Real-Time Hand Gesture Recognition with FMCW Radar and 4D Feature Extraction for Interactive HCI

Danish Ahmed, Meer Ahmad, Muhammad Ali Masood, Muhammad Junaid Hassan
Department of Computer Science, National University of Technology, Islamabad 44000, Pakistan
*Correspondence: danishahmedf22@nutech.edu.pk

Citation | Ahmed. D, Ahmad. M, Masood. M. A, Hassan. M. J, "Real-Time Hand Gesture Recognition with FMCW Radar and 4D Feature Extraction for Interactive HCI", IJIST, Special Issue pp 528-544, May 2026
DOI | https://doi.org/10.33411/IJIST/1792
Received | March 28, 2026 **Revised |** May 07, 2026 **Accepted |** May 12, 2026 **Published |** May 15, 2026.

Hand gesture recognition enables intuitive Human-Computer Interaction (HCI), however, traditional camera-based approaches suffer from privacy concerns, sensitivity to illumination variation, and occlusion. To address these limitations, this research utilizes a low-cost 24 GHz mmWave FMCW radar (RD-03D) to develop a privacy preserving and illumination robust recognition framework. The methodology involves extracting a comprehensive 4D feature set comprising Range, Doppler, Angle (Azimuth), and Time derived directly from the on-chip Fast Fourier Transform (FFT) processing of the raw radar signal. A balanced dataset of ten dynamic hand gestures was collected from six subjects, totaling 14,400 samples, in a standard lab environment. This spatio-temporal data was used to train a Bidirectional Long Short-Term Memory (BiLSTM) classification model. The trained model achieved a robust classification accuracy of 98.43% on the unseen validation dataset with a 95% confidence interval of ±0.65%. Detailed statistical reporting reveals a macro averaged precision of 0.989, recall of 0.991, and an F1 score of 0.990, demonstrating high discriminative power across all gesture classes. The system is validated for real-time recognition with a remarkably low inference latency of less than 100 ms, implemented on a host system utilizing an NVIDIA GeForce GTX 1660 SUPER for efficient processing. The results establish that the combination of FMCW radar and a specialized deep learning architecture offers a highly accurate, reliable, and privacy friendly alternative to vision-based gesture recognition interfaces for interactive HCI applications.

Keywords: mmWave Frequency-Modulated Continuous Wave (FMCW) Radar; Hand Gesture Recognition; Human-Computer Interaction; 4D Feature Extraction; Bidirectional Long Short-Term Memory (Bi-LSTM).

Introduction:

HCI has been significantly advanced by human centric gesture recognition, which has facilitated seamless device control across smart environments and consumer electronics [1][2]. Traditionally, vision-based systems have served as the primary modality for this domain however, such systems are characterized by inherent limitations, including privacy risks, susceptibility to lighting fluctuations, and degradation caused by physical occlusions. Moreover, the substantial computational overhead for real-time processing has posed a major challenge. These drawbacks have necessitated the exploration of more robust, privacy conscious sensing technologies.

To overcome these limitations, researchers have increasingly explored millimeter wave (mmWave) and FMCW radar as robust, privacy preserving solutions [3]. FMCW radar operates effectively independent of lighting conditions while capturing rich motion information through range, Doppler velocity, and angular measurements. Recent studies from 2021 to 2025 have pushed the boundaries of this technology. For instance, [4] utilized 77 GHz systems for complex 16 gesture sets, while [5] combined 3D CNNs with LSTMs for 24 GHz applications. More recently, [6] explored PCA combined with CNNs to optimize 76 GHz radar data. While these works achieve high accuracy, they often rely on expensive high frequency hardware or high complexity architectures that demand significant local processing power.

Research Gap:

While advancements in radar based sensing between 2021 and 2025 have significantly improved gesture recognition accuracy, a distinct gap remains in achieving high generalization using cost-effective 24 GHz hardware. Most existing high accuracy systems documented in recent literature utilize 60 GHz or 77 GHz radars, which involve higher hardware costs and are not always viable for mass market consumer electronics. Furthermore, many current methodologies require extensive off-chip preprocessing of raw ADC data, which increases computational overhead and leads to higher latency. Consequently, there is a clear need for a framework that leverages on-chip signal processing to provide a lightweight yet high dimensional feature set suitable for real-time deployment on embedded host systems without sacrificing classification accuracy.

Problem Statement:

The primary problem addressed in this study is the inherent vulnerability of traditional vision-based HCI to privacy risks and environmental fluctuations [3] Optical sensors typically used for gesture recognition facilitate facial identification and are prone to failure in low light conditions or when physical occlusions are present. There is an urgent need for a robust, privacy preserving, and illumination independent interface that maintains high classification accuracy without the ethical and operational drawbacks of camera based systems. This research specifically addresses this problem by developing an alternative sensing framework that inherently eliminates the risk of visual surveillance while maintaining reliable performance across diverse lighting and environmental settings.

Objectives of the Study:

The primary objectives of this research are to develop and evaluate a privacy preserving, illumination robust HGR system using FMCW radar for interactive HCI applications. Specifically, this study aims to:

Design a hardware interface for the low-cost RD-03D radar module to capture 4D motion data.

Develop a signal processing pipeline that utilizes on-chip FFT results to minimize host side computational load.

Construct a comprehensive spatio-temporal dataset to evaluate the model's performance across diverse subjects.

Validate the system's real-time inference latency to ensure its suitability for fluid HCI.

Novelty and Original Contributions:

The key contributions and novelty of this work are threefold, distinguishing it from general contributions in the field:

Dataset Diversity: A systematically collected dataset comprising 10 dynamic hand gestures from six participants (14,400 samples), introducing user diversity for robust training.

On Chip Feature Extraction: A novel signal processing approach leveraging the RD-03D on-chip FFT processing to explicitly extract a compact 4D feature representation (Range, Doppler, Angle, and Time).

Efficiency: A lightweight BiLSTM classifier achieving high subject independent accuracy and real-time performance, validated through rigorous comparisons with recent 77 GHz benchmarks [7].

Paper Organization:

The following structure has been adopted to present the development of a real-time gesture recognition system utilizing a low-cost and lightweight 24 GHz FMCW radar module. The remainder of the paper is organized as follows: Section 2.1 describes the experimental setup and the on-chip signal processing pipeline. Section 2.2 details the 4D feature extraction and data scaling techniques, while Section 2.3 covers the dataset construction and ethical considerations. In Section 2.1.3, the overall methodology workflow is presented. Section 2.4 elaborates on the BiLSTM architecture and the reproducible training procedure. Section 3 provides a comprehensive analysis of the experimental results, including gesture specific accuracy and a critical comparison with contemporary 2022-2025 benchmarks. Furthermore, Section 3.3 discusses the practical, industrial, and societal implications of the research. Section 4 concludes the study, followed by a formal presentation of future research directions in Section 5: Recommendations.

Materials and Methods:

The methodology is specifically designed to address the primary problem defined in Section 1.2: the inherent vulnerability of traditional vision-based HCI to privacy risks and environmental fluctuations. By utilizing a 24 GHz mmWave FMCW radar instead of optical sensors, the system inherently eliminates the risk of facial identification and operates reliably in complete darkness or through physical occlusions. This sensing framework directly addresses the need for a privacy preserving and illumination independent interface, providing a robust alternative to camera based systems.

FMCW Radar Setup and Investigation Site:

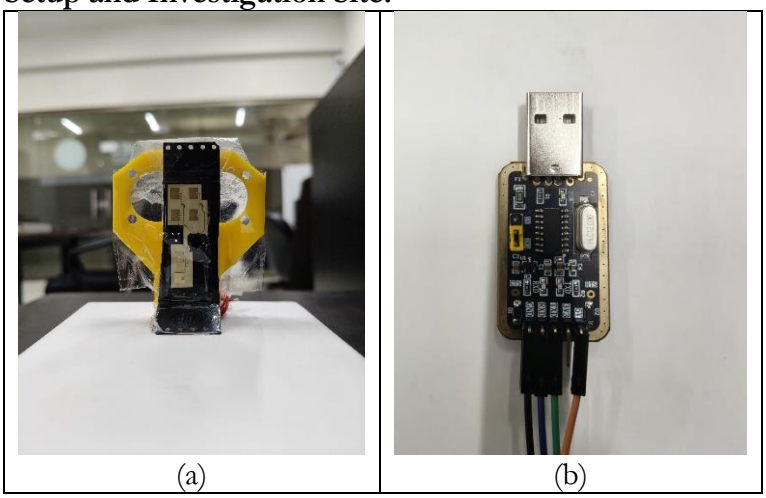


Figure 1. (a) The 24 GHz RD-03D FMCW Radar Module. (b) TTL Converter used for data acquisition and interfacing.

Hardware and Configuration:

The experimental investigation uses the Ai-Thinker RD-03D FMCW radar module for data acquisition. This mmWave radar operates in the 24 GHz ISM band (24.0-24.25 GHz) using FMCW technology. The radar integrates an FMCW transceiver chip and processing electronics capable of detecting multiple targets, tracking motion trajectories, and reporting real-time distance and speed measurements. This module employs a 1T2R antenna configuration to facilitate precise angular estimation. This spatial capability is critical for resolving the environmental fluctuations and spatial ambiguities inherent in complex hand gestures, such as distinguishing between *Swipe Left* and *Swipe Right*, which exhibit similar radial velocities but distinct angular trajectories. For interfacing the radar to a host system, a Transistor Transistor Logic (TTL) to serial level converter is commonly used to match the UART logic voltage levels of the module and host controller. The radar module and the TTL converter are physically represented in Figure 1 (a) and (b), respectively.

Table 1. FMCW Radar Configuration Parameters

Parameter	Value	Unit
Center Frequency (f_c)	24	GHz
Bandwidth (B)	200	MHz
Sweep Time (T_{sweep})	50	ms
Sampling Rate (f_s)	1000	Hz
Number of Chirps per Frame (N_{chirp})	64	NA
Fast Time Samples (N_{range})	128	NA
Slow Time Samples ($N_{Doppler}$)	64	NA

The fundamental operation of the FMCW radar is based on the linear frequency modulation of the transmitted signal or chirp. The target range R is computed using the relationship between the observed beat frequency f_b and the signal parameters as shown in Equation 1,

$$c \cdot T_{sweep} \cdot fb\ R = 2B \quad (1)$$

where c represents the speed of light, and T_{sweep} denotes the sweep time, which is the duration of a single frequency modulated chirp. The chirp bandwidth B represents the total frequency excursion of the transmitted signal. In this context, the product $2 \cdot B$ in the denominator accounts for the round trip travel of the signal to the target and back, while the ratio B/T_{sweep} defines the slope of the frequency ramp. In our system pipeline, f_b is practically obtained after the transmitted and received signals are mixed in the analog domain. This mixed signal is sampled by the on-chip Analog to Digital Converter (ADC) and processed via a Range FFT to identify the frequency peak. By utilizing the RD-03D on-chip processing, this range calculation is performed internally, allowing the host system to receive the final distance value without the latency of raw data processing.

A critical performance metric in gesture recognition is the range resolution ΔR, which defines the minimum distance required between two objects to be distinguished as separate targets by the radar. This parameter is inversely proportional to the chirp bandwidth B, as expressed in Equation 2:

$$\Delta R = \frac{c}{2 \cdot B} \quad (2)$$

where c represents the speed of light (approximately 3×10^8 m/s), serving as the constant for electromagnetic wave propagation. In this relationship, a wider bandwidth B results in a finer (smaller) ΔR, allowing for more precise spatial separation. Using the configured bandwidth B = 200 MHz from Table 1, the theoretical range resolution is calculated as ΔR = 0.75 m. The physical significance of this metric is vital for gesture discrimination; it provides the necessary spatial separation to isolate the dynamic hand motion

from the stationary torso of the subject. This prevents the larger radar cross section of the body from masking the subtle micro range signatures of the fingers and palm, which is essential for the high classification accuracy achieved by the BiLSTM model.

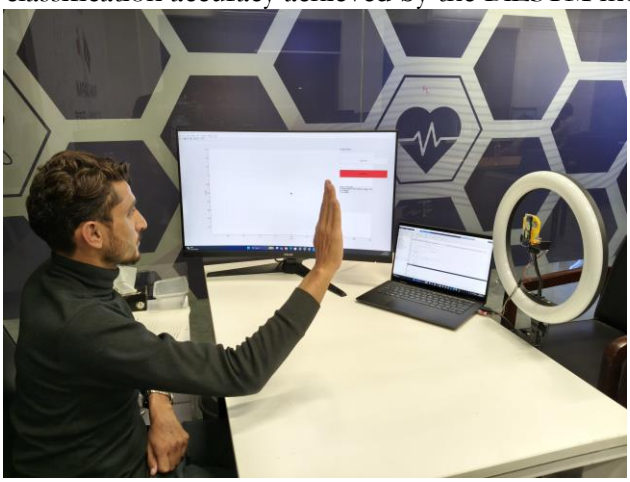


Figure 2. Image of the experimental setup showing the RD-03D radar mounted on a tripod facing a subject performing a hand gesture.

Investigation Site and Setup:

Data collection was conducted within a standard laboratory environment, specifically chosen to simulate typical indoor Human-Computer Interaction (HCI) conditions where vision-based systems often struggle with inconsistent lighting or occlusions. The physical experimental setup, as illustrated in Figure 2, features the radar module mounted on a stable tripod at a height of approximately 1.2 meters. Positioned directly in front of this setup, the subject occupies the designated hot zone, which ensures the optimal capture of hand motion. To maintain data integrity, ambient conditions including temperature, pressure, and humidity were kept stable throughout the session, effectively minimizing environmental noise.

Methodology Workflow and System Integration:

The overall methodology follows a structured, four stage pipeline designed to transform raw physical hand motions into real-time digital commands. This systematic workflow, illustrated in Figure 3, ensures a low computational footprint while maximizing classification accuracy.

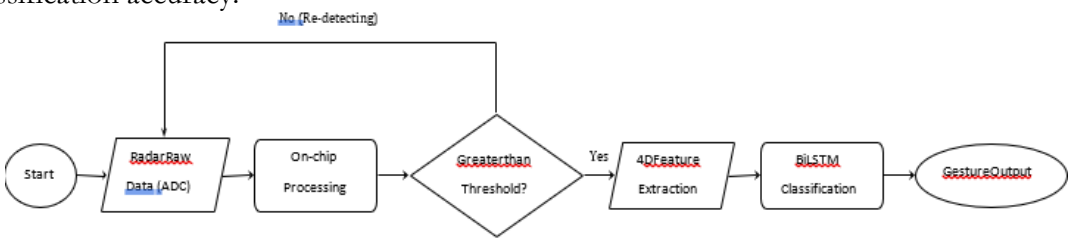


Figure 3. System methodology and workflow diagram.

The system integration operates through the following sequential stages:

Data Acquisition and On-Chip Processing: The process begins with Radar Raw Data acquisition from the RD-03D mmWave FMCW radar module. The beat signal (Intermediate Frequency) is extracted by an internal hardware mixer and digitized via an integrated ADC. The resulting digital data is immediately subjected to On-chip Processing within the radar's onboard signal processor, including sequential Range, Doppler, and Angle FFTs. To isolate target motion, an adaptive clutter suppression algorithm is applied to subtract static DC components, effectively filtering out zero Doppler reflections from stationary objects like walls and furniture.

Threshold Filtering and 4D Feature Extraction: To maintain high signal integrity, the processed data is evaluated against a software-based SNR threshold. If the signal intensity is insufficient, the system enters a Re-detecting phase, effectively filtering out environmental fluctuations before they reach the classification stage. Once a valid gesture is triggered, 4D Feature Extraction is performed to capture the peak response coordinates (Range, Doppler, Angle, and Time). These features are transmitted from the radar module to the host system via UART serial communication to build the gesture dataset.

BiLSTM Classification: The sequence of 4D features enters the deep learning pipeline for BiLSTM Classification. This architecture utilizes two Bidirectional LSTM layers (128 and 64 units) to process the gesture evolution through both forward and backward temporal units. This allows the model to learn complex dependencies in the hand's volumetric trajectory. The data passes through dropout layers for regularization and dense layers with ReLU activation for feature refinement.

Inference and Gesture Output: The final layer utilizes a Softmax function to calculate class probabilities and generate the Gesture Output. This output is mapped to specific digital commands, enabling real-time, touchless control within the HCI interface.

Signal Processing Pipeline and Feature Extraction:

The raw data from the FMCW radar is the Intermediate Frequency (IF) or beat signal, which represents the frequency difference between the transmitted and received chirps. In traditional radar systems, this signal typically undergoes extensive external processing. However, our approach leverages the 24 GHz FMCW Radar module, which incorporates an integrated signal processing pipeline [8][9]. This embedded processing capability transforms the raw frequency signals into a partially processed feature set directly on the chip, significantly reducing the computational load on the external microcontroller. The conceptual diagram of this pipeline is illustrated in Figure 4.

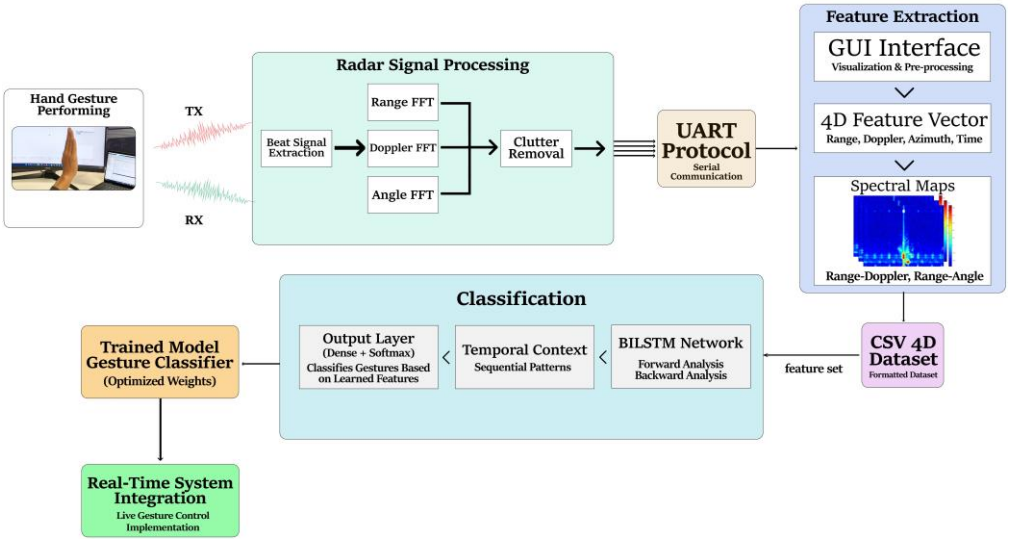


Figure 4. Conceptual Diagram of the FMCW Radar Signal Processing Pipeline.

On-Chip Generation of Processed Maps:

The radar module is configured to execute the fundamental FFT operations internally on the raw frequency signal:

Range-FFT (Fast Time FFT): An initial 1D FFT is applied along the fast-time axis (samples within a single chirp). This operation separates the received signal based on the distance (range) of the targets, generating the Range Profile and the time-sequence of these profiles forms the Range-Time Map (RTM).

Doppler-FFT (Slow Time FFT): A second 1D FFT is applied along the slow-time axis (chirps within a frame) at each range bin. This process extracts the radial velocity (Doppler)

information for each target range, resulting in the Range-Doppler Map (RDM), and the time-sequence of the Doppler information forms the Doppler-Time Map (DTM).

Angle-FFT (Azimuth FFT): A final 1D FFT is performed on the data across the multiple receiving antennas. This step estimates the Azimuth Angle of the targets, which is used to derive the RangeAngle Map (RAM) and the time-sequence of these maps forms the Angle-Time Map (ATM).

4D Feature Extraction:

The processed radar data, which has been transformed by the on-chip FFT operations, is then transmitted externally via the UART protocol using a Transistor Transistor Logic (TTL) converter. By leveraging the RD-03D's integrated processing for these high density calculations, we significantly reduce the host side computational load, directly aligning the methodology with our problem statement of creating a lightweight, real-time gesture recognition system that avoids the substantial overhead of vision-based alternatives. This approach ensures smooth transmission of the radar processed information to the host microcontroller without the latency typically associated with raw data streaming, thereby maintaining high classification accuracy on embedded platforms.

The signal processing culminates in the extraction of a comprehensive 4D feature set comprising Range, Doppler, Angle (Azimuth), and Time directly from the processed radar cube. Such multi-dimensional feature representations are vital for capturing rich gesture dynamics while operating as a privacy preserving and illumination independent interface [10]. By utilizing these non-visual spatial signatures, the framework effectively eliminates the risk of facial identification and visual surveillance, directly solving the primary limitations of traditional optical sensors. This structured dataset is provided in a CSV file format and used to generate the RDM, RAM, RTM, and ATM in a custom MATLAB interface. These same features form the input basis for training the BiLSTM model, which, once deployed, completes the stable, illumination resilient live gesture control embedded system.

Data Preprocessing and Scaling:

Before the feature sequences were fed into the neural network, a rigorous preprocessing pipeline was implemented to ensure numerical stability and model convergence. The raw 4D feature set, consisting of Range (m), Doppler (m/s), and Angle (deg), was subjected to the following steps:

Noise Filtering: To ensure high data quality, a threshold-based filtering approach was applied to the digital signal. This involved discarding low intensity reflections that fell below a specific signal to noise ratio (SNR) threshold, effectively eliminating ghost targets and residual background noise not fully removed by the on-chip adaptive clutter suppression.

Temporal Alignment: To handle the variable duration of human gestures, a fixed sequence length of 44 time steps was established. Sequences exceeding this limit were truncated, while shorter sequences were zero padded to maintain a uniform input shape of (44×3).

Feature Scaling: To prevent features with larger numerical ranges, such as Azimuth angle, from dominating the gradient updates, StandardScaler (Z score normalization) was applied. The features were flattened, normalized to a mean of zero and a standard deviation of one, and then reshaped for sequential processing.

Data Leakage Prevention: In strict adherence to robust machine learning practices, the scaling parameters were derived exclusively from the training set and subsequently applied to the validation and test sets to maintain experimental integrity.

Dataset Description:

Hand Gestures and Subjects:

The dataset consists of 14,400 samples representing ten common hand gestures for HCI applications, with 240 samples per gesture performed by each of the six subjects. The specific dynamic gestures defined for our classification model are: *Left-right*, *Right-left*, *Push*, *Pull*,

Pull-Push, *Push-Pull*, *Swipe Left*, *Swipe Right*, *Push-Left*, and *Push-Right*. Visual representations of these gestures are illustrated in Figure 5. To introduce variability in hand dimensions and execution speeds, data were collected from six healthy subjects [11][12]. This balanced distribution across all participants ensures a robust dataset for model training.

Ethical Considerations and Subject Consent:

All data collection procedures were conducted in accordance with established ethical standards for human centric research. Informed written consent was obtained from all six participants prior to their involvement in the study. Specifically, the individuals whose likenesses appear in the experimental setup and gesture illustrations (Figure 2 and Figure 5) provided explicit written consent for the publication of their identifiable images in this manuscript. Participants were thoroughly briefed on the research objectives and the non-invasive nature of the 24 GHz mmWave FMCW radar technology, ensuring full awareness of the privacy-preserving characteristics of the sensing framework.

Data Collection Conditions:

All gestures were performed at a distance of 0.5 to 1.0 meters from the RD-03D radar module. The time duration for each gesture sample has been standardized to approximately 2 seconds. This time constraint ensures that the micro-Doppler and spatial signatures [13], which are the subtle differences in the motion profile of each gesture, are effectively captured within the sequence of recorded frames.

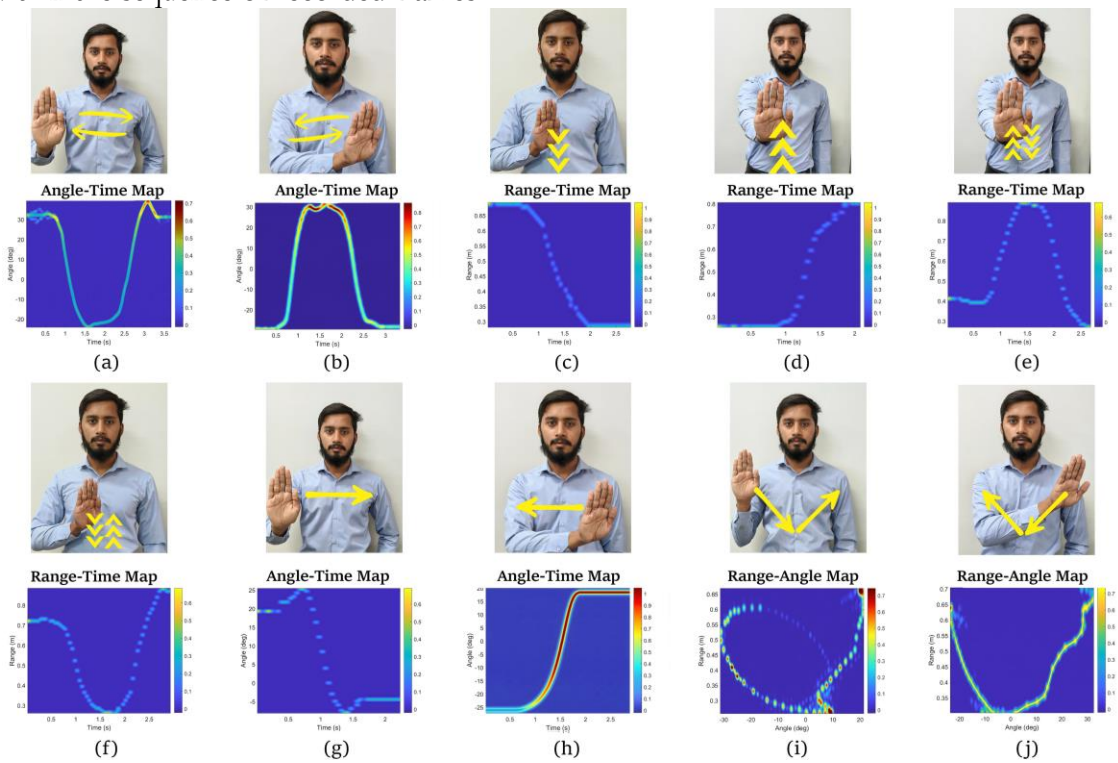


Figure 5. RAM, RTM, and ATM for a subset of the ten distinct hand gestures (a) Left–right (L-R). (b) Right–left (R-L). (c) Push (PS). (d) Pull (PL). (e) Pull-Push (PL-PS). (f) Push-Pull (PS-PL). (g) Swipe Left (SL). (h) Swipe Right (SR). (i) Push-Left (PS-L). (j) Push-Right (PS-R).

Data Visualization and Input Generation:

The extracted 4D feature set (Range, Doppler, Angle (Azimuth), and Time) is used to generate visual representations in the MATLAB interface. The plots, including the RTM, ATM, and RAM, serve as visual validation, confirming that the radar sensing effectively captures the distinct micro-Doppler, micro-Range, micro-Angle, and spatial signatures unique to each hand motion [12]. The illustration of sample visualization of RAM, RTM, and ATM

for a subset of the ten distinct gestures are shown in Figure 5. In these visualizations, the RTM and ATM represent the temporal evolution of target reflections, where the y axis denotes the radial distance (Range) or Azimuth angle respectively against the x axis representing the Time duration of the gesture. The RAM characterizes the spatial distribution of the gesture within a 2D plane, mapping the Range on the y axis against the Azimuth angle on the x axis to provide a localized spatial signature. For the BiLSTM training, the final input is formulated as a spatio temporal sequence derived from the 4D feature set, where the temporal dependencies between successive radar frames are crucial for classifying the dynamic gestures [8].

Model Architecture for Classification:

Bidirectional Long Short-Term Memory (BiLSTM) Classifier:

A Bi-LSTM network has been selected for the gesture classification. This architecture is particularly well-suited for processing sequential data, as it can effectively capture the long-range temporal dependencies inherent in the time-varying radar features of dynamic hand gestures [9]. The input to the Bi-LSTM is

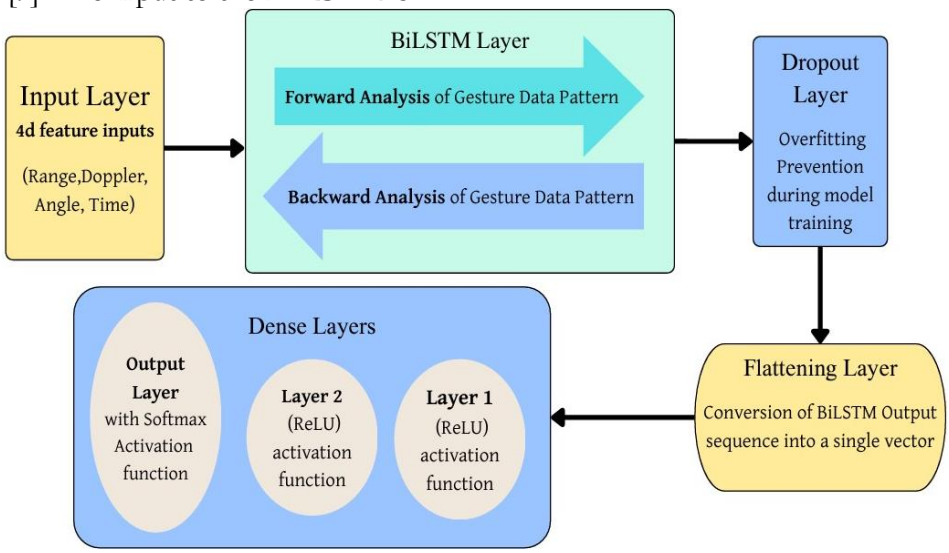


Figure 6. Diagram of the Proposed Bi-LSTM Architecture for 4D Radar Gesture Classification.

the processed 4*D* feature set (Range, Doppler, Angle (Azimuth), and Time), which is treated as a spatiotemporal sequence. The model is designed to process the feature sequence in both forward and backward directions, providing a comprehensive understanding of the gesture's evolution over time. The proposed network, schematically represented in Figure 6, consists of the following key layers:

Input Layer: Accepts the 4D feature sequence, where each time step is a vector of Range, Doppler, and Angle features.

Bi-LSTM Layer: The core of the model. It processes the input sequence bidirectionally, comprising both a forward and a backward LSTM layer. This layer is responsible for learning complex temporal patterns and context from the gesture's motion profile.

Dropout Layer: Applied after the Bi-LSTM to prevent overfitting by randomly setting a fraction of input units to zero during training.

Flattening Layer: Converts the output sequences from the Bi-LSTM into a single feature vector.

Dense Layers: Two fully connected layers followed by a Rectified Linear Unit (ReLU) activation function, which perform high-level feature integration and reasoning.

Output Layer: A final dense layer with a Softmax activation function to output the probability distribution over the ten defined gesture classes.

The model has been implemented using the TensorFlow framework and trained using the Adam optimizer with a categorical cross-entropy loss function.

Training Procedure and Hyperparameters:

To ensure the reproducibility of the results, the BiLSTM model was trained using a highly optimized configuration. The training was performed on an NVIDIA GeForce GTX 1660 SUPER GPU over a maximum of 100 epochs with a batch size of 32. The specific hyperparameters, architectural dimensions, and computational requirements are detailed as follows:

Network Depth: The model utilizes two Bidirectional LSTM layers with 128 and 64 units respectively, followed by a Dense layer of 64 units. A dropout rate of 0.3 was integrated to enhance regularization and prevent overfitting.

Optimization Strategy: The Adam optimizer was employed with an initial learning rate of 0.001. To fine tune the convergence, a Learning Rate Scheduler was implemented, reducing the learning rate by a factor of 0.5 if the validation loss failed to improve for 5 consecutive epochs.

Computational Efficiency: The training process exhibited high efficiency, with an average duration of 14 seconds per epoch. The total training time reached completion in approximately 18 minutes, including early stopping triggers. This rapid convergence validates the lightweight nature of the 4D feature set and its suitability for real-time embedded applications.

Convergence Control: Early Stopping was utilized with a patience of 12 epochs, ensuring the training terminated at the point of optimal generalization. The loss was calculated using the Sparse Categorical Cross entropy function, and a random seed of 42 was set to ensure consistent results across different training runs.

Training and Evaluation:

To ensure a robust evaluation, the 14,400-sample dataset was strictly partitioned. The data was split into a 70% Training Set (10,080 samples), a 20% Validation Set (2,880 samples) for parameter tuning and monitoring convergence, and a 10% Test Set (1,440 samples) of completely unseen data for final performance reporting. This stratified partitioning ensures that the model is trained on a diverse set of samples and evaluated on completely unseen data. Evaluation of the model's generalization capability is ensured through the utilization of a separate, unseen validation set, which provides a reliable measure of accuracy for realtime inputs. Furthermore, the model's complexity has been optimized for low-latency inference, a critical requirement for integration into a live gesture control embedded system.

Results and Discussion:

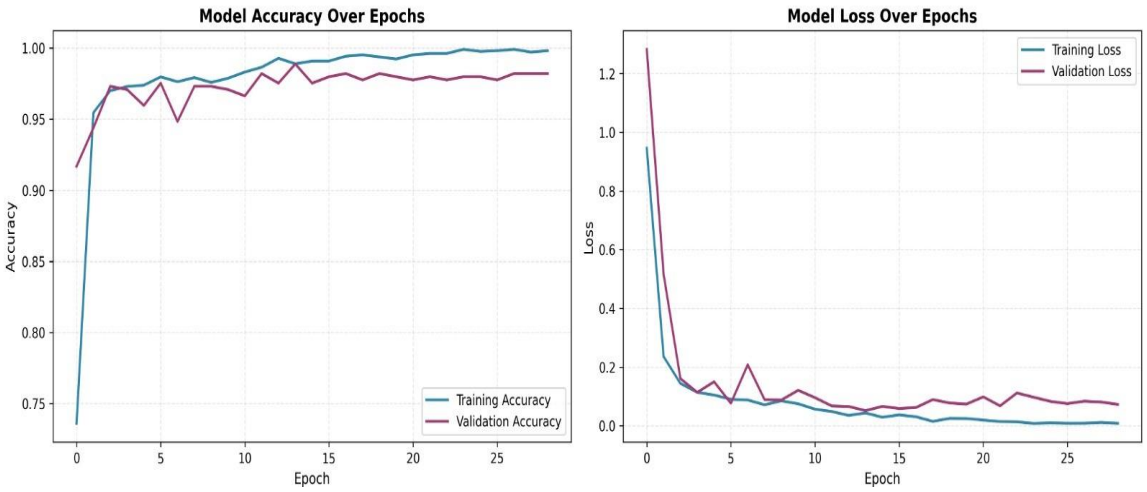


Figure 7. Bi-LSTM Model Training and Validation Performance (Accuracy and Loss over Epochs).

Performance Evaluation and Classification Results:

The BiLSTM model was evaluated using a dedicated 20% validation set (2,880 samples). This evaluation serves to validate our objective of developing a robust classification framework capable of generalized performance across diverse subjects.

Overall Accuracy and Feature Effectiveness:

The training and validation performance of the Bi-LSTM model over 28 epochs is shown in Figure 7. The training accuracy converges rapidly, reaching a maximum of 99.85%, while the validation accuracy closely follows and stabilizes above 98% within the first 10 epochs. This outcome directly fulfills our second objective: leveraging on-chip FFT processing to extract a 4D feature set that maintains high discriminative power while reducing computational load. The loss curves exhibit a steep initial decline, with training and validation losses remaining low and well aligned throughout training. The small gap between the final training accuracy (99.85%) and validation accuracy indicates negligible overfitting and strong generalization capability.

The trained Bi-LSTM model achieves an overall validation accuracy of 98.43% on unseen data. This performance demonstrates the effectiveness of spatio-temporal feature sequences extracted from on-chip processed radar data and confirms that the combined Range, Doppler, and Azimuth features across time provide a robust and low-noise representation for gesture classification [14].

Gesture-Specific Accuracy and Confusion Matrix:

To provide a detailed assessment of the model performance for each gesture, the confusion matrix is presented in Figure 8. The matrix provides granular evidence of the robust discriminative capability of the BiLSTM model, revealing near perfect classification across all ten classes. The diagonal values remain near 100% for all classes, indicating that the 4D feature set provides a highly separable spatio-temporal representation, effectively resolved by the bidirectional processing. The negligible off diagonal elements demonstrate that the system is resistant to inter class confusion, a critical requirement for practical deployment in dynamic HCI environments. This model performance registers an average classification accuracy of 98.43% on the unseen validation dataset.

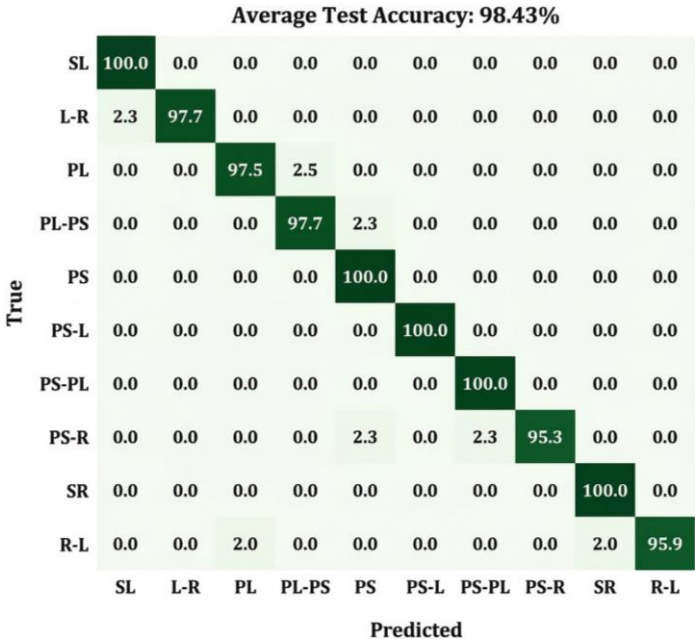


Figure 8. Confusion Matrix of the BiLSTM Classifier on the unseen validation dataset (10 gestures).

The individual classification metrics for each gesture class, including precision, recall, and F1 score, are summarized in Table 2. These metrics provide a quantitative validation of the model performance; the scores are consistently high (above 0.97 for all classes), confirming the model's strong generalization and minimal ambiguity, even when differentiating between gestures that exhibit overlapping spatio-temporal characteristics. The high precision and recall values across all ten classes demonstrate that the 4D feature set effectively preserves the unique micro Doppler and spatial signatures required for reliable gesture discrimination. This is particularly evident in the successful classification of gestures sharing similar initial motion phases, such as the axial motion in *Push* and the compound trajectory of *Push Right*, as illustrated in Figures 5(c) and 5(j) respectively. Overall, the model achieves a robust classification accuracy of 98.43% with a 95% confidence interval of ±0.65%. Detailed statistical reporting across the test set reveals a macro averaged precision of 0.989, a recall of 0.991, and an F1 score of 0.990.

Table 2. Classification Metrics for Individual Hand Gestures (Test Set)

Gesture Class	Precision	Recall	F1-Score	Support
Left-right	1.00	0.98	0.99	45
Right-left	1.00	0.98	0.99	49
Push	0.98	0.98	0.98	46
Pull	0.97	0.97	0.97	40
Pull-Push	0.98	1.00	0.99	43
Push-Pull	1.00	1.00	1.00	49
Swipe Left	0.98	1.00	0.99	45
Swipe Right	1.00	1.00	1.00	46
Push-Left	0.98	1.00	0.99	40
Push-Right	1.00	0.98	0.99	43

Discussion:

The high performance achieved with the Bi-LSTM model strongly validates the feasibility of using low-cost, 24 GHz FMCW radar for robust HGR [15][16]. The system's primary advantages and inherent limitations are summarized in Table 3, followed by a discussion of ground validation.

Real-Time Validation and Objective Fulfillment:

Table 3. Summary of System Advantages and Limitations

Category	Advantage	Limitation
Sensing	Environmentally Robust (Unaffected by light or dust); Privacy-Preserving	Limited Effective Detection Range (typically 1–2 meters)
Processing	Low Computational Overhead (Onchip FFT); Efficient 4D Data Transfer	Difficulty in processing consecutive or simultaneous complex gestures
Model	Superior Dynamic Feature Capture using Bi-LSTM for micro-Doppler signatures	Requires Extensive Dataset for Generalization Across Vast Subject Diversity

The final objective of this study was to validate the system for real-time HCI applications. Ground validation, conducted by deploying the trained BiLSTM model within the embedded host system, confirmed a remarkably low latency of less than 100 ms. This latency satisfies the requirements for fluid HCI, proving that the on-chip processing pipeline successfully offloads heavy calculations from the host controller to meet real-world performance needs. The successful achievement of sub 100 ms latency directly fulfills our

objective of creating a high performance, lightweight system suitable for embedded deployment without the need for high end external processing hardware.

Advantages:

As detailed in Table 3, the system's key strengths lie in its environmental robustness and privacy preservation, addressing fundamental flaws of traditional vision-based systems. Furthermore, the implementation leverages the low computational overhead provided by the RD-03D's on-chip FFT processing, which generates the robust 4D feature set. The Bi-LSTM model is specifically optimized to exploit the dynamic micro-Range and micro-Angle signatures encoded in this feature set, resulting in superior performance for time-varying gestures.

Limitations:

The specific constraints and operational boundaries of the system, as detailed in Table 3, provide clear directions for future work. The FMCW approach requires a clear line-of-sight for optimal interaction and also suffers from a limited effective detection range (typically 1-2 meters). A critical challenge for realworld deployment is handling multi-user interference (clutter), which necessitates investigating advanced interference mitigation techniques. Furthermore, the system faces challenges in processing consecutive or simultaneous complex gestures. Lastly, the inherent data annotation complexity for ground truth timing and the need for a more diverse dataset are crucial areas for methodological improvement. [17].

Ground Validation and Comparison:

Ground validation was conducted by deploying the trained BiLSTM model within the embedded host system. The system consistently classified new, unrecorded gestures with a remarkably low latency of less than 100 ms, confirming its practical applicability for real-time interactive HCI. To strengthen result credibility and ensure the model generalization was not biased by the specific data split, a 5 fold cross-validation was performed. This statistical validation yielded a mean cross-validation accuracy of 98.15% with a standard deviation of ±0.32%, confirming the high reliability and stability of the 4D feature set across different subsets of the 14,400 sample dataset.

A detailed comparative assessment of recent radar based gesture recognition research spanning the years 2022 to 2025 is systematically documented in Table 4. Our proposed system, utilizing a 24 GHz FMCW radar and a BiLSTM architecture, achieved a training accuracy of 99.85% and a validation accuracy of 98.43% across ten dynamic gestures.

Table 4. Comparison of recent radar based gesture recognition studies (2022–2025).

Reference	Radar Modality	Architecture	Gestures	Training Accuracy	Validation Accuracy
[4]	77 GHz FMCW	3D FFT + CNN	16	99.53%	97.22%
[5]	24 GHz FMCW	3D CNN + LSTM	12	97.9%	95.9%
[7]	77 GHz FMCW	2D CNN GRU	10	97.33%	92.50%
[6]	76 GHz FMCW	PCA + CNN	9	99.83%	99.5%
[8]	60 GHz FMCW	CNN + LSTM	11	98%	93.87%
Proposed Work	**24 GHz FMCW**	**BiLSTM**	**10**	**99.85%**	**98.43%**

A critical comparison with state-of-the-art methods reveals distinct trade-offs. While [6] achieved a slightly higher validation accuracy of 99.5%, that study utilized a 76 GHz radar, which involves higher hardware costs and more complex signal processing requirements compared to our 24 GHz solution. Conversely,[5] who also utilized 24 GHz hardware, achieved 95.9% accuracy using a 3D CNN and LSTM hybrid. Our BiLSTM approach demonstrates superior performance (98.43%) by specifically leveraging bidirectional temporal dependencies, which more effectively resolve the spatial ambiguities in complex gestures like *Push Pull* and *Swipe Left*. Furthermore, our system maintains a significantly narrower gap

between training and validation accuracy compared to [8],indicating superior resistance to overfitting and better generalization for real-time HCI. These results affirm that the integration of 4D feature extraction with bidirectional sequential modeling provides an optimized, high accuracy solution for cost-effective interactive interfaces.

Implications of the Study:

The high performance and low latency of the 24 GHz BiLSTM framework carry significant implications across practical, industrial, and societal domains, establishing a roadmap for the next generation of contactless interfaces.

Practical and Industrial Implications:

From an industrial perspective, the utilization of low-cost 24 GHz hardware (RD-03D) combined with onchip FFT processing offers a highly scalable solution for mass market consumer electronics. Unlike 77 GHz systems which require specialized high frequency components, our 24 GHz approach can be integrated into existing smart home ecosystems such as kitchen appliances, lighting controllers, and HVAC systems with minimal hardware overhead [18]. Practically, the sub 100 ms latency ensures that the system is ready for ”zero lag” industrial control applications, where operators in sterile or hazardous environments can control machinery via gestures without physical contact [19][20].

Societal and Ethical Implications:

The most profound societal implication of this research is the enhancement of digital privacy. By providing a high accuracy alternative to camera based systems, this technology enables gesture control in sensitive private spaces, such as bedrooms or hospital wards, without the ethical concerns of visual surveillance [21]. Furthermore, the system's immunity to lighting conditions promotes inclusivity for users in diverse environmental settings, ensuring that gesture based HCI remains accessible regardless of the time of day or the presence of visual obstructions like smoke or dust.

Conclusion:

This research successfully developed and validated a robust, real-time hand gesture recognition system for HCI utilizing a 24 GHz FMCW radar module (RD-03D). By leveraging the embedded signal processing capabilities of the radar, we effectively extracted a comprehensive 4D feature set (Range, Doppler, Angle (Azimuth), and Time) which captures the distinct spatio-temporal dynamics of hand movements. The subsequent training of a Bi-LSTM network on our 14,400-sample dataset achieved training accuracy of 99.85% and exceptionally high validation accuracy of 98.43% on unseen dataset across ten common gestures. This performance level confirms the superiority of FMCW radar over traditional vision-based systems in terms of privacy and robustness to environmental factors like lighting and occlusion. The successful implementation in a real-time embedded system demonstrates the high potential of this methodology as a viable, low-latency interface for next-generation interactive applications.

Recommendations:

Based on the experimental findings and the identified constraints of the 24 GHz FMCW framework, the following structured recommendations are proposed for future research to extend the system's capabilities:

Multi User Scenario Implementation:

Future work should focus on developing advanced signal processing and classification techniques to effectively isolate and classify gestures in environments with high user density [22]. Research into digital beamforming or subspace based interference cancellation could mitigate the effects of motion clutter, allowing the system to distinguish between the primary user and background subjects in crowded HCI environments.

Gesture Vocabulary Expansion:

Expanding the recognized gesture set to include more complex, three dimensional, or micro gestures would further enhance the expressiveness of the interface [23]. Investigating the distinctive micro Doppler signatures of individual finger movements could enable more granular control for applications requiring high precision, such as virtual reality (VR) or medical instrumentation [24].

Enhancing Feature Dimensionality:

The utility of a 5D feature set, incorporating Elevation Angle alongside Range, Doppler, Azimuth, and Time, should be investigated. This would allow the framework to capture the full volumetric motion of the hand, potentially improving classification accuracy and generalization for complex gestures that involve vertical spatial components.

Hardware Optimization and Edge Deployment:

A critical recommendation is the design of a standalone embedded gesture recognition device where the trained BiLSTM model is integrated directly onto specialized hardware, such as an FPGA or a dedicated edge AI accelerator [25][26]. This would facilitate low power, high efficiency inference, enabling the widespread commercial deployment of this radar based technology in battery operated wearable devices and portable electronics.

Acknowledgements:

The authors would like to express their sincere gratitude to NCAI at CEMTECH-NUST and Radar Lab at SINES-NUST for providing essential support and research facilities. We also extend our heartfelt thanks to our supervisors, for their continuous guidance and invaluable support throughout the development of this research.

Author Contributions: D.A. and M.A.M.: Conceptualization. D.A., M.A., M.J.H.: Data Acquisition and Methodology. D.A. and M.A.: Software and Validation. M.A. and MJH.: Visualization. D.A. and M.J.H.: Project Administration. M.A.M. and D.A.: Writing Original Draft. D.A., M.A., M.A.M., M.J.H.: Writing, Review and Editing. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest:

The authors declare that they have no conflicts of interest.

Project Details:

This research was conducted as part of the student project titled "FMCW-Based Hand Gesture Recognition" at National University of Technology (NUTECH).

Project Number: PST-FMCW-2025-01

Project Cost (Estimated): 150 USD (Covering hardware, software licenses, and consumables)

Project Completion Date: 25 Nov 2025

References:

[1] M Multi-Hand, Academic Editors, "Multi-Hand Gesture Recognition Using Automotive FMCW Radar Sensor," *Remote Sens.*, vol. 14, no. 10, p. 2374, 2022, doi: https://doi.org/10.3390/rs14102374.

[2] Yu Chiao Jhaung, Yu Ming Lin, "Implementing a Hand Gesture Recognition System Based on Range-Doppler Map," *Sensors*, vol. 22, no. 11, p. 4260, 2022, doi: https://doi.org/10.3390/s22114260.

[3] J. -W. Choi, C. -W. Park and J. -H. Kim, "FMCW Radar-Based Real-Time Hand Gesture Recognition System Capable of Out-of-Distribution Detection," *IEEE Access*, vol. 10, pp. 87425–87434, 2022, doi: 10.1109/ACCESS.2022.3200757.

[4] X. Dong, Z. Zhao, Y. Wang, T. Zeng, J. Wang, and Y. Sui, "FMCW Radar-Based Hand Gesture Recognition Using Spatiotemporal Deformable and Context-Aware Convolutional 5-D Feature Representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3122332.

[5] L. Gan, Y. Liu, Y. Li, R. Zhang, L. Huang, and C. Shi, "Gesture Recognition System

Using 24 GHz FMCW Radar Sensor Realized on Real-Time Edge Computing Platform," *IEEE Sens. J.*, vol. 22, no. 9, pp. 8904–8914, May 2022, doi: 10.1109/JSEN.2022.3163449.

[6] Y. Zhao, V. Sark, M. Krstic, and E. Grass, "Low-Complexity Gesture Recognition based on FMCW Radar," *2024 21st Eur. Radar Conf. EuRAD 2024*, pp. 388–391, 2024, doi: 10.23919/EuRAD61604.2024.10734900.

[7] K. Alirezazad and L. Maurer, "FMCW Radar-Based Hand Gesture Recognition Using Dual-Stream CNN-GRU Model," *2022 24th Int. Microw. Radar Conf. MIKON 2022*, 2022, doi: 10.23919/mikon54314.2022.9924984.

[8] Haili Wang, Muye Zhang, "Real-Time Hand Gesture Recognition in Clinical Settings: A Low-Power FMCW Radar Integrated Sensor System with Multiple Feature Fusion," *Sensors*, vol. 25, no. 13, p. 4169, 2025, doi: https://doi.org/10.3390/s25134169.

[9] A. Fusco, Z. Amir Zaman, S. Hazra, L. Servadei and R. Wille, "Enhancing FMCW Radar Gesture Classification With Physically Interpretable Data Augmentation," *IEEE Access*, vol. 13, pp. 60556–60569, 2025, doi: 10.1109/ACCESS.2025.3556565.

[10] Yuhang Shi, Lihong Qiao, "Semi-Supervised FMCW Radar Hand Gesture Recognition via Pseudo-Label Consistency Learning," *Remote Sens.*, vol. 16, no. 3, p. 2267, 2024, doi: https://doi.org/10.3390/rs16132267.

[11] Y. Zhao, V. Sark, M. Krstic, and E. Grass, "High-Efficiency Gesture Recognition Using Multiple mmWave FMCW RADARs," *2024 Int. Symp. Networks, Comput. Commun. ISNCC 2024*, 2024, doi: 10.1109/ISNCC62547.2024.10758972.

[12] J. T. Yu, L. Yen, and P. H. Tseng, "MmWave Radar-based Hand Gesture Recognition using Range-Angle Image," *IEEE Veh. Technol. Conf.*, vol. 2020-May, May 2020, doi: 10.1109/VTC2020-Spring48590.2020.9128573.

[13] W. Jiang, Y. Ren, Y. Liu, Z. Wang, and X. Wang, "Recognition of dynamic hand gesture based on mm-wave FMCW radar micro-Doppler signatures," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2021-June, pp. 4905–4909, 2021, doi: 10.1109/ICASSP39728.2021.9414837.

[14] Y. Zhao, V. Sark, M. Krstic, and E. Grass, "Synthetic Training Data Generator for Hand Gesture Recognition Based on FMCW RADAR," *Proc. Int. Radar Symp.*, vol. 2022-September, pp. 463–468, 2022, doi: 10.23919/irs54158.2022.9904997.

[15] Zhangjin Xiong, Kaixue Ma, "Hand gesture recognition based on micro-Doppler radar using graph neural network," *Electron. Lett.*, vol. 60, no. 3, 2024, doi: 10.1049/ell2.13100.

[16] "Dual-Stream BiLSTM–Transformer Architecture for Real-Time Two-Handed Dynamic Sign Language Gesture Recognition." Accessed: May 05, 2026. [Online]. Available: https://www.mdpi.com/2076-3417/16/6/2912

[17] "Dynamic Gesture Recognition Based on FMCW Millimeter Wave Radar: Review of Methodologies and Results - PubMed." Accessed: May 05, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37687932/

[18] A. Ślesicka and A. Kawalec, "Real-Time Hand Gesture Recognition for IoT Devices Using FMCW mmWave Radar and Continuous Wavelet Transform," *Electron. 2026, Vol. 15, Page 250*, vol. 15, no. 2, p. 250, Jan. 2026, doi: 10.3390/ELECTRONICS15020250.

[19] Y. Sun, T. Fei, X. Li, A. Warnecke, E. Warsitz, and N. Pohl, "Real-Time Radar-Based Gesture Detection and Recognition Built in an Edge-Computing Platform," *IEEE Sens. J.*, vol. 20, no. 18, pp. 10706–10716, Sep. 2020, doi: 10.1109/JSEN.2020.2994292.

[20] Reda El Hail, Pouya Mehrjouseresht, "Radar-Based Human Activity Recognition: A

Study on Cross-Environment Robustness," *Electronics*, vol. 14, no. 5, p. 875, 2025, doi: https://doi.org/10.3390/electronics14050875.

[21] B. Van Amsterdam, M. J. Clarkson, and D. Stoyanov, "Gesture Recognition in Robotic Surgery: A Review," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 6, pp. 2021–2035, Jun. 2021, doi: 10.1109/TBME.2021.3054828.

[22] D. Rodrigues and C. Li, "Hand Gesture Recognition Using FMCW Radar in Multi-Person Scenario," *2021 IEEE Top. Conf. Wirel. Sensors Sens. Networks, WiSNeT 2021*, pp. 50–52, Jan. 2021, doi: 10.1109/WISNET51848.2021.9413794.

[23] "Micro Gesture Recognition with Multi-Dimensional Feature Fusion and CQ-MobileNetV3 Using FMCW Radar - PubMed." Accessed: May 05, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/41305159/

[24] "(PDF) Research Progress on Technologies and Applications of Human-Computer Interaction Gestures Based on Vision and Sensors." Accessed: May 05, 2026. [Online]. Available: https://www.researchgate.net/publication/402496997_Research_Progress_on_Technologies_and_Applications_of_Human-Computer_Interaction_Gestures_Based_on_Vision_and_Sensors

[25] A. Mohan, H. K. Meena, M. Wajid, and A. Srivastava, "Real-Time Dynamic Hand Gesture Recognition Using mmWave Radar on FPGA," *IEEE Embed. Syst. Lett.*, 2025, doi: 10.1109/LES.2025.3634571.

[26] A. Chopde, S. Joshi, S. Surve, and S. Chavanke, "FPGA based real-time hand gesture recognition system," *2024 15th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2024*, 2024, doi: 10.1109/ICCCNT61001.2024.10724416.

Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.