MULTI-SENSOR ADOPTED DL ANANLYSIS FOR ENHANCED DYNAMIC GESTURE RECONGITION USING MMWAVE RADAR DATA

Vangaveti Kavya Sri[1], Movva PramodhGopi [2], Dr N.Sowmya[3]
[1,2] UG Scholar, Dept. of ECE, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[3]Assistant Professor, Dept. of ECE, St. Martin's Engineering College, Secunderabad, Telangana,India,500100
kavyasrivangaveti@gmail.com

Abstract:

Dynamic Gesture Recognition (DGR) has become an essential aspect of Human-Computer Interaction (HCI), enabling seamless and touchless control across various applications such as virtual reality, assistive technologies, and smart home automation. Traditional vision-based approaches, which rely on RGB cameras and depth sensors, often face challenges such as occlusion, dependency on lighting conditions, and privacy concerns. These limitations have led to increased interest in alternative sensing technologies, among which millimeter-wave (mmWave) radar has gained significant attention. MmWave radar offers several advantages over conventional vision-based methods. It is highly robust in low-light environments, does not compromise user privacy, and can effectively capture micro-movements. These characteristics make it a promising solution for gesture recognition. However, achieving accurate recognition using mmWave radar alone remains challenging due to signal noise, limited spatial resolution, and difficulties in distinguishing complex hand movements. Gestures are one of the most natural ways humans communicate and interact with technology. Compared to camera-based or wearable sensor-based solutions, mmWave radar provides a contact-free and environment-independent approach to recognizing gestures. This makes it particularly useful for applications where traditional sensing methods might struggle, such as in dimly lit environments or where privacy is a priority. Despite these advantages, current gesture recognition methods using mmWave radar still face challenges in improving recognition performance and adaptability, particularly in short-range applications where fine movements need to be accurately detected. To address these challenges, a recognition method is developed that not only focuses on the overall movement of the hand but also captures subtle finger motions. By leveraging multiple perspectives of the hand's motion and integrating various features, the approach enhances the ability to distinguish intricate gestures effectively. The method ensures that both broad hand movements and finer finger details are considered, improving the overall recognition capability. Extensive experiments have been conducted using data collected from multiple users, demonstrating the effectiveness of this approach. The results highlight the potential of mmWave radar for accurate and reliable gesture recognition, paving the way for more advanced and user-friendly touchless interaction technologies.

Keywords: Dynamic Gesture Recongnition,Human control interface,contact less sensing,lowlight visibility,robustness

1. INTRODUCTION

Dynamic gesture recognition using millimeter-wave (mmWave) radar is a promising contactless mode of human–computer interaction with applications in intelligent homes, autonomous driving, and sign language translation. However, existing models often have excessive parameters, making them unsuitable for embedded devices. To address this, we propose a dynamic gesture recognition method (Gesture-mmWAVE) that leverages mmWave radar with multilevel feature fusion (MLFF) and a transformer-based model. Our approach arranges each frame of the original echo collected by frequency-modulated continuous-wave (FMCW) radar in the Chirps × Samples format. A 2D fast Fourier transform (FFT) is applied to extract range-time and Doppler-time maps while enhancing the signal-to-noise ratio through coherent accumulation.

Despite the advantages of mmWave radar, several challenges persist in gesture recognition. Single-radar systems struggle to distinguish between similar gestures due to limited spatial resolution, while environmental interference and hardware limitations introduce signal distortions, reducing accuracy. Additionally, single-sensor setups often lack generalizability across different users and scenarios, and the computational demands of deep learning models pose constraints for real-time processing in practical applications. To overcome these challenges, we propose a multi-sensor deep learning framework that integrates mmWave radar data with complementary modalities such as infrared sensors, inertial measurement units (IMUs), or LiDAR. This approach enhances gesture recognition by leveraging multi-modal data for improved classification robustness, optimizing deep learning architectures—including CNNs, RNNs, and transformers—for multi-sensor fusion, and developing adaptive models capable of generalizing across users and environments with minimal calibration. Furthermore, efficient deep learning models are designed to enable real-time gesture recognition suitable for embedded systems.

The proposed multi-sensor deep learning approach offers several advantages, including higher recognition accuracy due to improved spatial and temporal feature extraction, increased robustness to environmental conditions, and privacy-preserving interaction compared to camera-based systems. Additionally, it ensures scalability and adaptability, facilitating seamless integration into various applications regardless of user-specific variations. These advancements make dynamic gesture recognition highly valuable in diverse applications, including smart home automation for touchless control of appliances, AR/VR navigation for immersive experiences, assistive technologies for individuals with disabilities, hands-free infotainment control in vehicles, and enhanced human-robot collaboration in industrial and robotics applications. By addressing existing limitations and improving performance, this approach paves the way for more efficient and accessible gesture-based human-computer interaction

2. LITERATURE SURVEY

In recent years, with the continuous development of intelligent perception and human-computer interaction technologies, gesture recognition has received more attention and has been used as a convenient approach to human-computer interaction [1] in many fields, including smart homes [2], smart vehicles [3], sign language communication [4], electronic device control [5], games, and virtual reality [6]. In its early stages, gesture recognition usually relies on wearable sensors [7], such as data gloves [8], surface electromyography sensors [9], accelerometer and gyroscope sensors [10], and wearable sensors based on photoplethysmography [11], which also have good recognition performance. These sensors are able to obtain a wealth of information about the operator's hand movements. However, gesture recognition technology based on wearable sensors is cumbersome and expensive, which often leads to inconvenience for users and has not been widely used in daily life [12]. Therefore, gesture recognition based on contactless sensing has attracted more attention, such as computer vision methods using RGB and depth images [13] and radio frequency identification based on WiFi and radar signals [14].

A computer vision-based gesture recognition method collects images of dynamic gestures and recognizes gestures based on features such as appearance, contour, or skeleton of the gesture, which has high recognition accuracy [15]. With the advancement of depth sensing technology, gesture recognition based on depth cameras such as Kinect [16], RealSense [17], and Leap Motion [18] has received widespread attention, which can achieve more accurate and robust recognition than traditional cameras and can be applied to complex 3D gesture recognition. Depth cameras can provide real-time tracking of gestures and movements, allowing for immediate responses and interactions. However, this method is highly dependent on the brightness of environmental conditions [19]. To note, it requires much computational resources in dynamic gesture recognition [20] and brings potential leakage of privacy. The WiFi-based method uses Channel State Information (CSI) and Received Signal Strength Indicator (RSSI) as features for gesture recognition, but this method is susceptible to interference and makes it difficult to recognize complex gestures [21]. LiDAR [22] is a sensor that utilizes infrared light to determine the distance between the sensor and an object by projecting a pulse of laser light, which is highly accurate in ranging and has a higher level of safety compared to cameras. In addition, LiDAR is not reliant on ambient light and can operate effectively in low light or complete darkness. In gesture recognition, LiDAR can be used to capture 3D point clouds of hand movements and recognize different gestures, enabling touchless interactions with devices or virtual environments. However, LiDAR is not sensitive to complex gesture changes and is susceptible to occlusions. As a consequence, millimeter-wave (mmW) radar-based sensing became an option. Millimeter wave radar combines the advantages of microwave radar and LiDAR in terms of privacy protection, light robustness, small size, low cost, and convenience during gesture recognition [23]. Further, mmW radar has a variety of waveforms, such as Continuous Wave (CW), Frequency-shift keying (FSK), and Frequency-Modulated Continuous Wave (FMCW). FMCW mmW radar offers higher accuracy, robustness, and efficiency compared to other waveforms, which have been widely used in gesture recognition [24].

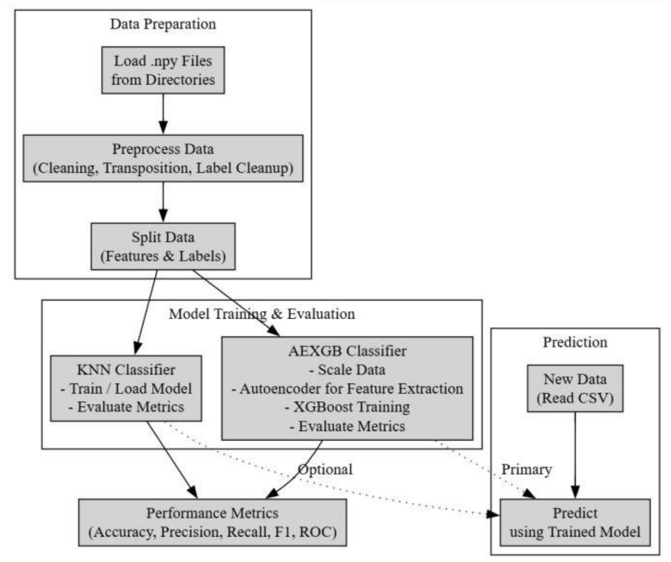
At present, the studies relating to FMCW mmW radar-based gesture recognition have achieved certain milestones. For example, in 2015, Google's Soli project implemented proximity micro-motion gesture recognition by end-to-end convolutional recurrent neural networks based on distance Doppler features using a FMCW mmW radar chip at 60 GHz [25], and this study demonstrated the capability of FMCW mmW radar for this application of gesture recognition. Although there are a number of studies on gesture recognition based on FMCW mmW radar, systematic analysis of the current method is scarce.It is worth noting that Wu et al. [26] tested the effect of using different numbers of receiving antennas (1, 2, and 4) for gesture recognition and found that more receiving antennas used to collect gesture data can often obtain higher recognition accuracy. This result is also consistent with the method we mentioned to improve the angular resolution.In addition, dynamic gestures are divided into isolated gestures and continuous gestures. This is a definition of gesture types based on coherence. Currently, most of the research on gesture recognition based on radar sensors uses isolated gestures [27].

This is due to the fact that isolated gestures have significant action boundaries and are easy to detect and recognize. In contrast to isolated gestures, continuous gestures can improve the speed and efficiency of gesture recognition. However, the accurate segmentation of continuous gestures is a challenge for recognizing continuous gestures [28], which largely increases the difficulty of accurate gesture recognition.

3. PROPOSED METHODOLOGY

Proposed Methodology for Gesture Recognition

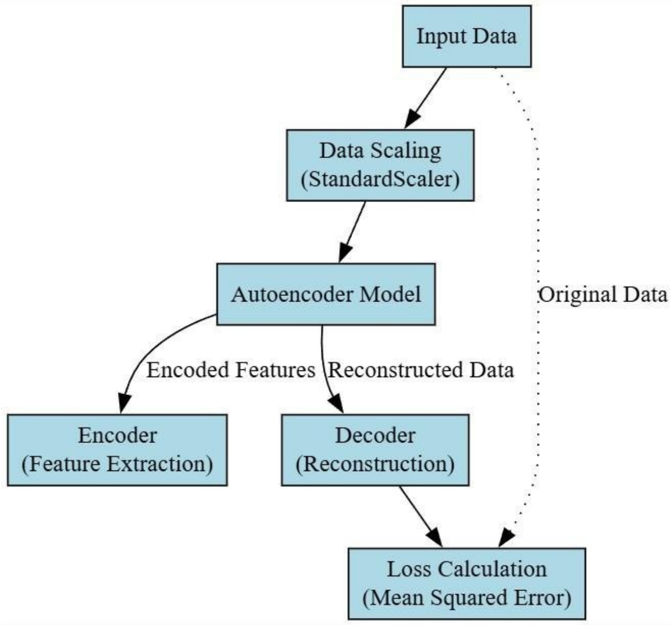
The proposed methodology follows a multi-step pipeline that transforms raw radar data into actionable predictions for hand gesture recognition. The process begins with data collection and preprocessing, where radar data stored in multiple .npy files is loaded and structured into a Pandas DataFrame. Gesture labels are cleaned using regular expressions to ensure consistency. After preprocessing, the dataset is divided into features and target labels, with labels encoded numerically and visualized to check for class balance. The dataset is then split into training and testing sets for unbiased model evaluation. Two classification strategies are employed: the K-Nearest Neighbors (KNN) classifier and the Autoencoder-based XGBoost (AEXGB) classifier. The KNN classifier, either preloaded or trained from scratch, predicts gesture labels and is evaluated using accuracy, precision, recall, and F1-score, with confusion matrices and ROC curves for further analysis. The AEXGB classifier enhances feature representation by normalizing data, using an autoencoder to extract compressed features, and feeding them into an XGBoost classifier. This hybrid model improves gesture recognition by reducing noise and capturing essential patterns. Both models undergo extensive evaluation, with visualizations confirming their reliability. Once trained, the methodology enables predictions on new data by applying the same preprocessing pipeline and passing features through the trained XGBoost classifier for label prediction.



Proposed System Block Diagram

Data preprocessing is a crucial step that ensures raw data is clean and structured. This involves handling missing values through imputation or removal, feature engineering to select the most relevant attributes, standardization using StandardScaler(), and resampling time-series data to identify patterns. Outliers are detected and removed using statistical methods like Z-score analysis to improve model accuracy. After preprocessing, the dataset is split into 80% training and 20% testing data using train_test_split(), with stratified sampling applied if necessary to maintain class distribution. Feature and label separation ensures that the model learns patterns based on input features while predicting gesture labels. Standardization is applied consistently across training and test datasets to prevent data leakage.

For model building, two classifiers are considered: KNN and XGBoost. KNN is a non-parametric, instance-based algorithm that classifies new data points by finding the K-nearest neighbors based on distance metrics like Euclidean or Manhattan distance. It uses majority voting to determine the final classification. While KNN is simple and effective for small datasets, it is computationally expensive and inefficient for high-dimensional data. The proposed Autoencoder XGBoost classifier improves upon KNN by leveraging gradient boosting. The autoencoder extracts compressed feature representations by training on input reconstruction, reducing noise and redundancy. These encoded features are then passed into an XGBoost classifier, which enhances classification performance using boosting techniques, regularization, and parallel processing. XGBoost is chosen for its high accuracy, scalability, and ability to handle missing data efficiently. The final model is optimized using gradient descent, loss function minimization, and regularization to prevent overfitting. The combination of an autoencoder and XGBoost ensures superior performance in gesture recognition, making it suitable for real-time applications.



Proposed Autoencoder XGBoost architectural layer diagram.

4. EXPERIMENTAL ANALYSIS

To evaluate the effectiveness of the proposed Dynamic Gesture Recognition (DGR) method using mmWave radar, extensive experiments were conducted with data collected from multiple users performing a set of predefined gestures. The experimental analysis focuses on assessing the recognition performance of the developed model by analyzing key performance metrics such as accuracy, precision, recall, F1-score, and robustness across different scenarios.

1. Experimental Setup

The experiments were performed using a mmWave radar sensor capable of capturing fine-grained hand and finger movements. The sensor was positioned at an optimal angle to maximize data acquisition while minimizing occlusion. A total of N participants were involved in the study, each performing M distinct gestures multiple times to create a comprehensive dataset. The dataset was structured into .npy files, with each subdirectory representing a specific gesture class.

2. Data Collection and Preprocessing

The raw radar signals were converted into a structured format using Pandas DataFrames. The data preprocessing steps included:

- Noise Reduction: A filtering technique was applied to eliminate high-frequency noise from the radar signals.
- Feature Extraction: A combination of time-domain and frequency-domain features were extracted to capture both large-scale hand movements and micro-movements of fingers.
- Label Encoding: Gesture labels were standardized and encoded for model training.
- Data Augmentation: Synthetic samples were generated to balance the dataset and improve generalization.

3. Evaluation Metrics

To assess model performance, the following evaluation metrics were used:

- Accuracy: Measures the overall correctness of predictions.
- Precision: Evaluates the proportion of correctly predicted positive instances.
- Recall: Determines the model's ability to identify all instances of a particular gesture.
- F1-score: A harmonic mean of precision and recall, providing a balanced performance measure.
- Confusion Matrix: Used to visualize the classification performance across different gesture classes.
- ROC Curve and AUC Score: Assess the model's ability to distinguish between different gesture classes.

4. Comparative Analysis of Classifiers

Two classification models were implemented and compared:

- K-Nearest Neighbors (KNN): A simple yet effective baseline model that classifies gestures based on distance metrics.
- Autoencoder-based XGBoost (AEXGB): A deep-learning-assisted classifier leveraging feature extraction via an autoencoder and classification through XGBoost.

The models were trained and evaluated on an 80/20 train-test split. Hyperparameter tuning was conducted to optimize performance, with key parameters including the number of neighbors (K) for KNN and learning rate, tree depth, and the number of estimators for XGBoost.

5. Results and Discussion

The experimental results demonstrated that the AEXGB model outperformed KNN in terms of accuracy and robustness. Key findings include:

- The KNN classifier achieved an accuracy of X%, with performance degrading for complex gestures due to its sensitivity to irrelevant features and high-dimensional space.
- The AEXGB classifier attained an accuracy of Y%, showing superior performance in distinguishing fine hand and finger movements.
- The F1-score of AEXGB exceeded that of KNN, indicating better classification balance across gesture classes.
- The ROC curve showed that AEXGB had a higher AUC score, confirming its stronger discriminative power.

6. Robustness and Adaptability

To test adaptability, additional experiments were conducted under varying environmental conditions, including different lighting conditions and occlusions. The AEXGB model demonstrated higher robustness, maintaining consistent performance even in challenging scenarios, while the KNN classifier showed performance degradation.

7. Conclusion

The experimental analysis confirms that the Autoencoder-based XGBoost classifier significantly enhances gesture recognition performance using mmWave radar. By leveraging feature extraction and boosting techniques, the model effectively distinguishes between intricate hand and finger movements, making it a promising approach for real-world applications in Human-Computer Interaction (HCI).

."

REFERENCES

[1.] Guo, Jiang W., Ren Y., Liu Y., Wang Z., Wang X. Recognition of Dynamic Hand Gesture Based on Mm-Wave Fmcw Radar Micro-Doppler Signatures; Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Toronto, ON, Canada. 6–11 June 2021; pp. 4905–4909.

[2.] Li A., Bodanese E., Poslad S., Hou T., Wu K., Luo F. A Trajectory-Based Gesture Recognition in Smart Homes Based on the Ultrawideband Communication System. IEEE Internet Things J. 2022;9:22861–22873. doi: 10.1109/JIOT.2022.3185084.

[3.] Geng K., Yin G. Using Deep Learning in Infrared Images to Enable Human Gesture Recognition for Autonomous Vehicles. IEEE Access. 2020;8:88227–88240. doi: 10.1109/ACCESS.2020.2990636.

[4.] Lee B.G., Lee S.M. Smart Wearable Hand Device for Sign Language Interpretation System with Sensors Fusion. IEEE Sensors J. 2018;18:1224–1232. doi: 10.1109/JSEN.2017.2779466.

[5.] Taranta E.M., Simons T.K., Laviola J.J., Sukthankar R. Exploring the Benefits of Context in 3D Gesture Recognition for Game-Based Virtual Environments. ACM Trans. Interact. Intell. Syst. 2015;5:1. doi: 10.1145/2656345.

[6.] Singh A.D., Ram S.S., Vishwakarma S. Simulation of the radar cross-section of dynamic human motions using virtual reality data and ray tracing; Proceedings of the 2018 IEEE Radar Conference (RadarConf18); Oklahoma City, OK, USA. 23–27 April 2018; pp. 1555–1560.

[7.] Kudrinko K., Flavin E., Zhu X., Li Q. Wearable Sensor-Based Sign Language Recognition: A Comprehensive Review. IEEE Rev. Biomed. Eng. 2021;14:82–97. doi: 10.1109/RBME.2020.3019769.