Journal of Electronics, Electromedical Engineering, and Medical Informatics
Multidisciplinary: Rapid Review: Open Access Journal Vol. 7, No. 2, April 2025, pp: 533-545; eISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Manuscript received November 30, 2024; Accepted March 1, 2025; date of publication April 18, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i2.642

How to cite: Raden Rofiq Yudha Setyawan, Fiky Y. Suratman, and Khilda Afifah, “Development of Human Activity Recognition (HAR) for Health Rehabilitation Using MMWAVE Radar with 3D Point Cloud Data”, Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 7, no. 2, pp. 533-545, April 2025.

Development of Human Activity Recognition (HAR) for Health Rehabilitation Using MMWAVE Radar with 3D Point Cloud Data

Raden Rofiq Yudha Setyawan[1], Fiky Y. Suratman[1, 2], and Khilda Afifah[1, 2]

[1] Telkom University Center of Excellence Intelligence Sensing-IoT, Telkom University, Bandung, Indonesia
[2] School of Electrical Engineering, Telkom University, Bandung, Indonesia

Corresponding author: Raden Rofiq Yudha Setyawan (e-mail: rofiqys@student.telkomuniversity.ac.id)

ABSTRACT Postoperative recovery is a crucial phase in ensuring successful rehabilitation. However, many healthcare facilities face challenges due to the limited availability of medical personnel, making routine patient monitoring difficult. This limitation can delay the early detection of complications and reduce overall recovery effectiveness. To address this issue, this study proposes a non-invasive, radar-based system for remote postoperative patient monitoring. The proposed system utilizes the IWR6843AOP radar to generate 3D point cloud data, spatially representing patient movements. This approach enables continuous monitoring without compromising patient privacy, allowing healthcare providers to offer more efficient care. The collected data undergoes preprocessing, including normalization, labeling, and dataset splitting, before being classified using deep learning models such as 3D CNN, 3D CNN+LSTM, 3D CNN+Bi-LSTM, PointNet, PointNet++, and RNN. The dataset consists of six activity categories: empty space, sitting, standing, walking, running, and squatting, recorded at a frame frequency of 18.18 Hz. Experimental results show that the 3D CNN combined with Bi-LSTM achieves the highest accuracy of 90%, surpassing models like PointNet and RNN. These findings indicate that a radar-based and deep learning-driven approach offers an accurate, efficient, and non-intrusive solution for postoperative monitoring, reducing the need for direct medical supervision. This technology has significant potential for broader healthcare applications, contributing to more advanced, accessible, and technology-driven patient monitoring systems. By integrating artificial intelligence and radar sensing, this research paves the way for innovative solutions in modern healthcare, ensuring better postoperative outcomes while optimizing medical resources.

INDEX TERM Postoperative recovery, 3D point cloud, Deep learning, Human activity recognition.

I. INTRODUCTION

Postoperative rehabilitation, especially after major surgeries, is a crucial part of the recovery process. Many patients know recovery is important, but may not realize that proper rehabilitation helps reduce complications, speed up healing, and support returning to normal activities. During rehabilitation, regular monitoring of a patient's condition is essential to assess recovery progress, determine when treatment can proceed to the next stage, or decide whether additional interventions are necessary [1], [2]. This stage includes physical, psychological, and social recovery. In this process, routine monitoring is essential for evaluating recovery progress, adjusting medical interventions, and early detection of signs of deterioration to identify potential complications that may delay recovery or endanger the patient’s health[3].

In practice, postoperative patients often exceed medical personnel capacity, especially in remote areas. This is a common challenge in many healthcare facilities, where physicians struggle to monitor patients regularly due to a shortage of healthcare workers[4]. In such situations, remote monitoring technology offers an effective solution, enabling doctors to access patient data and evaluate their condition without the need for daily in-person interactions. This technology reduces the burden on healthcare systems and facilitates better management of postoperative patients[5]. Remote monitoring technology offers an innovative solution. These systems allow healthcare providers to monitor patients’ physical conditions remotely and in real-time without physical presence, thus accelerating medical decision-making and reducing the need for resource-intensive visits.

However, most current remote monitoring systems still rely on surveillance cameras or wearable sensors based on Inertial Measurement Units (IMUs). Each of these approaches has limitations. Camera-based systems raise privacy concerns despite detecting behavior. Cameras continuously record patients’ faces and activities, which can be considered invasive and uncomfortable, particularly in private spaces like hospital rooms, bathrooms, or homes [6][7].

On the other hand, IMU-based sensors require direct attachment to the patient’s body. This can be uncomfortable, especially for postoperative patients who experience sensitivity or pain around the surgical site. Additionally, wearable devices require extra attention such as charging or periodic adjustments, which can burden both patients and healthcare providers [8], [9], To address these issues, researchers have explored the use of millimeter-wave (mmWave) radar as a more comfortable and non-invasive monitoring solution. Unlike cameras or wearable sensors, mmWave radar enables contactless monitoring without visual recording. The radar emits electromagnetic waves and captures their reflections to detect movement and body positions, without capturing images or touching the patient's body. This makes it ideal for preserving patient privacy and enhancing comfort during recovery[10], [11]. mmWave radar is capable of generating 3D point cloud data, which represents (x, y, z) positions along with additional parameters such as Signal-to-Noise Ratio (SNR) and Doppler frequency shifts to detect movement speed. This information is rich and can be used to deeply analyze patients' physical activity patterns. The radar is also resistant to changes in lighting or other environmental conditions, making it reliable for 24-hour monitoring[12].

Several studies have demonstrated the effectiveness of mmWave radar in Human Activity Recognition (HAR)—the process of identifying human activities such as sitting, standing, walking, lying down, or falling—using sensor data. HAR systems are crucial in rehabilitation contexts as they enable doctors to quantitatively assess the patient's recovery progress. Radar-based HAR systems offer the advantage of requiring no cameras or wearable devices, making them suitable for long-term use[13].

Despite the rapid development of radar technology for HAR, most existing studies have focused on fall detection for the elderly, security surveillance, or gesture control, rather than on monitoring the postoperative rehabilitation process. There is a notable lack of research that specifically applies mmWave radar for monitoring patients recovering from medical procedures. Additionally, most publicly available HAR datasets do not reflect actual rehabilitation conditions, such as limited mobility, the use of assistive devices, or variations in recovery stages.

Furthermore, the majority of radar-HAR studies emphasize classification accuracy without considering aspects such as usability, result interpretability, or integration challenges within clinical environments. For such systems to be widely adopted in healthcare practice, they must not only be accurate but also user-friendly and understandable to medical personnel without technical backgrounds. This study proposes a non-invasive patient activity monitoring system using high-resolution 3D point cloud data from mmWave radar to support Human Activity Recognition (HAR) in the context of postoperative rehabilitation, which requires regular motion monitoring. The radar system will be used to capture movement data of patients performing various common rehabilitation activities in a simulated clinical environment.

The data will train machine learning algorithms to classify activities. The dataset will reflect real patient conditions, including limited movement, use of assistive tools, and short-duration activities. The system evaluation will focus on accuracy, ease of use, and user privacy to ensure its feasibility for implementation in medical settings. The primary goal of this research is to develop and evaluate a radar-based HAR system to support physical activity monitoring of postoperative patients without requiring direct interaction with doctors non-invasive and efficient. The system is targeted to be an accurate, privacy-preserving, user-friendly solution that can be seamlessly integrated into existing healthcare systems. The contribution of this paper is as follows:

1. Development of a non-invasive HAR system based on mmWave radar specifically designed for postoperative rehabilitation monitoring.
2. Creation of a rehabilitation activity dataset using 3D point cloud data collected from volunteers or patients in simulated clinical scenarios.
3. Demonstration of the system's advantages in comfort and privacy compared to camera-based or wearable systems.
4. Comprehensive evaluation of system accuracy and reliability, including real-world trial scenarios and usability testing by healthcare personnel.
5. Design of an integration framework for incorporating the system into telemedicine platforms to extend access to rehabilitation services in remote or underserved areas.

Numerous studies have shown the effectiveness of radar technology in monitoring patient activity and vital signs, particularly in rehabilitation contexts[13]. This technology can identify changes in movement patterns, detect physical activity levels, and record motion variations all crucial for evaluating postoperative recovery progress[8]. Besides preserving patient privacy, radar-based remote monitoring expands healthcare reach, allowing physicians to provide optimal care without being physically present[14]. Therefore, radar technology proves to be highly effective for Human Activity Recognition (HAR). This study recommends the use of radar-based 3D point cloud data for non-invasive postoperative patient monitoring, enabling high-accuracy activity detection while maintaining patient privacy [15].

II. MATERIALS AND METHOD

A. RADAR DATA PROCESSING

This study utilizes the Texas Instruments IWR6843AOP radar, a compact and highly integrated millimeter-wave (mmWave) sensor specifically designed to generate 3D point cloud data directly in JSON format using Radar Toolbox version 2.20.00.05 from the Texas Instruments library, thereby reducing the need for extensive preprocessing and manual configuration. Unlike conventional radars that require complex raw data processing, the IWR6843AOP features integrated on-chip signal processing, which significantly simplifies the development workflow.

Data acquisition is carried out using the Radar Toolbox from Texas Instruments, which provides a user-friendly interface for data collection and visualization. Each output data frame contains essential information such as 3D coordinates, signal performance parameters, and object motion data. The use of this radar also facilitates system calibration, as the device is equipped with integrated supporting features, thereby making the experimental preparation process more efficient.

The main advantage of this system lies in its Multiple Input Multiple Output (MIMO) architecture, which enables more accurate and comprehensive spatial scanning. With its built-in signal processing capabilities, including Range FFT, Doppler FFT, and angle estimation algorithms, the radar can generate real-time spatial representations without the need for additional external processing, which typically requires high-performance computing hardware such as GPUs or FPGAs. This feature is particularly valuable in various fields such as human motion tracking, autonomous vehicle navigation, smart sensor-based security systems, gesture recognition interfaces, and indoor occupancy monitoring in smart buildings. Moreover, the data generated in JSON format is optimized for seamless integration with modern data processing platforms such as Python, MATLAB, and various cloud-based IoT systems. This facilitates faster, more flexible, and efficient data-driven research and development.

The direct integration of Doppler information into the point cloud output represents another significant benefit of this system. The velocity data embedded within each point enables more comprehensive motion analysis. The radar also features flexible communication interfaces and power-efficient operation, making it well-suited for embedded system implementations. In some preliminary tests, the radar has proven to be stable and consistent in capturing various human movement patterns, even when assistive devices are used by the individuals. By combining these advanced features, the IWR6843AOP delivers a powerful yet practical radar solution for sensor-based research and development applications. The system's balance of performance and usability makes it particularly effective for prototyping and deploying mmWave radar solutions across various domains.

$$\vartheta D = \frac{\lambda \cdot f_D}{2} \tag{1}$$

where f_D is the Doppler shift and λ is the wavelength of the transmitted signal Eq. (1) [16][12][17][18].

Unlike previous generations of radar, which only produced raw data requiring additional processing to obtain 3D point cloud representation as shown in FIGURE 1, the IWR6843AOP radar can directly produce 3D point cloud data[14], [19], [20]. The range R of a detected object is determined using the beat frequency f_b in a frequency modulated continuous Wave(FMCW) radar system, following the equation Eq. (2)[16]:

$$\vartheta = \frac{c \cdot f_b}{2S} \tag{2}$$

where c is the speed of light and S is the slope if transmitted frequency chirp. This capability streamlines the data processing workflow, reduces initial computational overhead, and provides greater flexibility for users to leverage the data for various applications [21][22].

As an initial step in this study, the JSON-formatted data will be converted into CSV files. This conversion is essential to streamline the process of analysis, labeling, and data organization, which are critical for the subsequent stages of model development. JSON format, while flexible and widely used for data interchange, is less suited for direct manipulation and batch processing in machine learning workflows. By transforming the data into CSV format, researchers can take advantage of tabular data structures that are more compatible with various data analysis tools and machine learning libraries.

Moreover, converting the data into CSV enables efficient handling of large datasets, which is crucial for training deep learning models. Tabular data facilitates processes such as filtering, searching for specific values, grouping, and generating descriptive statistics, all of which are essential during the initial exploration of the dataset. CSV files offer simplicity, faster parsing, and better integration with frameworks such as TensorFlow, PyTorch, and scikit-learn. This process is also crucial to ensure that the data can be efficiently pre-processed, visualized in the form of graphs or 3D visualizations, and accurately validated before being used for model training. For example, visualizing the distribution of points in a 3D point cloud, analyzing object movement patterns, or tracking changes in velocity over time becomes much easier when the data is in tabular form.

Through this conversion step, the research workflow is expected to become more efficient, the potential for labeling errors can be reduced, and the performance of the developed model for recognizing human activities can be optimized. With a simplified data structure and high compatibility with various analytical tools, the development process of AI models can proceed faster and more accurately aligned with the primary objectives of this study.

B. DATASET AND DATA COLLECTION

The aim of this study is to thoroughly investigate significant differences in the classification accuracy of human activity recognition models, taking into account various body positions and movement patterns performed by the subjects. This study includes six main categories of observed activities: empty space, sitting, standing, walking, running, and squatting, representing a spectrum of movements ranging from static to dynamic conditions. Data was collected for 15 seconds and resulted in 390-450 data points for each movement, for each activity, from three subjects of varying heights, weights, and ages, represented by labels 1 to 6 [23]. The participants involved in this study were aged 23 and 24 years old, all male, and in normal physical condition without any history of health issues that could affect their movement or posture. All participants provided consent for their movement data to be used in this research.

Data collection was conducted using the IWR6843AOP radar from Texas Instruments, supported by Radar Toolbox version 2.20.00.05. The radar produced 3D point cloud data in JSON format, including information such as frame number, timestamp, coordinates (x, y, z), Signal-to-Noise Ratio (SNR), and Doppler velocity. The radar was installed in a 6 x 3-meter room at a height of 1.5 meters, as depicted in Figure. 2. This setup was calibrated and strategically placed to ensure optimal capture of subject movements, utilizing the three-dimensional spatial representation offered by the radar. The radar was configured with a frame frequency of 18.18 Hz, sufficient to capture detailed patterns of human activity changes, such as transitions from sitting to standing or walking, without missing crucial details. This frequency offers an optimal balance between temporal resolution and data processing efficiency, ensuring the generated data is adequate for high-precision human activity analysis. Additionally, supplementary information such as movement speed (Doppler) and signal-to-noise ratio (SNR) adds valuable dimensions to the data, helping to assess signal quality and movement dynamics.

After collection, the JSON-formatted data was converted into CSV format to facilitate labeling, analysis, and deep learning model development. The collected data underwent multiple validation cycles to ensure accuracy and consistency. The resulting dataset provides a comprehensive and accurate depiction of the subjects' movement patterns within the room, including dynamic transitions between activities. By combining spatial information, velocity, and signal quality, this dataset enables in-depth analysis of various human activities. This research also opens opportunities for applications in healthcare monitoring, such as posture detection and movement transition analysis.

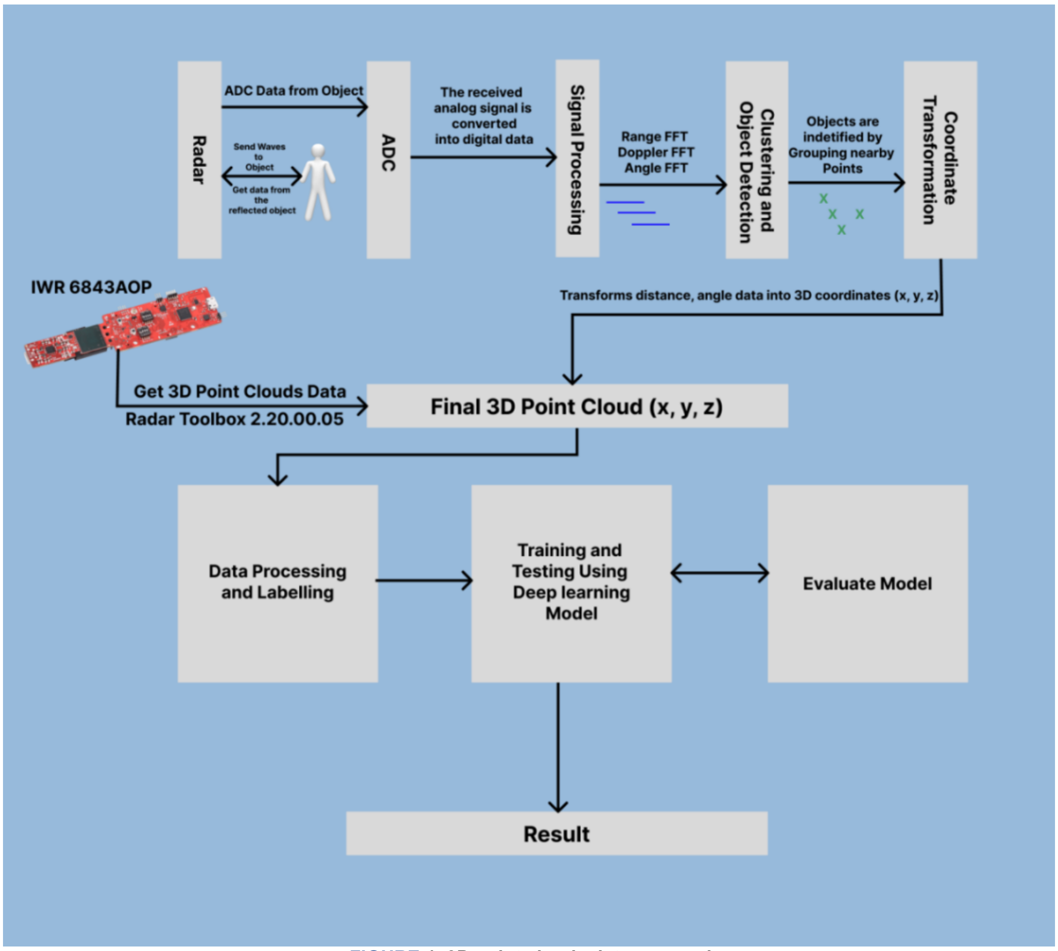

FIGURE 1. 3D point clouds data processing

FIGURE 2. Data Collecting Process

C. DATA PROCESSING

Data processing is a crucial step in preparing datasets for training machine learning models. This process aims to ensure that the data used in this research is of high quality, free from errors, and ready to be utilized in both the training and evaluation phases of the model to be developed. Without a systematic and thorough data processing stage, the model is at risk of experiencing performance degradation due to the presence of irrelevant data, noise, overfitting, or inconsistencies in the input format. Therefore, data processing is not only an initial step but also a fundamental stage that supports the overall success of the machine learning pipeline implemented in this study. The following are several stages carried out during the data processing procedure.

The first stage is data loading, which involves importing the collected dataset, followed by the removal of empty or NaN values and filtering out invalid entries. This step is essential to ensure that only clean and relevant data is fed into the training and testing systems, minimizing errors and improving model accuracy. Eliminating invalid data also helps reduce processing complexity and avoids potential biases caused by inappropriate or unrepresentative input. Furthermore, with cleaner data, the model can focus more effectively on learning important patterns without being distracted by uninformative data.

After the data has been cleaned, the next step is feature extraction, which plays a crucial role in determining the most optimal data representation for model training. Feature extraction is performed by identifying and selecting key attributes that are believed to significantly impact model performance. In the context of this research, the extracted features include point coordinates in three-dimensional space (x, y, z), Doppler magnitude—representing the relative motion speed of an object toward the sensor—and the signal-to-noise ratio (SNR). The SNR value is calculated using the formula in Eq. (3) [24]:

$$SNR = 10\log_{10}\left(\frac{Psignal}{\mathrm{Pnoise}}\right) \tag{3}$$

where $Psignal$ refers to the power of the signal received from the observed object, and Pnoise represents the power of disturbances or unwanted signals originating from the surrounding environment, interference, or sensor hardware limitations. SNR serves as an important indicator in assessing data quality, as a higher SNR indicates a greater proportion of useful information that the model can leverage for classification. The selection of these features is based on their functionality and contribution to enhancing the model's ability to accurately recognize and distinguish patterns of human activity. Following this, labeling is performed where each movement in the dataset is assigned a label for classification purposes. The dataset consists of five different movement types along with one empty state, leading to data labels ranging from 1 to 6. Proper labeling is crucial for supervised learning tasks as it allows the model to differentiate between various movement patterns accurately.

To maintain consistency in feature scales, normalization is applied, ensuring all feature values fall within the same range (e.g., 0 to 1). In Human Activity Recognition (HAR) tasks, normalization prevents models from being biased by varying feature scales, enhancing their ability to detect activities more effectively. Data splitting is then performed to divide the dataset into training and testing subsets. An 80:20 split is commonly used, where 80% of the data is allocated for training and the remaining 20% is reserved for testing. This split ensures that the model learns patterns while being evaluated for generalization on unseen data.

Lastly, one-hot encoding is applied to categorical labels, converting them into a binary format that is more suitable for model training. This encoding method enhances the learning process by ensuring the model interprets categorical data correctly. By following these systematic data processing steps, the dataset is well-prepared for training deep learning models, increasing the likelihood of successful point cloud feature classification.

D. ALGORITHM

In this study, we implemented several algorithms applicable to point cloud data, including 3D-CNN, 3D-CNN - BiLSTM, 3D-CNN - LSTM, PointNet, PointNet++, RNN, and RNN + LSTM. Based on previous research, the combination of CNN and BiLSTM has demonstrated high accuracy and robust model performance, as discussed in[25][26][27]. Other studies, such as[11][13], [19], [20], have also tested similar models with promising results. In this research, we used 100 epochs, considering it optimal to provide sufficient training time for the model. The number of epochs exceeds those used in prior studies, such as[9], [13], [28], aiming for more optimal results.

The 3D-CNN + BiLSTM model in this study utilized processed point cloud data, including feature normalization (pointX, pointY, pointZ, Doppler, SNR) and one-hot encoding of labels for multi-class classification. The data was then rearranged into a three-dimensional format to enable temporal analysis using LSTM-based architecture. The 3DCNN operation can be mathematically reoresented as :

$$\mathrm{Y(i,k,k,f)} = \sum_{h=0}^{kh-1}\sum_{w=0}^{kw-1}\sum_{c=0}^{c-1}\mathrm{X}(i+h,j+w,k+d,c)\cdot \mathrm{K(h,w,d,c,f+bf)} \tag{4}$$

In this equation, $\mathrm{X}(i+h,j+w,k+d,c)$ represents the input tensor corresponding to the normalized point cloud data at spatial position $(i+h,j+w,k+d,c)$ and input channel c. The kernel $\mathrm{K(h,w,d,c,f)+bf}$ denotes the 3D convolutional filter at position $\mathrm{(h,w,d)}$ for input channel c and output feuture index f. the resulting output feature map is denoted by Y(i, j, k, f), which represents the extracted feature at spatial location (i, j, k) for the f feature map. This 3D convolution operation captures local spatial features from the input point cloud by sliding the kernel over the input volume, performing element-wise multiplication and summation across spatial and channel dimensions, followed by bias addition, as expressed in Eq. (4) [29]. The extracted features are then passed to a Bidirectional LSTM (BiLSTM) layer for temporal analysis. The BiLSTM processes the sequence in both forward and backward directions, with the hidden states computed as:

$$\vec{h} = LSTM(x_t,\vec{h}_{t-1}),\ \ \overleftarrow{h_t} = LSTM(x_t,h_{t+1}) \tag{5}$$

The final hidden state h_t is the concatenation of the forward and backward states.

$$h_t = [\vec{h}_t;\overleftarrow{h}_t] \tag{6}$$

This bidirectional processing captures temporal dependencies in data effectively Eq. (5)&(6) [30]. The model architectufre consisted of several layers, including a Time Distributed Dense layer for capturing temporal features, a Bidirectional LSTM layer for bidirectional information capture, a dense layer with 128 neurons for feature abstraction, and a softmax output layer for six-class classification. The softmax function is defined as:

$$\mathrm{yi} = \frac{e^{zi}}{\sum_{j=1^{zj}}^{K}} \tag{7}$$

where yi is the predicted probability for class I, and zi is the logit for class I Eq. (7)[31] [20]. The model was trained for 100 epochs using the Adam optimizer, which updates the model parameters as follows:

$$\theta_t = \theta_{t-1} - \eta \frac{\widehat{m}_t}{\sqrt{\hat{v}_t} + \in} \tag{8}$$

where η is the learning rate, $\widehat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of the first and second moments of the gradients, respectively Eq. (8) [32]. The training utilized the categorical cross-entropy loss function, defined as:

$$\mathcal{L} = \sum_{i=1}^{N} . \sum_{j=1}^{K} \mathrm{y_{ij}} \log(\hat{\mathrm{y}}_{\mathrm{ij}}) \tag{9}$$

where y_{ij} is the true label (one-hot encoded) and $\hat{y}_{ij}$ is the predicted probability Eq. (9) [33][34]. A batch size of 32 used, woth validation on 20% if the training data.

The training results indicated strong performance with high accuracy, a robust F1 score, and solid precision and recall metrics, reflecting the model's ability to classify activities based on point cloud data effectively. The training history visualization also demonstrated stability and consistency in performance throughout the training process.

III. RESULT

A. ACCURACY

The classification accuracy of various methods evaluated in this study involved several models, namely 3D CNN, 3D CNN + LSTM, 3D CNN + Bi-LSTM, PointNet, PointNet++, RNN, and RNN + LSTM. Among all the models tested, the 3D CNN + Bi-LSTM architecture consistently demonstrated the best performance in terms of overall accuracy, precision, recall, and F1-score across all human activity classes, making it the most effective and reliable model recommended in this study for radar point cloud-based activity recognition. This model consistently delivers more stable classification results, even when dealing with varying data across different human activity classes.

Technically, the 3D CNN component plays a vital role in extracting crucial spatial features from the highly complex and unstructured point cloud data. The model is capable of identifying relevant spatial structures in the three-dimensional space, which is crucial for understanding the context of human activities. Meanwhile, the Bi-LSTM (Bidirectional Long Short-Term Memory) is responsible for capturing bidirectional temporal patterns, allowing the model to consider both past and future information simultaneously. This is critical because human activities that occur in real-time are usually interdependent on temporal relations, which can significantly affect the classification decision.

This combination makes the model particularly well-suited for data that inherently contains both spatial and temporal dimensions, such as the dataset used in this research, which reflects sequences of human activities in a three-dimensional space and temporally interconnected patterns. This combined approach significantly enhances the model's ability to understand complex and dynamic sequences of human movement.

As shown in TABLES 1, TABLE 2 and TABLE 3, the performance of the 3D CNN + Bi-LSTM model remains consistently high across all activity classes. For instance, in the walking class (class 2), the model achieved an F1-score of 0.92, significantly outperforming PointNet (0.84), PointNet++ (0.81), and even the standalone 3D CNN model (0.84). Although the 3D CNN + LSTM model also showed strong performance with an F1-score of 0.91, the addition of the Bi-LSTM layer still resulted in a measurable improvement. This performance difference indicates that the bidirectional information captured by Bi-LSTM provides a more comprehensive temporal context compared to the unidirectional LSTM. The consistency of this model is also evident in the standing class (class 5), where it achieved an F1-score of 0.88, indicating strong stability in recognizing various types of human activities, including low-movement actions that are often difficult to classify accurately by other models.

TABLE 4 further supports these findings by presenting a comprehensive comparison of the models based on key evaluation metrics such as accuracy, MAE, precision, recall, and F1-score, all of which collectively demonstrate the superiority of the 3D CNN + Bi-LSTM model over the others. The 3D CNN + Bi-LSTM model achieved the highest accuracy at 89.66%, the lowest Mean Absolute Error (MAE) at 29.38%, and superior F1-score, recall, and precision metrics compared to all other tested models. The dominance of these metrics demonstrates the superiority of the architecture in handling complex and high-dimensional radar point cloud data. From these results, it can be concluded that the 3D CNN + Bi-LSTM model delivers the most optimal performance for human activity classification based on radar point cloud data. Its strengths lie not only in high accuracy but also in the consistent and stable performance across all evaluation metrics used in this study. In other words, this model excels not only in terms of numerical performance but also in reliability, generalization, and potential applicability across various real-world scenarios.

B. PERFORMANCE

The classifier is designed to evaluate the accuracy of five sequentially performed movements and detect empty room conditions. Once data is collected from the six classification categories five types of movements and one empty room condition it is merged into a comprehensive dataset. This merging step is followed by data shuffling to increase variability during the model training and testing processes. This technique is applied to improve the model's classification accuracy and enhance its performance on a more diverse and realistic data distribution. A well-shuffled and representative dataset plays an important role in shaping a reliable classifier capable of handling subtle variations in activity execution

The testing results revealed that the 3D-CNN + BiLSTM model achieved the highest accuracy, approximately 90%, making it the best performer compared to the five other models evaluated on the same dataset. FIGURE 2 illustrates a comparative accuracy graph of the six tested models: 3D-CNN, 3D-CNN + BiLSTM, 3D-CNN + LSTM, PointNet, PointNet++, RNN, and RNN + LSTM, each tested across six model types (1, 2, 3, 4, 5, and 6). Among these models, 3D-CNN + BiLSTM demonstrated the most superior performance, achieving the highest accuracy of 90% on model type 2. This was followed by 3D-CNN + LSTM, RNN + LSTM, and RNN, which also showed relatively. High accuracy. Conversely, the 3D-CNN model

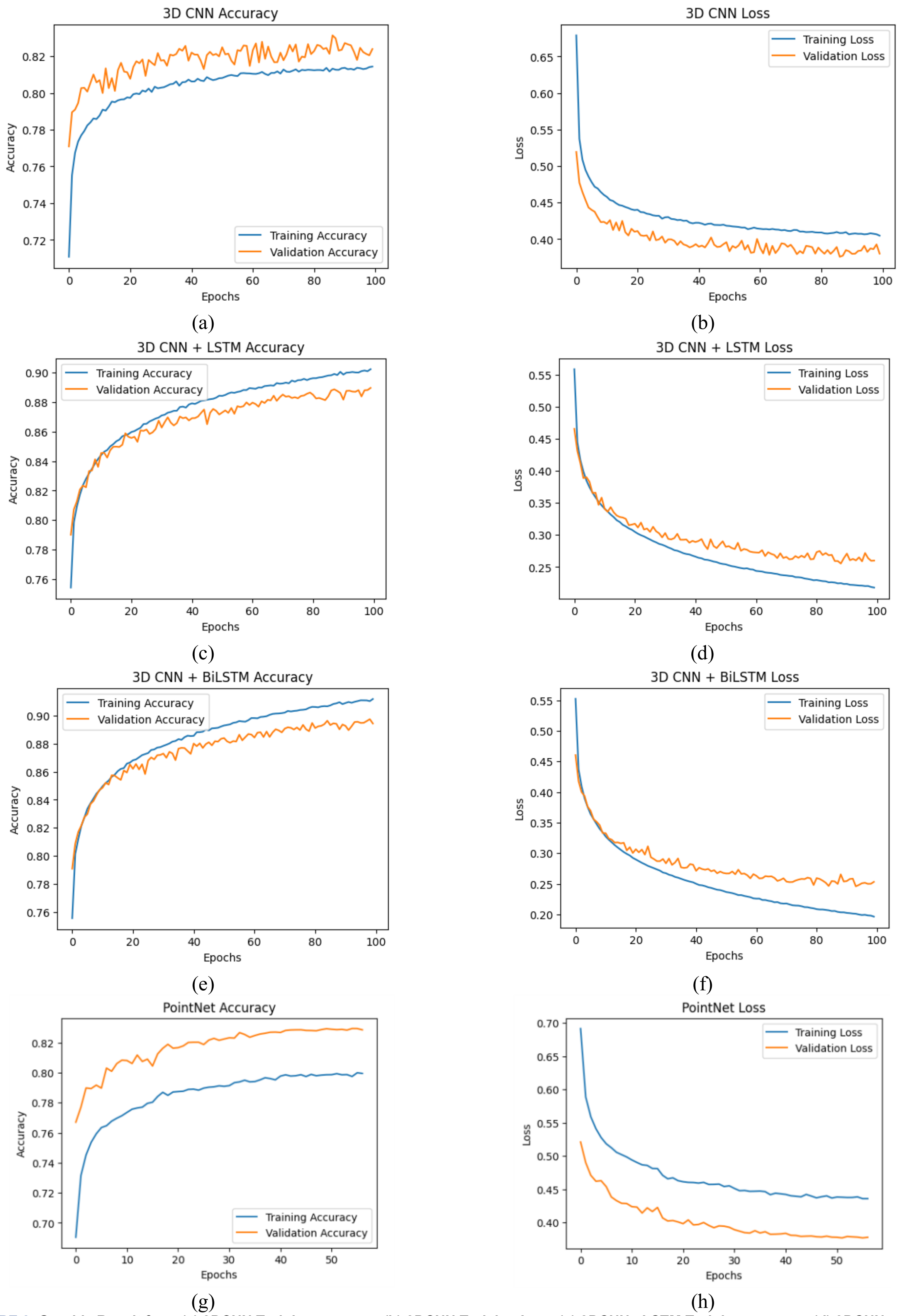

FIGURE 3. Graphic Result from (a) 3DCNN Training accuracy (b) 3DCNN Training Loss (c) 3DCNN +LSTM Training accuracy (d) 3DCNN +LSTM Training Loss (e) 3DCNN + BiLSTM Training accuracy (f) 3DCNN + BiLSTM Training Loss (g) PointNet Training accuracy (h) PointNet Training Loss

exhibited the lowest performance, with an accuracy of 82%, indicating that without further temporal processing such as LSTM or Bi-LSTM layers, this model lacks the capacity to capture the time dynamics inherent in complex sequences of human activities.

Further analysis indicated that model type 2 yielded the highest accuracy across all classifiers, followed by model types 6 and 5. FIGURE 2 provides a visual comparison of the accuracy of each model, offering a clearer view of the performance differences among the models and the tested movement categories. These findings underscore the importance of selecting the appropriate model to support effective and accurate motion recognition (TABLE 1).

FIGURE 3 (a) shows that the 3D CNN achieved a training accuracy of over 82%, with validation accuracy also increasing but slightly lower, indicating potential overfitting. The training loss gradually decreased to approximately 0.40, while the validation loss also declined but remained higher, reflecting challenges in generalization. The 3D CNN is designed to capture spatial features from 3D point cloud data effectively, such as object position and shape in space. However, this model has inherent limitations in processing temporal relationships between frames, as its architecture does not explicitly consider time sequences or motion dynamics. This becomes a major challenge in human activity recognition, where activity patterns often involve a sequence of movements that change continuously over time. Additionally, the 3D CNN is prone to overfitting, especially when processing high-dimensional point cloud data with a limited amount of training data[35].

In contrast, FIGURE 3 (c) shows that the 3D CNN + LSTM achieved a training accuracy of nearly 90%, with validation accuracy almost parallel, indicating better generalization capabilities. This combination is effective because the 3D CNN can extract spatial features such as structure and point distribution in three-dimensional space, while the LSTM complements this by capturing temporal sequences from frame to frame, enabling the model to understand the context of motion that occurs progressively over time. Thus, enabling the model to not only understand the position of objects in space but also how those objects move over time, which is highly relevant for human activity recognition applications. The training loss stabilized at around 0.30, Which indicates that this model is more stable and efficient in learning human activity patterns involving changes in time and space.

The combination of 3D CNN + BiLSTM in FIGURE 3 (e) demonstrated more optimal performance, with training accuracy reaching around 90% and validation accuracy closely matching, which reflects the model's remarkable ability not only to generalize well but also to capture more complex temporal information, which is crucial for recognizing a wide range of human movement patterns. The training loss dropped to approximately 0.25, and the validation loss also declined, indicating that this combination is highly effective in maintaining model stability and efficiency. BiLSTM offers an additional advantage over standard LSTM because it can process information in both directions, reading the data sequence forward and backward, allowing it to capture more complex and relevant temporal patterns in human activity data.

On the other hand, FIGURE 3 (g) shows that PointNet achieved a training accuracy of around 80%, which indicates a slight overfitting, where the model fits the training data more closely than it can generalize to broader data. The training loss decreased to about 0.50, while the validation loss stagnated at around 0.40, highlighting instability in generalization. This limitation of PointNet arises because the model processes each point independently without capturing local spatial or temporal relationships in the point cloud data, which are essential for recognizing complex human activity patterns[20].

TABLE 1
Result Classification

Class	Precision	Recall	F1-Score	Support	Class	Precision	Recall	F1-Score	Support
		3DCNN					**3DCNN + LSTM**		
0	0.62	0.52	0.57	101	0	0.70	0.73	0.72	101
1	0.69	0.64	0.67	4432	1	0.76	0.70	0.73	4432
2	0.79	0.89	0.84	11050	2	0.90	0.92	0.91	11050
3	0.92	0.95	0.93	5763	3	0.93	0.97	0.95	5763
4	0.87	0.77	0.82	11034	4	0.92	0.90	0.91	11034
5	0.85	0.85	0.85	8988	5	0.88	0.88	0.88	8988
		3DCNN + Bi-LSTM					**PointNet**		
0	0.78	0.74	0.76	101	0	0.68	0.52	0.59	101
1	0.73	0.77	0.75	4432	1	0.71	0.54	0.62	4432
2	0.91	0.93	0.92	11050	2	0.82	0.86	0.84	11050
3	0.94	0.96	0.95	5763	3	0.90	0.97	0.94	5763
4	0.93	0.90	0.92	11034	4	0.85	0.81	0.83	11034
5	0.90	0.86	0.88	8988	5	0.82	0.87	0.85	8988

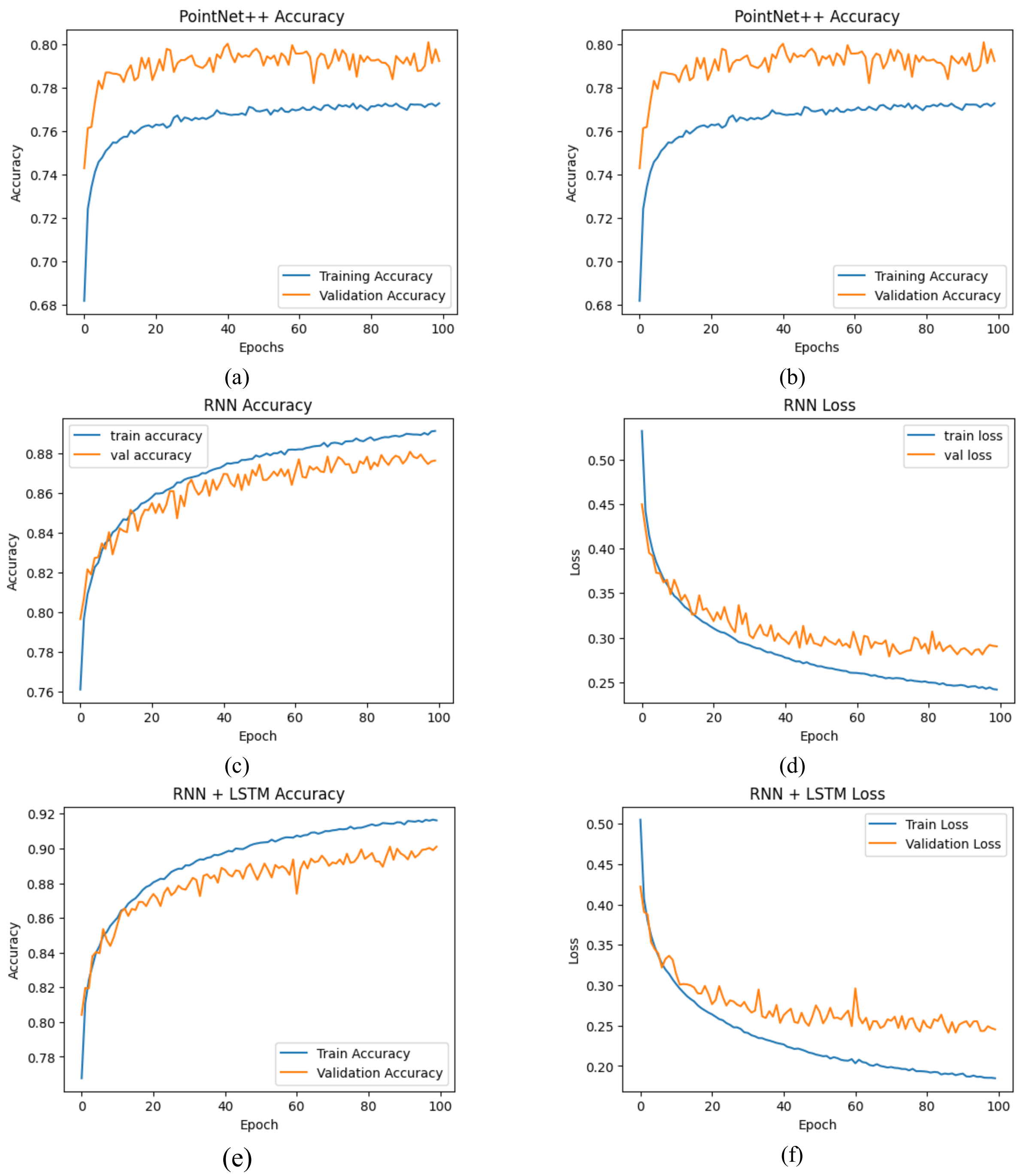

Figure 4. Graphic Result from (a) PointNet++ Training accuracy (b) PointNet++Training Loss (c) RNN Training accuracy (d) RNN Training Loss (e) RNN+LSTM Training accuracy (b) RNN+LSTM Training Loss

In FIGURE 4 (a) PointNet++ achieved a training accuracy of approximately 80%, while validation accuracy was slightly lower at around 77%, indicating potential overfitting. The training loss remained stable at around 0.50, reflecting instability in generalization. In contrast, as shown in FIGURE 4(d), the RNN model showed better performance, achieving 88% training accuracy with closely matched validation accuracy, indicating balanced generalization. The consistent accuracy suggests the model could adapt well to new input. The training loss also decreased to 0.30, while the validation loss remained low, showing that the model effectively learned and retained important features throughout training. The combination of RNN + LSTM in FIGURE 4(e) showed optimal performance, achieving nearly 89% training accuracy with closely matched validation accuracy. The training loss was as low as 0.20, with validation loss significantly reduced, reflecting good model stability.

IV. DISCUSSION

This study aims to analyze the significant differences in classification accuracy among various deep learning models for Human Activity Recognition (HAR) based on 3D point cloud data derived from mmWave radar sensors, which are increasingly utilized in non-intrusive patient monitoring systems during the medical rehabilitation process. The results show that the combined 3D CNN and BiLSTM model achieved the highest accuracy at 89.66%, outperforming other models such as PointNet, RNN, and standalone 3D CNN.

TABLE 2
Result Classification

Class	Precision	Recall	F1-Score	Support	Class	Precision	Recall	F1-Score	Support
PointNet++					RNN				
0	0.79	0.11	0.19	101	0	0.65	0.70	0.67	101
1	0.69	0.51	0.59	4432	1	0.73	0.72	0.72	4432
2	0.77	0.84	0.81	11050	2	0.88	0.91	0.89	11050
3	0.88	0.97	0.93	5763	3	0.92	0.97	0.95	5763
4	0.82	0.76	0.79	11034	4	0.91	0.87	0.89	11034
5	0.81	0.86	0.83	8988	5	0.88	0.86	0.87	8988

TABLE 3
RNN + LSTM

Class	Precision	Recall	F1-Score	Support
RNN + LSTM				
0	0.74	0.66	0.70	101
1	0.73	0.77	0.75	4432
2	0.91	0.92	0.92	11050
3	0.94	0.96	0.95	5763
4	0.92	0.91	0.92	11034
5	0.90	0.86	0.88	8988

0 =	Empty Room	3 =	Squatting
1 =	Sitting	4 =	Running
2 =	Walking	5 =	Standing

The strength of this hybrid model lies in its ability to process spatial and temporal information simultaneously, enabling more accurate detection of complex motion sequences typically observed in real-world rehabilitation scenarios. CNN effectively extracts spatial features from radar point cloud data, while BiLSTM captures temporal changes from past and future data sequences.

This combination enables the 3D CNN + BiLSTM model to recognize human movement patterns more comprehensively, by capturing not only the structure of poses from spatial data but also transitions between activities over time, which are critical for understanding patient recovery progression. In contrast, using CNN or 3D CNN alone tends to be limited in performance, as these models focus solely on extracting spatial information without incorporating the temporal dynamics that define motion sequences, thus reducing their effectiveness in HAR applications where activity duration and transition are essential features.

Meanwhile, RNN- or LSTM-based models also face considerable challenges in handling the irregular distribution and sparse nature of point cloud data, which can lead to unstable learning and reduced classification reliability without appropriate feature preprocessing. Therefore, integrating CNN and BiLSTM presents an appropriate approach, leveraging the strengths of both architectures to process the spatial and temporal dimensions of 3D radar point cloud data. When compared to previous studies summarized in TABLE 5, the proposed model in this research demonstrates competitive advantages in both accuracy and efficiency. First, a deep learning framework based on TimeDistributed CNN-LSTM architecture was implemented to process voxelized radar data from the MMActivity dataset [9], aiming to leverage spatial-temporal features for improved activity classification. This method integrated Convolutional Neural Networks (CNN) to extract spatial patterns from radar data slices and bidirectional Long Short-Term Memory (BiLSTM) networks to capture forward and backward temporal dependencies, achieving a classification accuracy of 90.47%, although the system's performance was not validated using point cloud data, potentially limiting its application in depth-aware HAR scenarios.

In a subsequent study, a hybrid model combining one-dimensional Convolutional Neural Networks (1D CNN) and LSTM was employed to classify Micro-Doppler signatures captured by radar sensors, resulting in a high recognition accuracy of 98.28% [21]. Although the reported performance is notable, the experimental setup was conducted using a relatively small and homogeneous dataset, and notably lacked the incorporation of 3D point cloud information, which is crucial for applications in HAR within complex, real-world environments where depth and spatial orientation vary significantly.

Another investigation explored the application of a TimeDistributed CNN-LSTM network on three-dimensional MRI scan data [5], yielding an accuracy of 98.90%. This result reinforces the efficacy of integrating spatial and temporal analysis for medical pattern recognition tasks. However, the domain-specific nature of the dataset medical imaging renders the model less transferable to radar-based HAR contexts, where data characteristics and environmental variability differ substantially. In [25], a feature fusion model integrating Principal Component Analysis Network (PCANet) with CNN-BiLSTM was applied to mmWave radar data, achieving the highest reported accuracy of 99.75% in the reviewed studies.

TABLE 4
Algorithm Comarison

Classifier	Accuracy(%)	MAE (%)	F1-Score	Recall	Precision
3DCNN	82.99	45.24	0.8287	0.8299	0.8310
3DCNN+LSTM	89.06	31.02	0.8898	0.8906	0.8894
3DCNN+Bi-LSTM	89.66	29.38	0.8968	0.8966	0.8978
PointNet	83.04	45.88	0.8269	0.8304	0.8270
PointNet++	80.45	51.69	0.7996	0.8045	0.8011
RNN	87.72	34.02	0.8768	0.8772	0.8770
RNN+LSTM	89.59	29.72	0.8961	0.8959	0.8968

TABLE 5
Comparison

Study	Model	Accuracy	Dataset	Limitations
This Study	3DCNN + BiLSTM	89.66%	PointCloud data (6 activities)	Limited to offline analysis; real-time performance not evaluated
Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar [9]	TimeDistributed CNN-LSTM	90.47%	MMActivity Dataset (voxelized radar data)	Limited to specific radar data; not generalizable to other modalities.
A Hybrid CNN-LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar [21]	1D CNN + LSTM	98.28%	Micro-Doppler radar data	Limited to small datasets; no point cloud data used.
TimeDistributed-CNN-LSTM: A Hybrid Approach Combining CNN and LSTM to Classify Brain Tumor on 3D MRI Scans Performing Ablation Study [5]	TimeDistributed CNN-LSTM	98.90%	3D MRI scans	Focused on medical imaging, not HAR.
Human Activity Recognition Method Based on FMCW Radar Sensor with Multi-Domain Feature Attention Fusion Network [22]	CNN+SMAN	97.58%	FMCW radar data	Complex architecture; high computational cost.
Human Multi-Activities Classification Using mmWave Radar: Feature Fusion in Time-Domain and PCANet [25]	PCANet + CNN-BiLSTM	99.75%	mmWave radar data	Limited to specific radar data; not generalizable to other modalities.
Multi-Sensor Data Fusion and CNN-LSTM Model for Human Activity Recognition System [26]	CNN-LSTM + Multi-Sensor Fusion	98.26%	Multi-sensor data (camera + radar)	Requires multiple sensors; not applicable to single-modality systems.

While the approach demonstrated robust HAR performance by utilizing statistical offset features and multi-level temporal-spatial fusion, its heavy reliance on specific radar configurations and high model complexity limits its generalizability across diverse environmental conditions and hardware platforms. A multi-modal sensor fusion approach integrating CNN-LSTM architectures with visual data from cameras and radar signals was explored in [26], producing an impressive accuracy of 98.26%. However, the necessity of synchronizing multiple sensors increases hardware complexity and deployment costs, thereby reducing the practicality of such systems in scenarios where minimal sensor infrastructure such as radar-only setups is preferred. In comparison, the model proposed in this study 3D CNN combined with BiLSTM achieved 89.66% accuracy using only point cloud data. It strikes a balance between performance and architectural simplicity, making it suitable for practical applications in medical rehabilitation monitoring. Unlike more complex fusion-based systems, it relies solely on radar point cloud input, avoiding the need for additional sensors, extensive calibration, or computationally intensive pre-processing pipelines. Nevertheless, this study has several limitations. First, the model has not yet been tested in real-time scenarios, despite the fact that real-time responsiveness is a critical requirement for health monitoring systems, especially when used for continuous assessment of patient activity or emergency detection in home-care settings. Second, experiments were conducted in a controlled environment with a limited number of participants (only three individuals), so broader generalization to larger populations requires further investigation. Third, the system has not been tested under varying environmental conditions, such as different room layouts, potential signal interference, or the use of mobility aids. These factors are crucial to ensure the system's applicability in diverse clinical settings. The implications of this study are highly promising in supporting more adaptive and personalized healthcare services. With its ability to classify patient activities accurately without the need for wearable devices or cameras, this system opens significant opportunities for developing remote rehabilitation services—especially for post-operative patients with limited mobility or in rural areas lacking access to rehabilitation specialists. The use of mmWave radar also allows for continuous monitoring without intruding on patient privacy, thereby improving both comfort and safety during recovery. In the future, this system has the potential to be integrated into telemedicine platforms or Electronic Health Records (EHRs), supporting smarter, more efficient, and targeted medical decision-making.

V. CONCLUSION

This study aimed to analyze the significant differences in classification accuracy among various deep learning models applied to point cloud data for human activity recognition (HAR). The results demonstrate that the 3D CNN + BiLSTM model outperformed other models, achieving an accuracy of 89.66%, while the 3D CNN model had the lowest accuracy of 82.99%. These findings highlight the effectiveness of combining spatial feature extraction (via 3D CNN) with temporal sequence modeling (via BiLSTM) for HAR tasks. Below, we provide a deep interpretation of the results, compare this study with previous works, discuss limitations, and explore the implications of the findings. This study aimed to analyze the significant differences in classification accuracy among various deep learning models applied to point cloud data for human activity recognition (HAR). The results demonstrate that the 3D CNN + BiLSTM model outperformed other models, achieving an accuracy of 89.66%, while the 3D CNN model had the lowest accuracy of 82.99%. These findings highlight the effectiveness of combining spatial feature extraction (via 3D CNN) with temporal sequence modeling (via BiLSTM) for HAR tasks. Below, we provide a deep interpretation of the results, compare this study with previous works, discuss limitations, and explore the implications of the findings.

REFERENCES

[1] F. Brodersen, J. Wagner, F. G. Uzunoglu, and C. Petersen-Ewert, "Impact of Preoperative Patient Education on Postoperative Recovery in Abdominal Surgery: A Systematic Review," *World J. Surg.*, vol. 47, no. 4, pp. 937–947, 2023, doi: 10.1007/s00268-022-06884-4.

[2] T. C. De Klerk, D. M. Dounavi, D. F. Hamilton, N. D. Clement, and K. T. Kaliarntas, "Effects of home-based prehabilitation on pre- and postoperative outcomes following total hip and knee arthroplasty: A systematic review and meta-analysis," *Bone Jt. Open*, vol. 4, no. 5, pp. 315–328, 2023, doi: 10.1302/2633-1462.45.BJO-2023-0021.

[3] K. N. Dainty, M. Bianca Seaton, and P. Richard Verbeek, "Moving from physical survival to psychologic recovery: a qualitative study of survivor perspectives on long-term outcome after sudden cardiac arrest," *Resusc. Plus*, vol. 5, no. September, p. 100055, 2021, doi: 10.1016/j.resplu.2020.100055.

[4] W. H. O. (WHO), *Global lob ppatient safety report 2024*, vol. 15, no. 1. 2024.

[5] S. Montaha, S. Azam, A. K. M. R. H. Rafid, M. Z. Hasan, A. Karim, and A. Islam, "TimeDistributed-CNN-LSTM: A Hybrid Approach Combining CNN and LSTM to Classify Brain Tumor on 3D MRI Scans Performing Ablation Study," *IEEE Access*, vol. 10, pp. 60039–60059, 2022, doi: 10.1109/ACCESS.2022.3179577.

[6] A. K. Alhazmi *et al.*, "Intelligent Millimeter-Wave System for Human Activity Monitoring for Telemedicine," *Sensors*, vol. 24, no. 1, pp. 1–23, 2024, doi: 10.3390/s24010268.

[7] F. Yan, N. Li, A. M. Iliyasu, A. S. Salama, and K. Hirota, "Insights into security and privacy issues in smart healthcare systems based on medical images," *J. Inf. Secur. Appl.*, vol. 78, no. October, p. 103621, 2023, doi: 10.1016/j.jisa.2023.103621.

[8] A. P. N.Lyons, A. Santra, "Improved Deep Representation Learning for Human Activity Recognition using IMU Sensors," *2021 20th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA).*, vol. 10.1109/IC, pp. 326–332, 2021, [Online]. Available: https://ieeexplore.ieee.org/document/9680027

[9] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," *Proc. Annu. Int. Conf. Mob. Comput. Networking, MOBICOM*, pp. 51–56, 2019, doi: 10.1145/3349624.3356768.

[10] K. Boikanyo, A. M. Zungeru, B. Sigweni, A. Yahya, and C. Lebekwe, "Remote patient monitoring systems: Applications, architecture, and challenges," *Sci. African*, vol. 20, 2023, doi: 10.1016/j.sciaf.2023.e01638.

[11] W. Gu, Zhanzhong, He, Xiangjian, Fang, Gengfa, Xu, Chengpei, Xia, Feng. Jia, "MMWAVE radar based HAR for health care monitoring robot.pdf," 2024, *Archive Prefix: arXiv*.

[12] Texas Instruments, "Texas Instruments Official Website," 2024, [Online]. Available: https://www.ti.com/

[13] S. An and U. Y. Ogras, "MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare," *ACM Trans. Embed. Comput. Syst.*, vol. 20, no. 5s, pp. 1–22, 2021, doi: 10.1145/3477003.

[14] K. Mirzaei, M. Arashpour, E. Asadi, H. Masoumi, Y. Bai, and A. Behnood, "3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review," *Adv. Eng. Informatics*, vol. 51, p. 101501, 2022.

[15] A. Gorji, H. U. R. Khalid, A. Bourdoux, and H. Sahli, "On the Generalization and Reliability of Single Radar-Based Human Activity Recognition," *IEEE Access*, vol. 9, pp. 85334–85349, 2021, doi: 10.1109/ACCESS.2021.3088452.

[16] V. C. Chen, F. Li, S. S. Ho, and H. Wechsler, "Micro-doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, 2006, doi: 10.1109/TAES.2006.1603402.

[17] C. Zhou, "Review: mmWave Radar Point Cloud Processing Technology for Human Activity and Posture Recognition," 2023.

[18] A. Prabhakara *et al.*, "High Resolution Point Clouds from mmWave Radar," *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 2023-May, pp. 4135–4142, 2023, doi: 10.1109/ICRA48891.2023.10161429.

[19] I. G. Van De Zande, "3D Point Cloud Object Detection for Millimeter Wave Radar: a Synthesis Study," no. September, 2023.

[20] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 77–85, 2017, doi: 10.1109/CVPR.2017.16.

[21] J. Zhu, H. Chen, and W. Ye, "A Hybrid CNN-LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar," *IEEE Access*, vol. 8, pp. 24713–24720, 2020, doi: 10.1109/ACCESS.2020.2971064.

[22] L. Cao, S. Liang, Z. Zhao, D. Wang, C. Fu, and K. Du, "Human Activity Recognition Method Based on FMCW Radar Sensor with Multi-Domain Feature Attention Fusion Network," *Sensors*, vol. 23, no. 11, 2023, doi: 10.3390/s23115100.

[23] Y. Wang, H. Liu, K. Cui, A. Zhou, W. Li, and H. Ma, "m-Activity : ACCURATE AND REAL-TIME HUMAN ACTIVITY RECOGNITION VIA MILLIMETER WAVE RADAR Beijing University of Posts and Telecommunications , School of Computer Science," no. Dl, pp. 8298–8302, 2021.

[24] B. J. Gluckman, T. I. Netoff, E. J. Neel, W. L. Spano, M. L. Spano, and S. J. Schiff, "Stochastic Resonance in a Neuronal Network from Mammalian Brain," *Phys. Rev. Lett.*, vol. 77, no. 19, pp. 4098–4101, 1996, doi: 10.1103/PhysRevLett.77.4098.

[25] Y. Lin, H. Li, and D. Faccio, "Human Multi-Activities Classification Using mmWave Radar: Feature Fusion in Time-Domain and PCANet,"

in *Sensors*, 2024, pp. 1–24. doi: 10.3390/s24165450.
[26] H. Zhou, Y. Zhao, Y. Liu, S. Lu, X. An, and Q. Liu, "Multi-Sensor Data Fusion and CNN-LSTM Model for Human Activity Recognition System," *Sensors*, vol. 23, no. 10, 2023, doi: 10.3390/s23104750.
[27] H. Khalid *et al.*, "Arsitektur Multi-View CNN-LSTM untuk Pengenalan Aktivitas Manusia Berbasis Radar," pp. 24509–24519, 2022.
[28] A. H. Victoria, S. V Manikanthan, M. A. Wildan, and K. H. Kishore, "International Journal of Communication Networks and Information Security Radar Based Activity Recognition using CNN-LSTM Network Architecture Article History," vol. 14, no. 3, pp. 303–312, 2022, [Online]. Available: https://ijcnis.org
[29] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 4489–4497, 2015, doi: 10.1109/ICCV.2015.510.
[30] Z. Hameed and B. Garcia-Zapirain, "Sentiment Classification Using a Single-Layered BiLSTM Model," *IEEE Access*, vol. 8, pp. 73992–74001, 2020, doi: 10.1109/ACCESS.2020.2988550.
[31] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, 2010.
[32] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
[33] A. Mao, M. Mohri, and Y. Zhong, "Cross-Entropy Loss Functions: Theoretical Analysis and Applications," *Proc. Mach. Learn. Res.*, vol. 202, pp. 23803–23828, 2023.
[34] R. O. Sinnott, F. Wu, and W. Chen, "A Mobile Application for Dog Breed Detection and Recognition Based on Deep Learning," *Proc. - 5th IEEE/ACM Int. Conf. Big Data Comput. Appl. Technol. BDCAT 2018*, pp. 87–96, 2018, doi: 10.1109/BDCAT.2018.00019.
[35] M. Trăscău, M. Nan, and A. M. Florea, "Spatio-temporal features in action recognition using 3D skeletal joints," *Sensors (Switzerland)*, vol. 19, no. 2, 2019, doi: 10.3390/s19020423.

AUTHOR BIOGRAPHY

Raden Rofiq Yudha Setyawan was born in Sleman, Indonesia, in 2001. He received his B.Sc. degree in Computer Engineering from Telkom University, Bandung, Indonesia, in 2023. Currently, he is pursuing his M.Sc. degree in Electrical Engineering at the same university. His research focuses on radar signal processing, deep learning, and human activity recognition using 3D point cloud data from mmWave radar. He also has strong interests in mobile application development, user interface and user experience (UI/UX) design, Internet of Things (IoT) systems, and cybersecurity. Raden has participated in several national programs and certifications, including two talent development programs organized by the Ministry of Education, Culture, Research, and Technology of Indonesia. He has also been involved in various software development projects related to Android applications and data-driven systems. His current research explores the integration of deep learning models with sensor-based human activity recognition systems for smart healthcare environments.

Fiky Y. Suratman was born in Jakarta, Indonesia, in 1976. He received the bachelor's degree in engineering physics and the master's degree from the School of Electrical and Informatics (STEI), Institut Teknologi Bandung, Bandung, Indonesia, in 1998 and 2006, respectively, and the Dr. Ing. degree from the Technische Universitaet Darmstadt (TU-Darmstadt), Darmstadt, Germany, in 2014.,Early placement in industry (Astra Microtronics Technology, Batam, Indonesia), from 1998 to 2001, he was then followed by lectureship at Universitas Komputer Indonesia (UNIKOM), Bandung. In 2007, he joined the Faculty of Electrical Engineering, Telkom University, Bandung, as a Faculty Member. He has been the Head of the Department of Electrical Engineering, Telkom University, since 2014. His research interests include statistical signal processing, radar signal processing, and machine learning.,Dr. Suratman is currently a member of the IEEE Signal Processing Society and Communications Society. He is also the Chairman of the Signal Processing Society (SPS)'s Indonesia Chapter. He was awarded the scholarship from the German Academic Exchange Service (DAAD) to continue his Dr. Ing. degree.

Khilda Affah was born in Karawang, Indonesia. She received her B.Ed. in Electrical Engineering Education from Universitas Pendidikan Indonesia (UPI) in 2014, followed by an M.Sc. in Microelectronics Engineering from Institut Teknologi Bandung (ITB) in 2017. She earned her Ph.D. in the Department of Functional Control Systems from Shibaura Institute of Technology (SIT), Japan, under the JICA Innovative Asia Scholarship. Since 2021, she has been a lecturer in Electrical Engineering at Telkom University. Her research interests include analog integrated circuits, engineering education, switched-capacitor circuits, smart card systems, smart home technologies, the Internet of Things (IoT), and applications in biomedical engineering