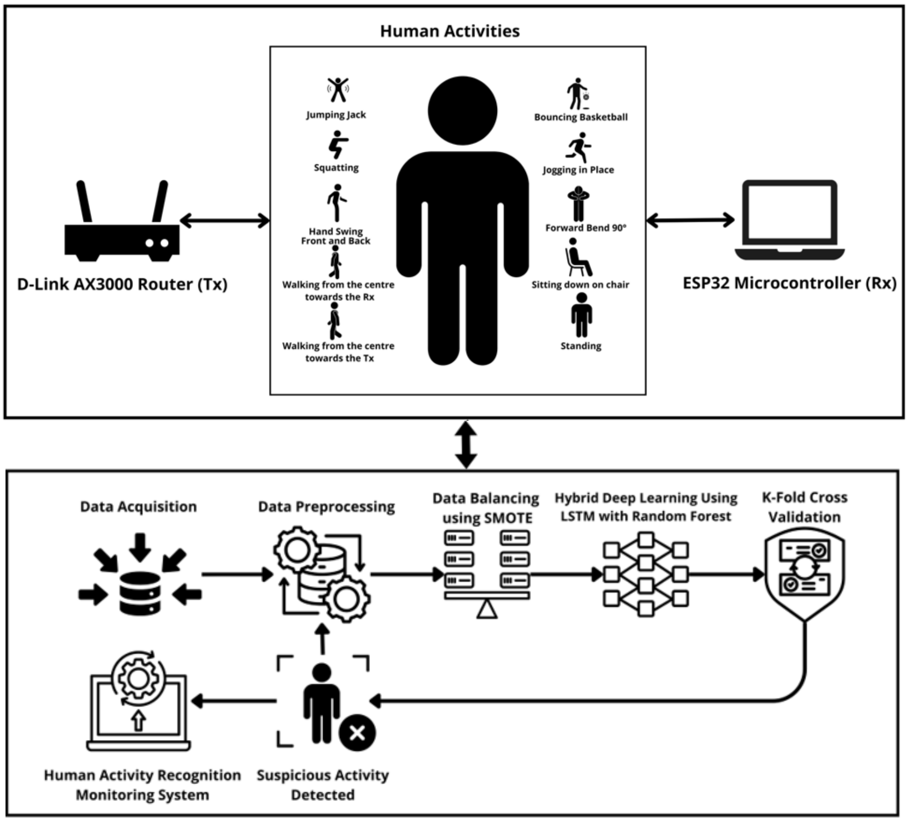

Figure 2. Workflow of the hybrid RF-based HAR model

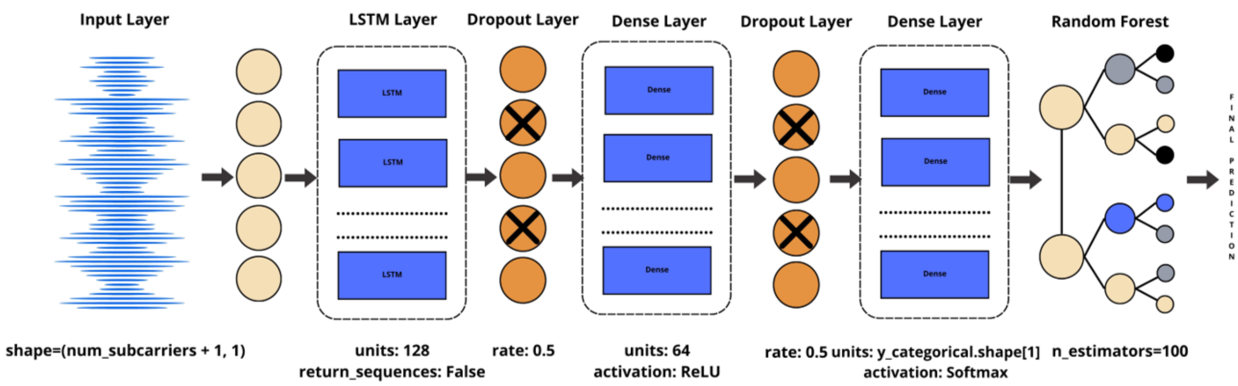

Figure 3. LSTM with random forest model layer architecture

architecture begins with acquiring and preprocessing RF signals to prepare data for temporal feature extraction. The LSTM network captures sequential dependencies in the CSI and RSSI features, followed by Dropout layers to mitigate overfitting. Finally, the random forest classifier takes the extracted features for robust and accurate activity classification.

The proposed framework integrates LSTM and random forest for HAR using RF signals. This hybrid deep learning model is designed to capture temporal dependencies through LSTM and enhance classification robustness using random forest. The framework is for LOS and NLOS environments, leveraging CSI and RSSI as input features. The framework is implemented in the following operational steps:

1. Data Preprocessing

Data preprocessing involves loading and combining datasets from various activities stored in structured directories. CSI features are extracted and combined with RSSI values to create a comprehensive feature representation. The features are standardized to ensure uniformity and enable practical model training. Activity labels are encoded into the numerical format and converted into categorical labels to ensure compatibility with the LSTM network. Additionally, the SMOTE is applied to balance the dataset.

2. Feature Extraction Using LSTM and random forest Classification

The LSTM-based model used a 128-unit LSTM layer to extract temporal features, followed by two dropout layers with a 50% dropout rate to mitigate overfitting. It includes a 64-unit dense layer and an output layer with SoftMax activation for multi-class classification. The model is compiled using the categorical cross-entropy loss function and the Adam optimizer. The LSTM model is trained for 50 epochs on the pre-processed dataset, using 80% of the data for training and the remaining 20% for testing. Finally, training and validation accuracy and loss are visualized across epochs to monitor the model's performance.

The trained LSTM model extracts temporal features from RF signal data for training and testing sets. A random forest classifier is then trained with 100 decision trees. Finally, the trained random forest model predicts activities on the test set.

3. Performance Evaluation

The performance evaluation of the hybrid deep learning model for RF-based HAR carried out methodically to ensure a thorough understanding of its capabilities. The model's effectiveness was assessed using standard metrics, including precision, recall, F1-score, and accuracy. These metrics offer valuable insights into the model's ability to classify human activities across various scenarios.

Precision indicates the proportion of correctly identified instances (true positives) among all the cases classified as positive, as shown in (8). A high precision score suggests that the model accurately identifies activities with minimal misclassification. It reflects the model's ability to avoid false positives and is calculated as follows:

$$\text{Precision (\%)} = \frac{TP}{TP+FP} \times 100. \tag{8}$$

Recall, also known as sensitivity, measures the proportion of true positives correctly identified among all actual positives, as shown in (9). A high recall value indicates that the model effectively detects all occurrences of the target activity. It assesses the model's ability to capture all relevant instances and is computed as:

$$\text{Recall (\%)} = \frac{TP}{TP+FN} \times 100. \tag{9}$$

The F1-score is the harmonic mean of precision and recall, as shown in (10), providing a balanced measure of the model's accuracy. A high F1-score reflects the model's robustness in balancing precision and recall. It is beneficial when there is an uneven class distribution:

$$\text{F1 - score (\%)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100. \tag{10}$$

Accuracy represents the proportion of correctly classified instances out of the total cases in the dataset, as shown in (11). High accuracy suggests that the model performs well across all activities and conditions. It provides a general measure of the model's overall performance and is expressed as:

$$\text{Accuracy (\%)} = \frac{TP+TN}{TP+TN+FP+FN} \times 100. \tag{11}$$

A confusion matrix was used to thoroughly analyze the model's performance for each activity. It provides a tabular representation of predicted versus actual classifications, which includes four key components: true positives (TP), where positive cases are correctly predicted; true negatives (TN), where negative cases are correctly predicted; false positives (FP), where negative cases are incorrectly predicted as positive; and false negatives (FN), where positive cases are incorrectly predicted as negative.

4. K-Fold Cross-Validation (CV)

Stratified 5-fold cross-validation is applied to evaluate the model's robustness. As shown in (12), the mean cross-validation accuracy is calculated by averaging the accuracy scores obtained from each fold in the K-fold cross-validation process, where represents the total number of folds:

$$\text{Mean CV (\%)} = \frac{1}{K}\sum_{i=1}^{K} \text{Accuracy}_i \times 100. \tag{12}$$

In this process, the dataset is divided into equal parts, and for each fold, the model is trained on part of the data while the remaining part is used as the validation set. The accuracy for each fold is computed, and then these individual accuracy scores are averaged to provide the mean cross-validation accuracy, giving a more robust estimate of model performance across different subsets of the data. The proposed advanced RF-based HAR model, incorporating all the procedures tailored explicitly for RF-based applications, is developed using the detailed method presented in Algorithm 1.

C. Experimental Setting

1. Experimental Setup

The experiments were conducted in a controlled setting using an NVIDIA Tesla T4 GPU, with CUDA version 12.2 and NVIDIA-SMI version 535.104.05 managing the system. The operating system was Ubuntu 22.04.3 LTS (Jammy Jellyfish), known for its compatibility and performance for computational tasks. The hybrid deep learning model was implemented using Python, leveraging TensorFlow for the LSTM network and Scikit-learn for the random forest classifier. All experiments were executed on a dedicated machine to ensure optimal utilization of computational resources, with no interference from other processes running on the GPU. The data preprocessing, model training, and evaluation were carried out under these conditions to guarantee the reproducibility and consistency of the results.

2. Dataset Description

This study used a recent data set that is publicly available radio frequency-based human activity [24]. Dataset collected using ESP32 microcontrollers and

Algorithm 1. RF-based HAR using LSTM with random forest

Require: The dataset D contains features $X \in R^{m.n}$ and labels $Y \in Z^{m}$, where *n* represents the number of features per sample and *m* indicates the number of samples.
Ensure: Trained LSTM model and random forest classifier.

1 Import and prepare data
2 Import dataset D and extract CSI and RSSI features.
3 *features = ['rssi'] + csi_columns*
4 Standardize features using *StandardScaler*.
5 Encode activity labels into the numerical format and convert to a categorical format.
6 *y = label_encoder.fit_transform(dataset['activity'])*
7 Prepare data for LSTM:
8 Reshape features into $X_{reshape} \in R^{m.n.1}$ for LSTM input.
9 Split the dataset into training and testing sets:
10 $X_{train}, X_{test}, y_{train}, y_{test}$ = *train_test_split* ($X_{lstm}, y_{categorical}$, *test_size*=0.2,*random_state*=42)
11 Apply SMOTE to balance the dataset by oversampling the minority class:
12 $X_{train\ smote}, y_{train\ smote}$ = *smote_fit.resample* ($X_{train\_reshape}$=($X_{train\_reshape}$ [0], – 1), y_{train}))
13 X_{train_smote}= $X_{train\_smote.reshape}$=($X_{train\_smote.shape}$ [0], *num_subcarriers* + 1, 1)
14 Train LSTM model:
15 Train the model using the training set and validate on the testing set:
16 *history= lstm_model.fit*($X_{train\_smote}, y_{train\_smote}$, *epochs=50, batch_size=64,validation_data=*(X_{test}, y_{test}),*verbose=1)*
17 Plot training and validation accuracy and loss metrics.
18 Extract features using LSTM:
19 Use the trained LSTM model to extract temporal features from X_{train_smote} and X_{test}.
20 Train random forest classifier:
21 Train a random forest classifier using LSTM-extracted features from the training set.
22 Optimize parameters: *n_estimators*=100.
23 K-fold cross-validation:
34 Apply Stratified 5-Fold Cross-Validation to evaluate the model's robustness.
25 *kf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)*
26 Compute the mean accuracy across all folds to assess the generalization capability.
27 Evaluate model:
28 Predict activities on the test set using the random forest model.
29 Generate a classification report with precision, recall, F1-score, and accuracy metrics.
30 Visualize the confusion matrix to analyze misclassifications.

D-Link AX3000 routers. It contains RF data captured in indoor environments under both LOS and NLOS scenarios. The dataset includes data from 8 subjects performing 10 activities, each repeated in 20 trials across three experimental setups. This results in 1,600 trials per setup, providing a comprehensive dataset for human activity recognition. Key features of the dataset include:

Table 1. Sample count activity per environment

Activity	Environment 1	Environment 2	Environment 3
Jumping jack	23,716	23,840	23,687
Squatting	23,734	23,687	24,151
Hand swing	23,714	23,879	23,981
Walk from center to Rx	23,734	24,048	23,633
Walk from center to Tx	23,850	23,835	24,650
Bouncing ball	22,896	23,554	23,321
Jogging in place	23,367	23,883	23,696
Forward bend	23,396	23,907	23,628
Sitting down	24,379	23,618	23,703
Standing	23,740	23,785	23,608

a. Activities

The ten recorded activities include jumping jacks, squatting, swinging hands front and back, walking towards the receiver (Rx), walking towards the transmitter (Tx), bouncing a basketball, jogging in place, bending forward at a 90° angle, sitting down on a chair, and standing.

b. Environment Scenarios

- Environment 1 (LOS Scenario): The Tx and Rx were placed 7 m apart, with volunteers performing activities at the midpoint.
- Environment 2 (LOS Scenario): The Tx and Rx were 5.2 m apart, with the same activity midpoint setup.
- Environment 3 (NLOS Scenario): A wooden wall was a barrier between the Tx and Rx, simulating real-world signal obstructions. The Tx was placed in a common area, while the Rx was in an inner room.

c. Data Features

Each trial provides one RSSI value and 52 CSI subcarriers captured using the ESP-CSI Tool. This configuration allows for a detailed analysis of RF signal characteristics during activities.

d. Hardware and Configuration

The ESP32 microcontroller (Rx) operated at a baud rate of 115,200 and a CPU frequency of 240 MHz. The D-Link AX3000 router (Tx) transmitted RF signals using the IEEE 802.11n (Wi-Fi 4) standard with a 40 MHz channel bandwidth. Before applying the SMOTE, the dataset exhibited an imbalanced class distribution across different activities, with certain activities having significantly fewer samples than others. The sample counts for the 10 activity classes represent the data collected from 8 subjects within a single environment after preprocessing, including noise reduction and data filtering. Table 1 presents the sample count activity per environment, illustrating the initial imbalance in the dataset.

SMOTE was applied to the training data to address this issue and prevent model bias towards majority classes, ensuring a more balanced distribution. Figure 4 illustrates the number of samples before and after applying SMOTE,

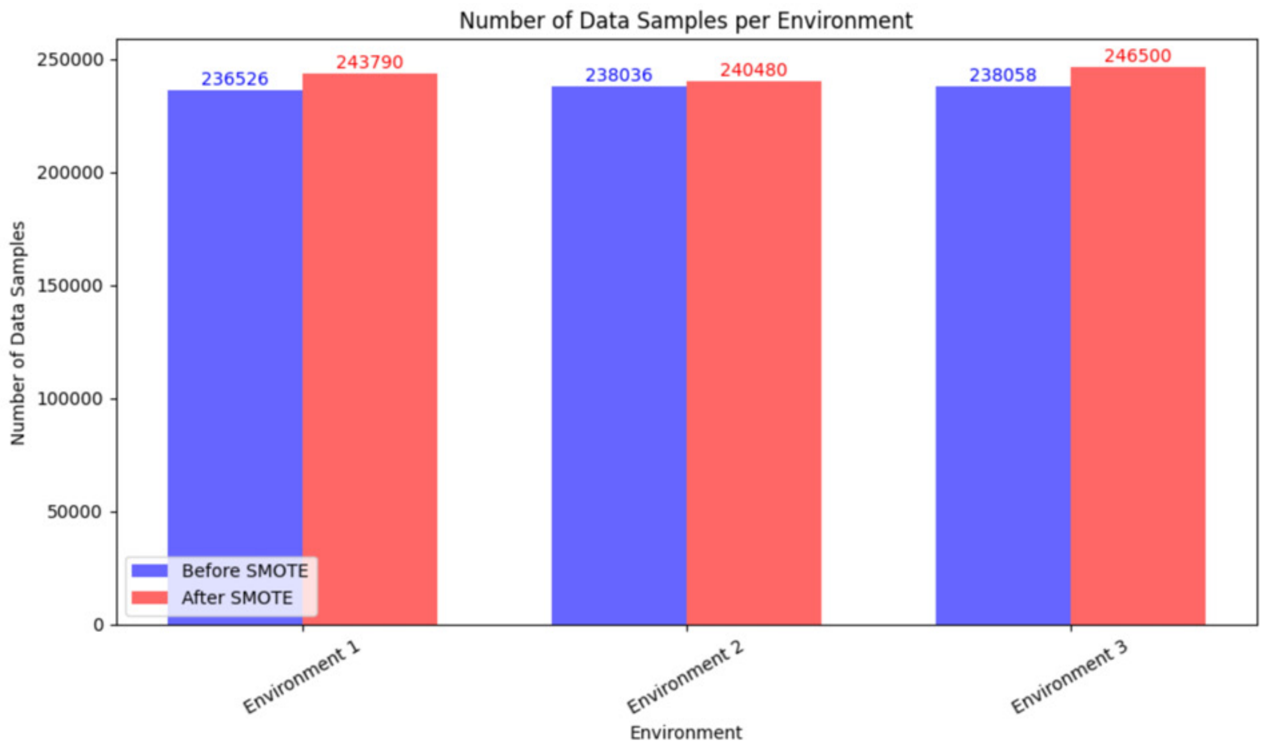

Figure 4. Data sample per environment before and after SMOTE

highlighting the initial imbalance and subsequent equalization, where all activity classes in the training set were adjusted to have equal representation. The post-SMOTE dataset was standardized based on the class with the highest number of samples, ensuring a uniform distribution across all activity classes. In the figure, the blue bars represent the sample count before SMOTE, while the red bars indicate the sample count after SMOTE, providing a clear visualization of the balancing effect.

IV. Experimental and Result Analysis

A. Hybrid Deep Learning Models Using LSTM with Random Forest

The proposed hybrid deep learning, combining LSTM with random forest, is designed to effectively classify various human activities across different environments in the context of RF-based HAR. The model was tested using datasets from three experimental environments, each presenting unique challenges in signal propagation and noise. RF datasets underwent preprocessing steps to optimize computational efficiency and improve feature representation, including data dimensionality reduction and normalization. The processed data was then input into the LSTM for temporal sequence modeling, taking advantage of its ability to capture complex temporal dependencies in RF signals. Finally, the data was classified using the random forest, known for its strength in decision-based ensemble learning. Figures 5-10 illustrate the training and validation accuracy and loss graphs, highlighting the model's learning progression and convergence over time.

Table 2. Performance metrics per environment

Activity	Environment 1	Environment 2	Environment 3
Precision	99.40%	97.52%	98.30%
Recall	99.40%	97.57%	98.29%
F1-score	99.40%	97.53%	98.27%
Accuracy	99.40%	97.58%	98.30%
Mean CV accuracy	99.51%	98.03%	98.72%

B. Model Performance Evaluation

Table 2 below presents the classification performance of the hybrid Deep Learning model across three experimental environments, each presenting distinct challenges. The model's performance is assessed using key metrics, including precision, recall, F1-score, overall accuracy, and mean cross-validation accuracy. These metrics offer valuable insights into the model's effectiveness in classifying human activities under various conditions.

The model's performance was evaluated under three different environments, each with varying conditions influencing signal propagation and classification accuracy. These scenarios include:

- Environment 1 (LOS Scenario)

In this LOS scenario, the Tx and Rx were placed 7 m apart, with volunteers performing activities at the midpoint. This setup allowed for direct signal propagation, resulting in high classification accuracy and strong overall performance, with an accuracy of 99.40% and a mean

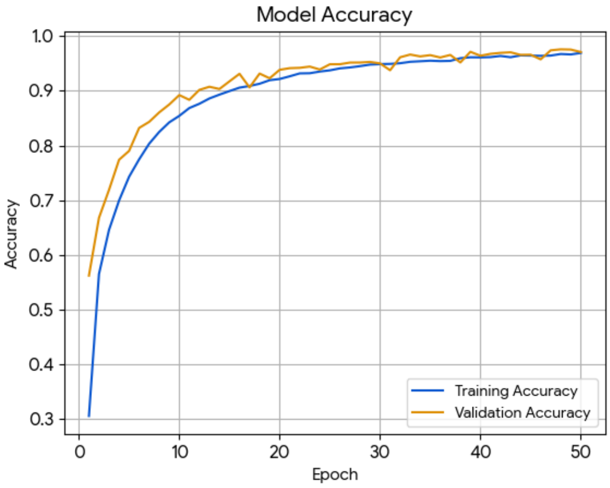

Figure 5. Accuracy graph environment 1

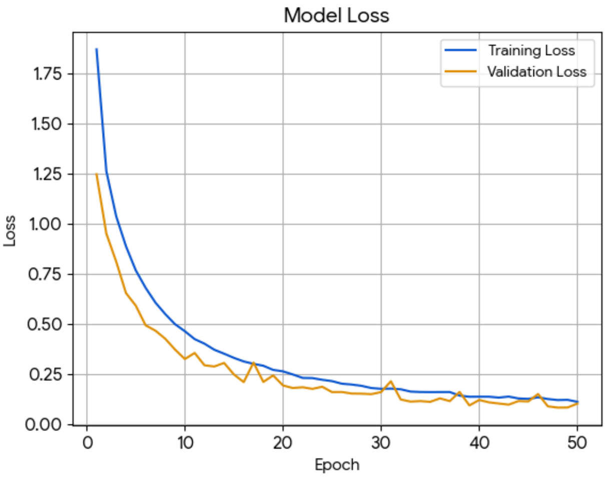

Figure 6. Loss graph environment 1

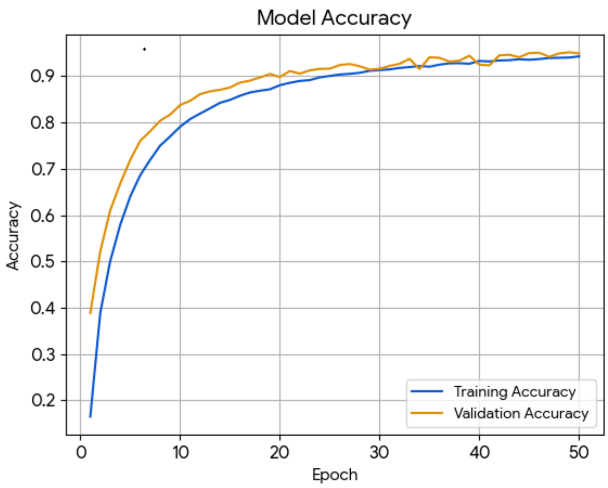

Figure 7. Accuracy graph environment 2

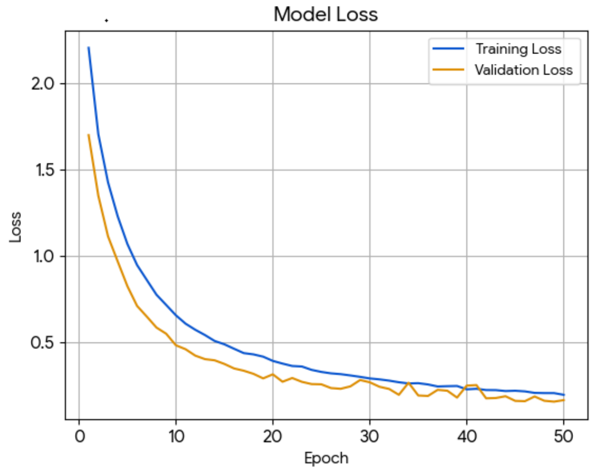

Figure 8. Loss graph environment 2

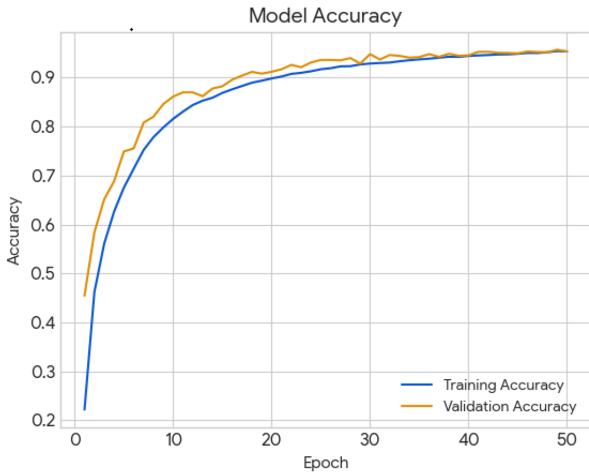

Figure 9. Accuracy graph environment 3

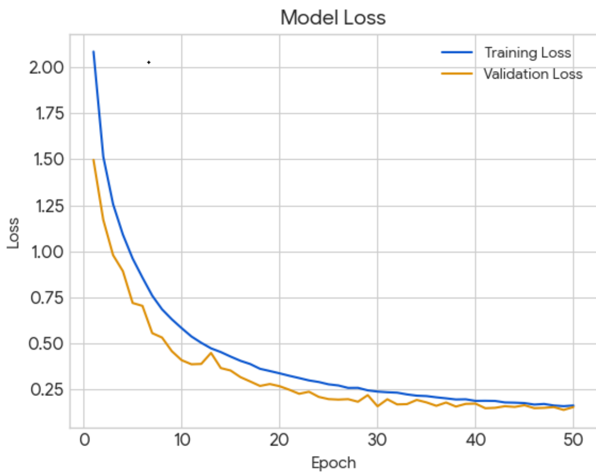

Figure 10. Loss graph environment 3

cross-validation accuracy of 99.51%.

- Environment 2 (LOS Scenario)

Similar to Environment 1, this LOS scenario also featured a Line-of-Sight condition, but with the Tx and Rx positioned 5.2 m apart. Despite the shorter distance, the model achieved an accuracy of 97.58%, with a mean cross-validation accuracy of 98.03%. The decrease in performance is likely due to the reduced distance between the Tx and Rx, which can lead to slight signal variations.

- Environment 3 (NLOS Scenario)

In the NLOS scenario, the Tx and Rx were separated by a wooden wall, simulating real-world signal obstructions. This setup significantly impacted signal propagation and resulted in a slightly lower accuracy of 98.30%. Despite

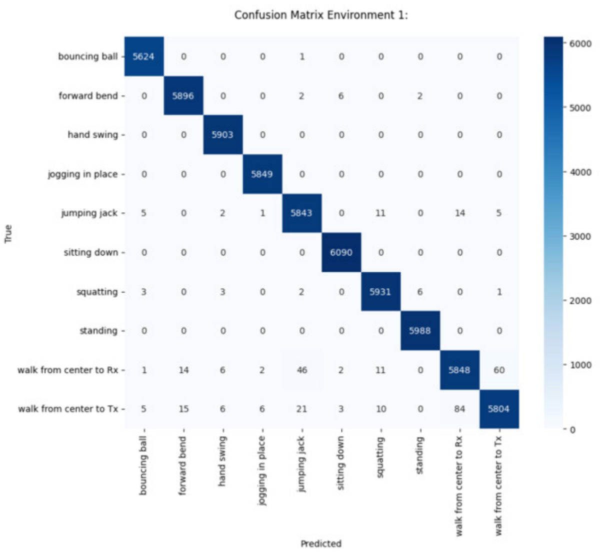

Figure 11. Confusion matrix environment 1

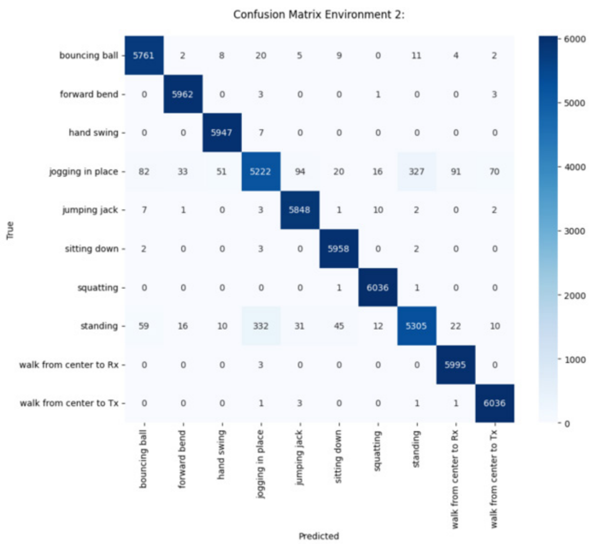

Figure 12. Confusion matrix environment 2

this challenge, the model maintained robust performance, with a mean cross-validation accuracy of 98.72%, demonstrating its adaptability to more complex conditions.

The confusion matrix was used to evaluate the classification performance of the hybrid deep learning model. It provides a detailed view of each activity's true positives, false positives, true negatives, and false negatives. The confusion matrix in Figures 11-13 shows the model's performance in classifying human activities across three different environments. The model accurately identifies activities such as bouncing ball, forward bend, hand swing, jumping jack, sitting down, and squatting, with high values along the diagonal in each confusion matrix.

However, some classification errors occur more frequently with activities and similar movement characteristics, such as jogging in place, which is often misclassified as standing in Environment 2 and Environment 3. This indicates that the model struggles to distinguish between these similar movements. In contrast,

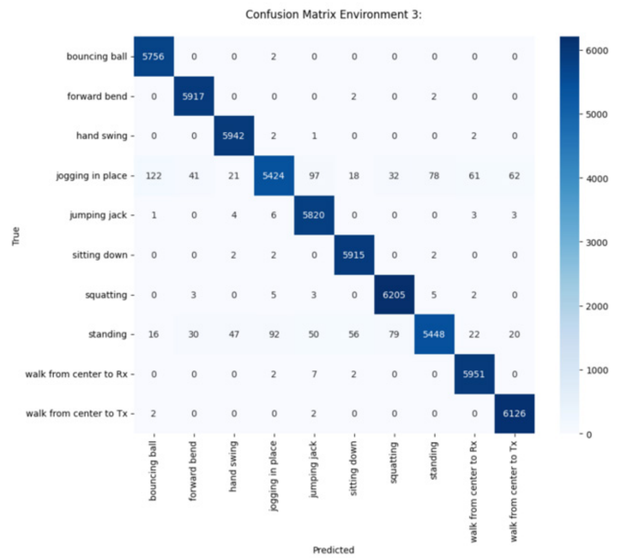

Figure 13. Confusion matrix environment 3

the activities that walk from center to Rx and walk from center to Tx are identified very accurately in Environment 2 and Environment 3 with minimum classification errors, suggesting that the model finds these activities easier to distinguish than other activities in Environment 1. Nevertheless, there are performance differences between the environments. Environment 1 performs better with fewer classification errors, while Environment 2 and Environment 3 have more errors.

C. The Variety of Approaches Developed to Address Challenges in HAR

Table 3 presents an overview of existing studies in HAR, as discussed in the Related Works section. It highlights essential elements of each study, such as the algorithms used, types of datasets, number of recognized activities, accurate metrics, and specific research objectives. By outlining these aspects, the table illustrates the methodological progression in the field and showcases the variety of approaches developed to address challenges in HAR. This overview provides a broader context for identifying research gaps and distinguishing each study's focus or unique contributions. It shows how the proposed method addresses key issues related to algorithmic, activity diversity, environment, and model adaptability in real-world scenarios.

The results of this study highlight the strength of the proposed hybrid model, which integrates LSTM networks with a random forest classifier for robust RF-based HAR. Unlike previous works, this model leverages a combination of CSI and RSSI features, enabling it to capture both temporal and spatial signal characteristics effectively. Tested across multiple environments, including LOS and NLOS conditions, the model consistently achieves high accuracy, with performance metrics reaching up to 99.40% in LOS and 98.30% in NLOS settings. This demonstrates the model's adaptability in complex environments and suitability for real-time human activity monitoring.

Table 3. The variety of approaches developed to address challenges in HAR

Reference	Employed Algorithm	Dataset	Activities	Accuracy	Focus Research
Bokhari et al. [14]	Deep gated recurrent unit (DGRU) model	Wi-Fi transceiver based on SDR (USRP)	7	95-99%	CSI-based HAR with passive monitoring for privacy-preserving recognition
Khan et al. [11]	LSTM, CNN, LSTM-CNN	Wi-Fi CSI (USRP devices)	7	LSTM: 95.3% CNN: 86.5% LSTM-CNN: 91.0%,	Wi-Fi CSI-based contactless HAR using deep learning models to recognize multiple activities
Mohtadifar et al. [15]	MLP, SVM, random forest, ERT, KNN, GTB	Radio frequency (VNA) + acoustic (microphone array)	4	Over 98%	Hybrid method combining RF and acoustic signals for HAR, improving recognition accuracy over single sensor systems
Strohmayer & Kampel [16]	EfficientNetV2 (CNN)	Wallhack1.8k dataset	3	LOS: 89.4–90.2% NLOS: 86.8–92.0%	Long-range through-wall Wi-Fi CSI-based HAR using low-cost ESP32-S3 with directional antennas
Huang & Dai [17]	KNN, SVM (SMO)	PRR and packet state series (USRP nodes)	4 classifications	Over 76.7%	Device-free human activity recognition using RF signal in indoor NLOS environments
Liu et al. [5]	One-class support vector machine (OSVM)	CSI data from Intel 5300 NIC and OpenWRT routers	3 categories	90.89%	CSI phase-based detection of strenuous motion under complex LOS/NLOS scenarios
Moghaddam et al. [6]	SVM, GNB, DT, LR, LDA, KNN, random forest	Wi-Fi-based human-to-human interaction dataset	7	random forest: 94.16%	Feature-level fusion of CSI & RSSI signals for fine-grained recognition of multi-subject human interactions using ML classifiers
Chen et al. [18]	HAR-SAnet: Signal adapted CNN	UWB radio module	7	Precision: 96.9% Recall: 96.5%	Real-time HAR using UWB and lightweight dual-domain CNN optimized on edge devices
Islam et al. [10]	STC-NLSTMNet (CNN + NLSTM)	Multi-environment dataset StanWiFi dataset	6	LOS: 98.20%, NLOS: 94.68% (Multi-env) 99.88% (StanWiFi)	Joint spatial and temporal feature extraction for HAR using CSI in LOS environment
This study	LSTM + random forest	RF-based human activity dataset	8	99.40% Environment 1, 97.58% Environment 2, 98.30% Environment 3	Hybrid Deep Learning model RF-based HAR using CSI and RSSI across LOS and NLOS environments for real-time monitoring system

Its integration of deep learning and ensemble learning components gives it a clear edge over prior approaches, positioning it as a promising solution for real-world IoT applications.

V. Conclusion

This study introduced a hybrid deep learning model combining LSTM networks with random forest classifiers for RF-based HAR. The model was evaluated in three experimental environments: two LOS scenarios and one NLOS scenario, demonstrating high classification accuracy, achieving 99.40% in Environment 1 (LOS), 97.58% in Environment 2 (LOS), and 98.30% in Environment 3 (NLOS). To address the class imbalance, the SMOTE was applied, enhancing model performance. Additionally, k-fold cross-validation was used to assess the model's robustness, yielding an average cross-validation accuracy of 99.51% in Environment 1, 98.03% in Environment 2, and 98.72% in Environment 3, confirming the model's consistency across different data splits.

The results demonstrate the potential of the hybrid LSTM with random forest model for reliable and scalable RF-based HAR systems, with implications for healthcare, security, and smart living systems applications. Additionally, further exploration of advanced signal processing techniques and integrating more complex architectures could help mitigate environmental variability and challenges in dynamic conditions, thereby improving system performance in diverse and unpredictable settings.

Acknowledgments

Thank you to Dr. Feddy Setio Pribadi, S.Pd., M.T., for the invaluable guidance, support, and insights provided during the preparation of this article. His contributions were instrumental in achieving the objectives of this research.

References

[1] L. Bibbò and M. M. B. R. Vellasco, "Human activity recognition (HAR) in healthcare," *Applied Sciences*, vol. 13, no. 24, p. 13009, Dec 2023.

[2] S. Sakka, V. Liagkou, and C. Stylios, "Exploiting security issues in human activity recognition systems (HARSs)," *Information (Switzerland)*, vol. 14, no. 6, p. 315, May 2023.

[3] G. Diraco, G. Rescio, A. Caroppo, A. Manni, and A. Leone, "Human action recognition in smart living services and applications: Context awareness, data availability, personalization, and privacy," *Sensors*, vol. 23, no. 13, p. 6040 Jun 2023.

[4] M. Karim, S. Khalid, A. Aleryani, J. Khan, I. Ullah, and Z. Ali, "Human action recognition systems: A review of the trends and state-of-the-art," *IEEE Access*, vol. 12, pp. 36372-36390, Mar 2024.

[5] J. Liu, L. Wang, L. Guo, J. Fang, B. Lu, and W. Zhou, "A research on CSI-based human motion detection in complex scenarios," In Proc. 19th International Conference on e-Health Networking, Applications and Services, Healthcom, IEEE, Dec. 2017, pp. 1-6.

[6] M. G. Moghaddam, A. A. N. Shirehjini, and S. Shirmohammadi, "A WiFi-based method for recognizing fine-grained multiple-subject human activities," *IEEE Trans Instrum Meas*, vol. 72, no. 2520313, pp. 1–13, Jun 2023.

[7] M. Waqas *et al.*, "Advanced line-of-sight (LOS) model for communicating devices in modern indoor environment," *PLOS ONE,* vol. 19, no. 7, p. e0305039, Jul 2024.

[8] J. S. Choi, W. H. Lee, J. H. Lee, J. H. Lee, and S. C. Kim, "deep learning based NLOS identification with commodity WLAN devices," *IEEE Trans Veh Technol*, vol. 67, no. 4, pp. 3295-3303, Apr 2018.

[9] C. Huang *et al.*, "Machine learning-enabled LOS/NLOS identification for MIMO systems in dynamic environments," *IEEE Trans Wirel Commun*, vol. 19, no. 6, pp. 3643-3657, Jun 2020.

[10] M. S. Islam, M. K. A. Jannat, M. N. Hossain, W. S. Kim, S. W. Lee, and S. H. Yang, "STC-NLSTMNet: An improved human activity recognition method using convolutional neural network with NLSTM from WiFi CSI," *Sensors*, vol. 23, no. 1, p. 356, Dec. 2022.

[11] M. Z. Khan *et al.*, "Contactless human activity recognition using deep learning with flexible and scalable software define radio," In Proc. International Wireless Communications and Mobile Computing, IWCMC, Jun 2023, pp. 126-131.

[12] 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*, vol. 5, pp. 3595-3623, Jun 2024.

[13] M. M. U. Khan, A. Bin Shams, and M. S. Raihan, "A prospective approach for human-to-human interaction recognition from Wi-Fi channel data using attention bidirectional gated recurrent neural network with GUI application implementation," *Multimed Tools Appl*, vol. 83, no.22, pp. 62379-62422, Jul 2024.

[14] S. M. Bokhari, S. Sohaib, A. R. Khan, M. Shafi, and A. ur R. Khan, "DGRU based human activity recognition using channel state information," *Measurement*, vol. 167, no. 2021, p. 108245, Jan 2021.

[15] M. Mohtadifar, M. Cheffena, and A. Pourafzal, "Acoustic-and radio-frequency-based human activity recognition," *Sensors*, vol. 22, no. 9, p. 3125, Apr 2022.

[16] J. Strohmayer and M. Kampel, "Directional antenna systems for long-range through-wall human activity recognition," In Proc. IEEE International Conference on Image Processing (ICIP), Jan 2024, pp. 3594-3599.

[17] X. Huang and M. Dai, "Indoor device-free activity recognition based on radio signal," *IEEE Trans Veh Technol*, vol. 66, no. 6, pp. 5316-5329, Jun. 2017.

[18] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, "RF-based human activity recognition using signal adapted convolutional neural network," *IEEE Trans Mob Comput*, vol. 22, no. 1, pp. 487-499, Jan. 2021.

[19] A. Gaikwad, "A comparative study of machine learning techniques for human activity recognition," *Journal of Informatics Education and Research*, vol. 4, no.2, pp. 1427–1434, Dec 2024.

[20] H. A. N. Huimei, Z. H. U. Xingquan, and L. I. Ying, "Generalizing long short-term memory network for deep learning from generic data," *ACM Trans Knowl Discov Data*, vol. 14, no. 2, pp. 1-28, Feb 2020.

[21] C. Shiranthika, N. Premakumara, H. L. Chiu, H. Samani, C. Shyalika, and C. Y. Yang, "Human activity recognition using CNN & LSTM," In Proc. 5th International Conference on Information Technology Research: Towards the New Digital Enlightenment, Jan 2020, pp. 1-6.

[22] Y. A. Khan, S. Imaduddin, Y. P. Singh, M. Wajid, M. Usman, and M. Abbas, "Artificial intelligence based approach for classification of human activities using MEMS sensors data," *Sensors*, vol. 23, no. 3, p. 1275, Jan 2023.

[23] V. Ghate and C. Sweetlin Hemalatha, "Hybrid deep learning approaches for smartphone sensor-based human activity recognition," *Multimed Tools Appl*, vol. 80, no. 8, pp. 35585-35604, Nov 2021.

[24] Z. Y. Lim, L. Y. Ong, and M. C. Leow, "Radio frequency-based human activity dataset collected using ESP32 microcontroller in line-of-sight and non-line-of-sight indoor experiment setups," *Data in Brief*, vol. 57, no. 2024, p. 111101, Dec 2024.