International Journal of Applied Science and Engineering

Human flow detection using CSI technology combined with STFT and machine learning

We-Ling Lin [1], Li-Syuan Chen [2], Jun-Jia Ou [3*]

[1] *Department of Intelligent Production Engineering, National Taichung University of Science and Technology, Taichung, Taiwan*
[2] *Department of Computer Science and Information Engineering, National Chung Hsing University, Taichung, Taiwan*
[3] *Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taichung, Taiwan*

ABSTRACT

This study investigates the significance of pedestrian flow detection technologies in highly populated areas such as museums, exhibition centers, and amusement parks, particularly in the fields of enterprise management, smart healthcare, and public safety. Traditional detection methods, such as cameras and infrared sensors, are often constrained by privacy concerns and environmental factors. In contrast, Channel State Information (CSI) technology utilizes variations in wireless communication signals, offering a privacy-preserving and cost-effective solution. To validate this approach, the study employs lightweight WiFi transceivers to capture signal perturbations caused by human activities, with a custom labeling and balancing method applied to the collected data. Short-Time Fourier Transform (STFT) is then used to convert the data into time-frequency domain representations for feature extraction. The processed dataset is subsequently fed into machine learning models for training and prediction. Four machine learning algorithms—Random Forest Classifier (RandomForestClassifier), Support Vector Classifier (SVC), XGBoost Classifier (XGBClassifier), and Gradient Boosting Classifier (GradientBoostingClassifier)—were evaluated, all demonstrating excellent performance. Among these, the Random Forest Classifier achieved 99% accuracy in scenarios detecting 0–2 people passing through the monitored area. The results indicate that integrating WiFi-based CSI technology with machine learning models can enable precise and efficient real-time pedestrian flow monitoring, showcasing promising applications in museum and healthcare environments.

Keywords: Artificial intelligence of things (AIoT), Channel state information (CSI), Human flow detection, Machine learning, Short-time fourier transform (STFT).

OPEN ACCESS

Received: December 27, 2024
Revised: January 24, 2025
Accepted: February 7, 2025

Corresponding Author:
Jun-Jia Ou
s1311232029@nutc.edu.tw

Publisher:
Chaoyang University of Technology
ISSN: 1727-2394 (Print)
ISSN: 1727-7841 (Online)

1. INTRODUCTION

In the context of advancing digital transformation and the growing demand for intelligent solutions, pedestrian flow detection technology plays a crucial role in enterprise operations management, smart healthcare, and public safety (Espinosa et al., 2021). From office buildings to hospitals, accurately tracking the dynamic distribution of people helps prevent overcrowding (Azizi et al., 2020) and improves energy efficiency (Shah et al., 2019). In public spaces such as hospitals, art galleries, exhibition halls, and amusement parks, understanding pedestrian flow patterns allows managers to identify hotspots and sparsely populated areas, optimizing flow planning and ensuring efficient crowd evacuation during emergencies. However, traditional pedestrian flow detection systems face numerous challenges in practical applications.

Currently, common pedestrian flow detection technologies include the cameras, infrared sensors, and radio-frequency identification (RFID) tags (Fitwi et al., 2021; Rangdale et al., 2023), which have been widely adopted in the commercial areas, public facilities, and increasingly apparent. Image-based detection systems typically require high-resolution cameras and good lighting conditions to accurately capture pedestrian flow changes (Winkler et al., 2014; Kwon et al., 2024; Wang et al., 2024). This not only raises hardware costs but also makes the systems vulnerable to environmental factors such as poor lighting. Moreover, the use of cameras poses significant privacy concerns, particularly in public spaces, potentially eliciting public resistance and skepticism toward surveillance systems. In enterprise environments, compliance with privacy protection laws further complicates the adoption of such technologies.

In contrast, while infrared sensors (Narayana et al., 2015) mitigate the direct privacy issues associated with image-based systems, their detection range is limited, and they are susceptible to external environmental interference, which can lead to sensor inaccuracies. More importantly, infrared sensors struggle to accurately distinguish individuals in crowded or rapidly changing environments, making it difficult to meet the growing demand for diverse and precise pedestrian flow detection solutions.

RFID, as a detection method based on electronic tags and reading devices, avoids some limitations of cameras and infrared sensors (Want, 2006). For example, RFID systems are independent of lighting conditions and less sensitive to environmental interference, improving detection stability and accuracy. However, the application of RFID technology also has challenges (Jia et al., 2012). First, each detected individual requires an RFID tag, which can increase costs and complexity for large crowds. Additionally, deploying RFID reading devices requires professional setup, and their range and accuracy may be limited by device specifications, potentially creating blind spots in large or complex scenarios. The effectiveness of RFID tags may also decline in cases of rapid movement or tag obstruction, further limiting its flexibility in various applications.

To address the challenges of traditional crowd detection methods, the application of Channel State Information (CSI) technology has emerged as a novel solution. By analyzing variations in wireless communication signals, CSI technology can accurately capture changes in the spatial environment. Unlike conventional detection systems, CSI technology does not rely on video surveillance or additional tagging devices but instead utilizes existing WiFi signals for real-time monitoring, thereby reducing deployment costs and effectively protecting user privacy (Moshiri et al., 2021; Ge et al., 2023). These characteristics make CSI technology particularly suitable for privacy-sensitive environments such as corporate office buildings, exhibition halls, and healthcare facilities. Numerous CSI-based crowd detection methods have been developed, typically using Access Point (AP) devices as data sources combined with machine learning (Xiao et al., 2019) or deep learning techniques (Liu et al., 2017) to develop recognition models. These methods leverage AP devices to obtain stable and high-dimensional CSI data, achieving high accuracy. However, systems relying on such equipment face high costs, increasing the difficulty of deployment. Moreover, many scenarios, such as corridors and stairwells, do not require such powerful equipment. This study aims to design and validate a lightweight CSI-based crowd detection technology, offering an efficient and cost-effective solution while exploring its feasibility in real-world scenarios.

Expected contributions are,

- Creation of a low-cost, non-contact crowd detection system: Propose an innovative solution based on MCU-enabled CSI technology, utilizing WiFi signal variations for crowd monitoring to achieve a low-cost, privacy-preserving deployment model.
- Integration of lightweight hardware architecture: Implement a compact design using the ESP32 module as the core device, enabling collaborative operation of AP and STA (Station Mode). The ESP32 STA transmits signals, and the AP receives and analyzes CSI data, providing an efficient and flexible hardware design.
- Introduction of advanced feature extraction techniques: Employ short-time fourier transform (STFT) to extract time-frequency features from raw CSI data, improving classification accuracy and reducing noise interference.
- Comprehensive model evaluation and comparison: Validate the effectiveness of CSI data processing by comparing the performance of Random Forest Classifier (RandomForestClassifier), Support Vector Classifier (SVC), XGBoost Classifier (XGBClassifier) and Gradient Boosting Classifier (GradientBoostingClassifier).

2. MATERIALS AND METHODS

2.1 Literature Review

In recent years, research has increasingly focused on utilizing modern communication technologies for detecting and tracking individuals in the various environments (Qian et al., 2017; Shi et al., 2022; Wu et al., 2023). Among these technologies, CSI refers to detailed information about a wireless communication channel obtained at the receiver, including features such as signal strength, phase, and frequency variations. These data provide insights into the state of the communication channel, enabling analysis of the spatial variations of wireless signals over time. The unique characteristics of CSI make it particularly effective in sensing minute environmental changes within a wireless network, such as detecting the presence and movement of obstacles, or even human postures and actions (Ma et al., 2021; Wang et al., 2021; Zhang et al., 2022; Zhou et al., 2023).

The application of CSI technology relies on Multiple-Input Multiple-Output (MIMO) systems. By transmitting and receiving signals through multiple antennas, MIMO systems allow the receiver to gather detailed information about the channel. This not only enhances the capacity of wireless communication but also improves the ability to monitor spatial changes with precision. As a result, CSI technology has demonstrated significant potential in nonvisual surveillance, privacy protection, and intelligent environmental sensing (Lin et al., 2015; Chen et al., 2019). One of the earliest and most extensively studied applications of CSI technology is activity recognition. Researchers have analyzed variations in WiFi signals caused by human movements to identify specific actions. For instance, Zeng et al. proposed a CSI-based activity recognition system (Zeng et al., 2016; Ding et al., 2019), which differentiates various body movements by analyzing changes in WiFi signals. This approach eliminates the need for wearable devices while ensuring privacy, as it relies on wireless signal data sequences rather than image data. The system effectively identifies basic actions such as walking, sitting, and standing, achieving an accuracy rate of over 80%.

With recent technological advancements, cutting-edge artificial intelligence and imaging technologies have been widely applied in the medical field, driving innovations in medical diagnostics and healthcare. For instance, Laghari et al. (2023) proposed a model based on the Deep Residual-Dense Network (DRDN) and Bidirectional Recurrent Neural Network (BiRNN), successfully applied to the detection of atrial fibrillation (AF). Leveraging deep learning techniques, this model significantly improved the early identification rate of arrhythmias. Additionally, Saeed et al. (2023) developed a convolutional neural network (CNN) model named DeepLeukNet for classifying microscopic images, aiding in the detection of acute lymphoblastic leukemia (ALL) and demonstrating the potential of AI in cancer diagnosis.

In the domain of medical imaging, Laghari et al. (2022) explored how to collect and interpret medical images ranging from the nanoscale to hyperspectral imaging in highly challenging environments. These technologies enable detailed analysis of internal human structures, thereby enhancing diagnostic accuracy. Another review by Laghari et al. (2024) focused on the applications of virtual and augmented reality in healthcare, particularly in the quality assessment of user experiences in serious games. These technologies have been employed in patient rehabilitation training and psychotherapy. Furthermore, Das et al. (2023) proposed a remote healthcare system combining electroencephalogram (EEG) and kinect sensors. This system not only provides convenient health monitoring for the elderly and disabled but also facilitates remote diagnostics. The value of such technologies became increasingly apparent during the pandemic, highlighting their importance for modern healthcare.

The application of CSI technology in the field of healthcare is growing, achieving notable advancements in health monitoring and intelligent medical research. CSI can accurately capture environmental information around the human body by detecting subtle changes in wireless signals, eliminating the need for wearable devices, and making it ideal for contactless monitoring. This feature is particularly useful for detecting physiological indicators such as falls, respiratory rates and heart rates (Wang et al., 2020). For instance, traditional fall detection methods rely on cameras or wearable devices, often facing privacy issues or usage inconvenience. In contrast, CSI-based technologies have been applied to monitor vital signs such as breathing and heart rate (Liu et al., 2023; Lei et al., 2024; Sun et al., 2024). By analyzing minor changes in wireless signals, CSI can identify slight chest movements, thereby inferring respiratory and heart rates. This technology is not only suitable for daily health monitoring but also advantageous for remote monitoring scenarios.

In the realm of CSI applications for human flow, various studies have focused on developing technologies for spatial crowd counting. For example, Xiao et al. (2019) proposed an AI-based mobile sensing technology that uses wireless signals for crowd detection. This method employs machine learning models to analyze signal features, achieving efficient flow monitoring in dynamic environments. However, it relies on high-performance hardware and complex data processing architectures, leading to high system deployment costs and limited adaptability. Similarly, Liu et al. (2017) introduced a deep learning model, WiCount, which utilizes WiFi signals for crowd counting by analyzing high-dimensional data. While this approach demonstrates good accuracy, it demands significant device resources and faces challenges related to noise during the feature extraction process.

In comparison, the method proposed in this study leverages lightweight CSI devices combined with STFT for feature extraction. This approach effectively reduces noise interference in the data while enhancing the accuracy of classification models. Moreover, the system has relatively low hardware requirements, enabling efficient and cost-effective deployment, particularly suitable for crowd counting in narrow spaces such as corridors or staircases. Compared to the aforementioned studies, this method emphasizes practicality and feasibility, achieving a balance between performance and cost, thereby providing an innovative solution for crowd detection.

2.2 Research Methodology

With advancements in wireless communication technology, environmental monitoring through CSI signals has emerged as a novel and practical approach. In existing studies, human flow detection often relies on image recognition methods, which are typically complex and costly to deploy in real-world scenarios. This study aims to develop a low-cost, simplified CSI-based system to replace image-based human flow detection. Therefore, we set up a simulated environment to capture human flow data and used this data to build an efficient, cost-effective, and adaptable human flow detection system. The overall flowchart is illustrated in Fig. 1.

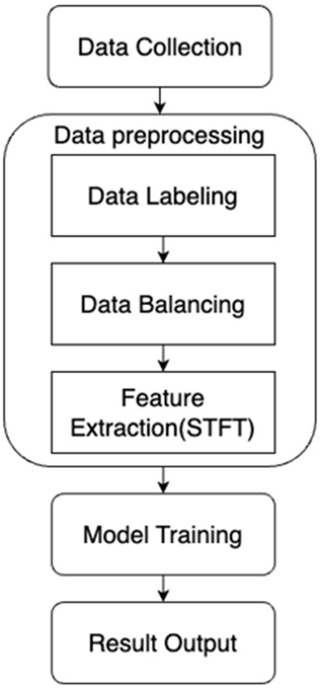


Fig. 1. Overall flowchart

2.2.1 Environment Setup

The experimental environment was set up in an open corridor where we placed our equipment. The setup consists of four single-chip microcontrollers, a camera, and a computer. Each single-chip unit includes two signal receivers and transmitters made from single-chip technology, capturing signal variations as solid objects pass through the space between them. A camera was also included in the setup to record video, facilitating later data annotation. A schematic of the complete environment setup is shown in Fig. 2, while the experimental setup in the actual environment is depicted in Fig. 3.

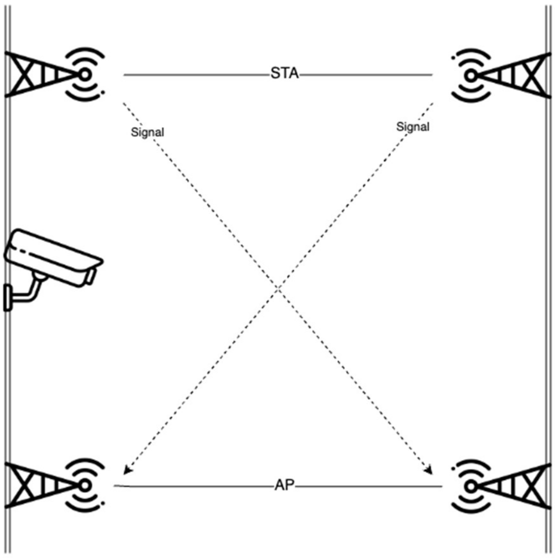


Fig. 2. Environment setup diagram

Fig. 3. Experimental setup in the actual environment

As seen in Fig. 2 and Fig. 3, to ensure the captured data has sufficient signal strength and distinguishable features, we opted for two sets of transmitter/receiver pairs positioned at an angled transmission layout, with signals crossing each other. The distance between each unit is two meters, which extends the signal coverage area and ensures a more noticeable signal variation when people pass through. During the experiment, data collection was conducted across different numbers of people passing through the corridor, covering scenarios from zero to two individuals. as shown in Fig. 4 (a, b). This setup allows the system to operate effectively under varying crowd sizes. For data accuracy, each experiment was performed under consistent environmental conditions.

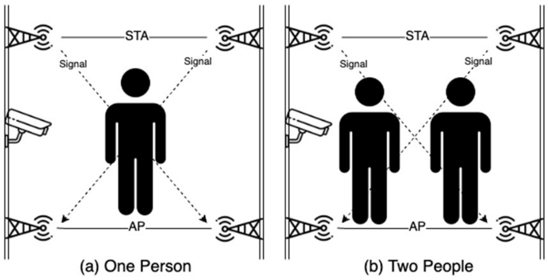


Fig. 4. Data collection environment diagram

2.2.2 Data Processing

CSI data provides high-resolution insights into the propagation characteristics of signals within a wireless channel. This data typically encompasses frequency domain features such as amplitude and phase. Because wireless signals experience multipath effects like reflection, refraction, and diffraction as they propagate through space, CSI is adept at capturing dynamic environmental changes and subtle channel perturbations caused by human activities (Cominelli et al., 2023). Consequently, CSI data can deliver detailed temporal and frequency characteristics useful for recognizing human activity patterns. However, due to its high-frequency variability and nonstationary nature, analyzing CSI data directly in the time or frequency domain can be challenging and may not fully reveal its underlying behavioral patterns. Therefore, in this study, we utilize time segments labeled with human activity characteristics and employ STFT techniques to transform the data. This approach allows us to extract effective time-frequency features, thereby supporting subsequent classification tasks.

2.2.2.1 Feature Labeling

Once the raw data has been collected, the first step is to label the data by marking the characteristics of crowd size, such as no people, one person, or two people, at the time points they appear. These labels are then categorized into arrays corresponding to the different crowd sizes for storage. Subsequently, based on the labeled time intervals, the corresponding CSI data is extracted and stored. To prevent imbalances in feature occurrence frequency that could affect predictive outcomes, we impose a data quantity limit on each feature to ensure a consistent sample size. This is essential for subsequent data processing and analysis.

2.2.2.2 Data Transformation

To address the non-stationary characteristics of CSI data and extract its time-frequency features, this study employs the STFT as the primary data preprocessing method (Baba, 2012). By applying a sliding window in the time domain, STFT segments the signal into localized intervals and performs the fourier transform on each. This results in a joint distribution of time and frequency, effectively revealing the frequency components that vary over time. This capability provides significant advantages in handling non-stationary signals and is particularly suitable for extracting the time-frequency features of human flow activities. The short-time Fourier transform is defined as:

$$X(t,f)=\int_{-\infty}^{\infty} x(\tau)\cdot w(\tau-t)\cdot e^{-j2\pi f\tau}d\tau \tag{1}$$

The formula referenced in (Allen et al., 1997). In this equation:

$x(\tau)$: the time-domain signal.

$w(\tau - t)$: the window function used to localize the signal in time.

$e^{-j2\pi f\tau}$: the kernel function of the Fourier transform.

The resulting $X(t, f)$ represents the energy distribution of the signal at time t and frequency f. This formula illustrates that the STFT decomposes a time-domain signal into its time-frequency domain representation. By adjusting the window function's length and type, a balance between time resolution and frequency resolution can be achieved.

2.2.3 Model Selection

This section introduces the models used in this research. Observing the characteristics of CSI data reveals that signal transmission is affected by factors such as multipath effects, reflection, attenuation, interference, and noise, resulting in high-dimensional and nonlinear properties. Thus, four machine learning classifiers were chosen for comparison: RandomForestClassifier, SVC, XGBClassifier and GradientBoostingClassifier. Each classifier possesses unique operating features, making them well-suited for addressing different types of issues in CSI signals. This study trains and compares models using each of these classifiers.

The first classifier, RandomForestClassifier, comprises multiple decision trees, each learning from a randomly selected subset of CSI features and then using a voting system among the trees to make the final decision. RandomForestClassifier achieves model diversity by resampling data multiple times during training, effectively identifying the CSI features most influential in classification.

SVC, known for its strong performance in high-dimensional data, classifies data by finding a decision boundary that maximizes the margin between classes. The model maps input data to a high-dimensional space and identifies an optimal plane that distinguishes different classes. Focusing on support vectors—data points near the decision boundary with the most significant impact on classification. SVC avoids overfitting, making it ideal for small-scale, high-dimensional CSI data classification.

Finally, XGBClassifier and GradientBoostingClassifier are both gradient-boosting-based classifiers, iteratively generating decision trees to improve classification accuracy. Despite sharing a common principle, they differ in implementation. XGBClassifier includes optimization features like automated missing-value handling, L1/L2 regularization, parallel computation, and GPU acceleration, making it suitable for large datasets and reducing overfitting risk. In contrast, GradientBoostingClassifier relies on manual hyperparameter tuning and manual missing value imputation, offering greater flexibility. Both models are well-suited for CSI data processing.

3. RESULTS AND DISCUSSION

This study conducted experiments on CSI human flow data using four different machine learning classifiers and compared their performance. The purpose of the experiments was to evaluate each classifier's accuracy and stability when handling CSI signals with high dimensionality, multipath effects, and nonlinearity. Classification was performed for scenarios with zero to two individuals passing through the corridor, and the predictive abilities of the models were quantified through confusion matrices and accuracy rates.

3.1 Results of STFT Feature Analysis

After pre-processing and transforming the raw data, we extracted three segments corresponding to different numbers of people passing through the observation area. These segments were analysed using time-frequency spectrograms. Fig. 5 presents the data from left to right, showing scenarios of zero, one, and two individuals passing, respectively. As observed in the figure, when no people are within the detection range, the signal exhibits minimal fluctuations and remains stable. In contrast, when one person is present, the signal shows more variations in both high and low frequencies compared to when no one is present. With two people when passing through the signals,

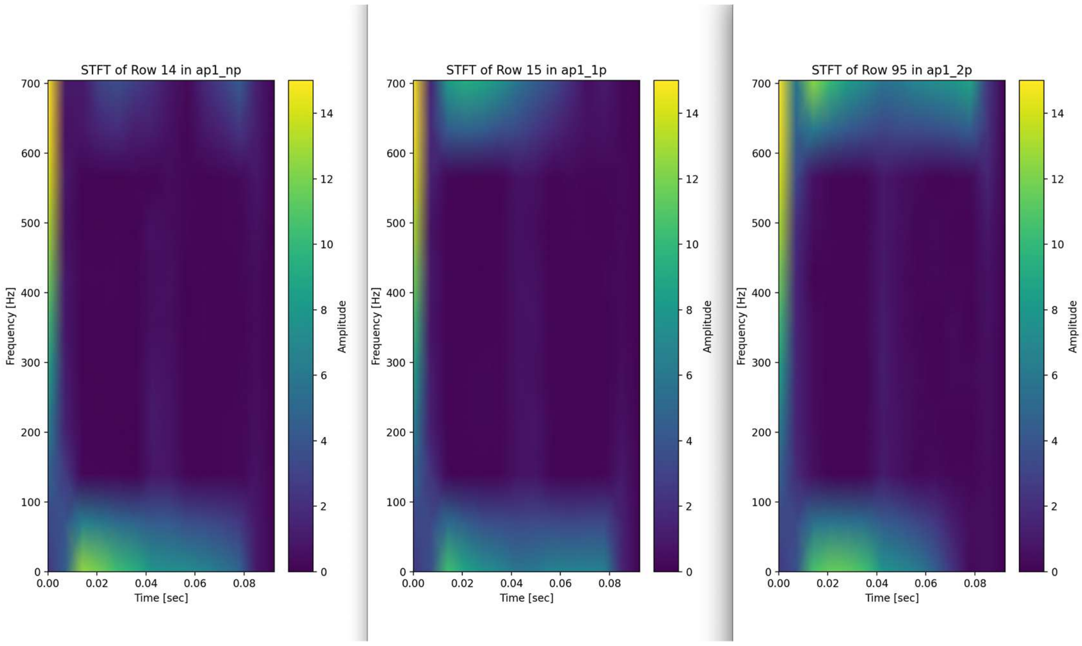


Fig. 5. Comparison of STFT spectrograms

becomes even more dynamic, with noticeably brighter colours in the high and low-frequency regions. These results indicate that the CSI data processed through STFT contain sufficient features for the model to learn effectively.

3.2 Accuracy Analysis

In this section, accuracy was used as the primary evaluation metric to compare the performance of four models RandomForestClassifier, SVC, XGBClassifier, and GradientBoostingClassifier under varying numbers of individuals and different data processing methods. The overall accuracy of each classifier summarized in Table 1.

Table 1. Transfer the sensor log data into training data

Model / Data	Raw data accuracy (%)	After STFT accuracy (%)
RandomForestClassifier	50	99
SVC	44	98
XGBClassifier	73	98
GradientBoostingClassifier	69	98

From the table, it can be observed that there is a notable difference in accuracy among the models when applied to raw CSI data. XGBClassifier achieved the highest performance with an accuracy of 73%, while SVC had the lowest at 44%. This suggests that with unprocessed CSI data, XGBoost's inherent capabilities to handle high-dimensional data and noise, as well as its automatic handling of missing values, allow it to perform relatively well. On the other hand, although SVC generally performs well with high-dimensional data, the noise and nonlinearity of CSI data may make it challenging for SVC to identify an appropriate boundary.

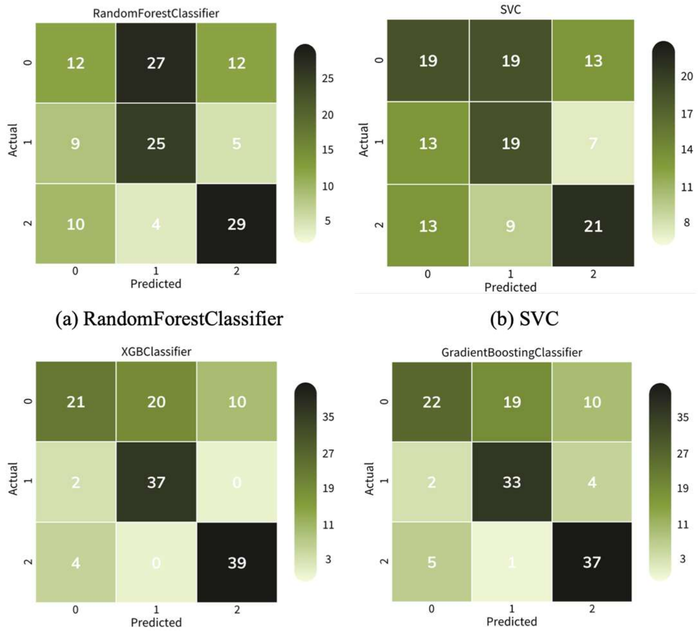
In order to further compare the performance of the proposed method in this study with existing technologies, we selected data processing and prediction methods from two related studies and applied them to the dataset used in this research. The first study is the machine learning-based approach proposed by Xiao et al. (2019), while the second is the deep learning-based approach introduced by Liu et al. (2017).

The method proposed by Xiao et al. (2019) uses a sliding window for noise reduction and extracts statistical features (skewness and kurtosis) as the basis for classification, employing a SVM for crowd detection. However, when applied to the dataset in this study, the accuracy achieved was only 29%, indicating limitations in handling the specific environment and data characteristics of this research.

On the other hand, Liu et al. (2017) method employs Butterworth filtering and phase correction for noise elimination, extracting phase and amplitude from CSI data as feature inputs, and utilizes a fully connected Backpropagation Neural Network (BP Neural Network) for crowd detection. When tested on this study's dataset, this method also demonstrated limited applicability, achieving an accuracy of just 30%, highlighting its inadequacies in addressing the unique environment and data characteristics of this research.

3.3 Confusion Matrix Analysis

In this section, confusion matrices are used to illustrate each model's performance in predicting different numbers of individuals. Confusion matrices of each model trained and tested on raw data are shown in Fig. 6.



(a) RandomForestClassifier (b) SVC

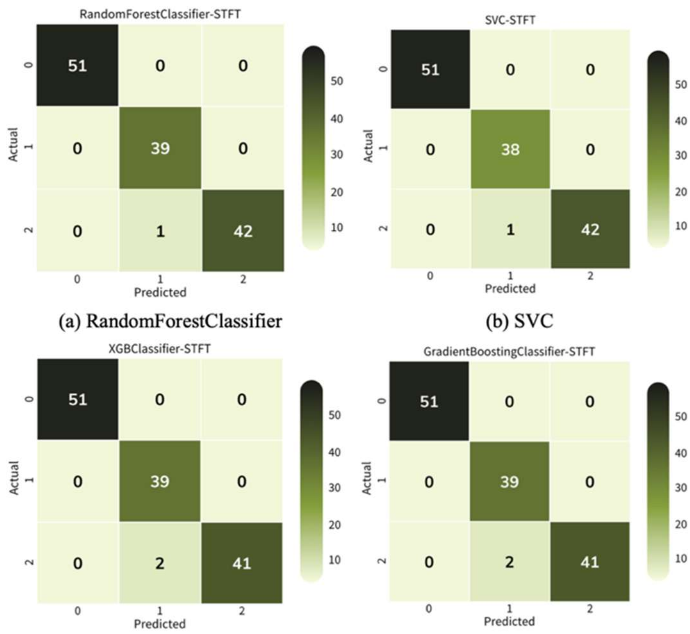
(c) XGBClassifier (d) GradientBoostingClassifier

Fig. 6. Confusion matrix of raw data

As seen in Fig. 6, although each model achieves a certain level of accuracy, they demonstrate varying performance levels due to differing adaptability to CSI data. While some models perform better than others, there is significant room for improvement overall. Therefore, we applied STFT to preprocess the data and then trained the models, resulting in the confusion matrices shown in Fig. 7.

From Fig. 7, it is clear that after training on STFT-processed data, the accuracy of all four models improved significantly. This is because the STFT processing effectively filtered out noise and interference from the raw data, making the essential and distinguishing features in the data more prominent and easier for the models to recognize.

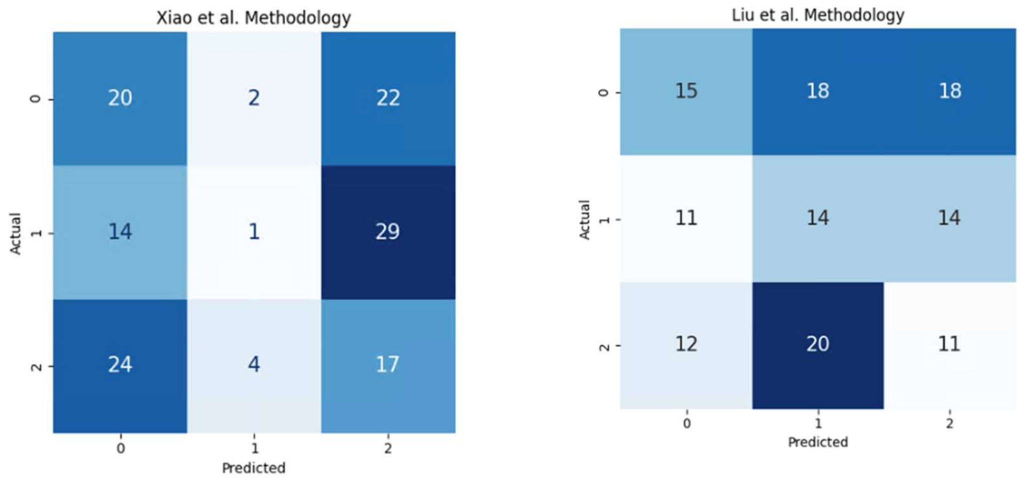
Next, we will present the results of applying the methods proposed by Xiao et al. (2019) and Liu et al. (2017) through confusion matrices, as shown in Fig. 8.



(a) RandomForestClassifier (b) SVC

(c) XGBClassifier (d) GradientBoostingClassifier

Fig. 7. Confusion matrix after STFT



(a) The method by Xiao et al. (b) The method by Liuet al.

Fig. 8. Confusion matrix after STFT

From Fig. 8, it can be observed that the recognition performance of these two methods is significantly poor when applied to the data collected under the equipment and environmental settings of this study. The confusion matrix reveals evident classification errors, particularly with higher error rates in cases with fewer individuals or boundary categories, resulting in low overall accuracy. This may be attributed to the noisy characteristics of the dataset and environmental variations, indicating that these two methods are unsuitable for the experimental environment and equipment used in this study.

Finally, we used three common quantitative metrics—Precision, Recall, and F1-Score, to summarize the results of all methods into a table. The table presents the Macro average values of these three metrics for comparison. The summarized results are shown in Table 2.

From Table 2, it can be observed that the machine learning method proposed by Xiao et al. (2019) achieved a Precision of only 25%, indicating a high proportion of misclassifications when identifying individuals, resulting in low accuracy. The Recall was only 29%, demonstrating that this method had low sensitivity for this classification task and failed to effectively detect most targets. The combined average F1-Score for Precision and Recall was also low at just 24%, suggesting that this method is unable to meet the requirements of this equipment and experimental environment.

Similarly, the deep learning method proposed by Liu et al. (2017) performed poorly in the application scenario of this study. It achieved a Precision of 31%, Recall of 30%, and an F1-Score of only 30%. Despite employing advanced deep learning techniques, this method failed to adapt to the lightweight equipment and data characteristics used in this study, leading to suboptimal performance and accuracy.

In contrast, in this study's experimental results, the RandomForestClassifier combined with STFT feature extraction on lightweight hardware achieved the best performance, with Precision, Recall, and F1-Score all reaching 99%. Other machine learning classifiers (SVC, XGBClassifier, GradientBoostingClassifier) also performed well, maintaining results at a high level of 98%, demonstrating the feasibility of using STFT combined with machine learning to detect CSI data in this environment.

These experimental results highlight that the application of the STFT technique in this study effectively filtered out environmental noise from the CSI data collected by lightweight devices and extracted distinguishable time-frequency features. This approach significantly improved the detection accuracy and stability of the models, showcasing the potential of combining lightweight hardware with optimized data processing techniques for CSI data applications.

Table 2. Quantitative metrics for each method

Method / Metrics	Precision (%)	Recall (%)	F1- Score (%)
STFT + RandomForestClassifier	99	99	99
STFT + SVC	98	98	98
STFT + XGBClassifier	98	98	98
STFT + GradientBoostingClassifier	98	98	98
Xiao method	25	29	24
Liu method	31	30	30

4. CONCLUSION

This study successfully designed and validated an efficient, lightweight, and cost-effective WiFi CSI-based human flow detection system centered on the ESP32 module. Compared to devices used in other studies, the adoption of ESP32 significantly reduced the system's hardware cost and size. By integrating STFT analysis with machine learning techniques, the system effectively filtered environmental noise and extracted key features for accurate human flow detection. Experimental results demonstrated that the system is particularly suitable for simple scenarios such as corridors and staircases, highlighting its advantages in privacy protection, low cost, and non-contact sensing. This work lays a solid foundation for the development of privacy-friendly human flow detection systems.

Future research can further enhance this lightweight device by integrating deep learning techniques, such as autoencoders or time-frequency domain fusion analysis, to improve its data representation capabilities and detection performance in more complex scenarios. Additionally, optimizations can be made for multi-target detection and dynamic environments, especially to address challenges posed by multipath effects and environmental variations. Overall, this study achieved an efficient application of WiFi CSI technology in human flow detection using low-cost, lightweight hardware and provides a reference for future advancements and extensions of CSI-based technologies.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENT

This work was supported by the National Science and Technology Council, R.O.C. under Grant NSTC 113-2622-E-025-001 - and NSTC 112-2222-E-025-002 -

REFERENCES

Allen, J.B., Rabiner, L.R. 1977. A unified approach to short-time fourier analysis and synthesis. Proceedings of the IEEE, 65, 1558–1564.

Azizi, S., Nair, G., Rabiee, R., Olofsson, T. 2020. Application of internet of things in academic buildings for space use efficiency using occupancy and booking data. Building and Environment, 186, 107355.

Baba, T. 2012. Time-frequency analysis using short-time fourier transform. The Open Acoustics Journal, 5, 1–10.

Chen, Z., Zhang, L., Jiang, C., Cao, Z., Cui, W. 2019. WiFi CSI-based passive human activity recognition using attention-based BLSTM. IEEE Transactions on Mobile Computing, 18, 2714–2724.

Cominelli, M., Gringoli, F., Restuccia, F. 2023. Exposing the CSI: A systematic investigation of CSI-based Wi-Fi sensing capabilities and limitations. Proceedings of the 2023 IEEE International Conference on Pervasive Computing, 1–8.

Das, S., Adhikary, A., Laghari, A.A., Mitra, S. 2023. Eldo-Care: EEG with kinect sensor based Telehealthcare for the disabled and the elderly. Neuroscience Informatics, 100130.

Daware, T., Dhote, T. 2023. Enhancing public safety through real-time crowd density analysis and management. Proceedings of the 5th International Conference on Inventive Research in Computing Applications (ICIRCA), 1040–1046.

Ding, J., Wang, Y. 2019. WiFi CSI-based human activity recognition using deep recurrent neural network. IEEE Access, 7, 174257–174269.

Espinosa, Á., Lopez Ruiz, J., Mata Mata, F., Espinilla, M. 2021. Application of IoT in healthcare: Keys to implementation of the sustainable development goals. Sensors, 21, 2330.

Fitwi, A., Chen, Y., Sun, H., Harrod, R. 2021. Estimating interpersonal distance and crowd density with a single-edge camera. Computers, 10, 143.

Ge, Y., Taha, A., Shah, S.A., Dashtipour, K., Zhu, S., Cooper, J., Imran, M.A. 2023. Contactless WiFi sensing and monitoring for future healthcare: Emerging trends, challenges, and opportunities. IEEE Reviews in Biomedical Engineering, 16, 171–191.

Jia, X., Feng, Q., Fan, T., Lei, Q. 2012. RFID technology and its applications in Internet of Things (IoT). IEEE Pervasive Computing, 5, 1282–1285.

Kwon, H., An, S., Lee, H.-Y., Cha, W.C., Kim, S., Cho, M., Kong, H.-J. 2024. Beyond surveillance: Privacy, ethics, and regulations in face recognition technology. Frontiers in Big Data, 7, 1337465.

Laghari, A.A., Yin, S. 2022. How to collect and interpret medical pictures captured in highly challenging environments that range from nanoscale to hyperspectral imaging. Current Medical Imaging, 20, 281222212228.

Laghari, A.A., Estrela, V.V., Li, H., Shoulin, Y., Khan, A.A., Anwar, M.S., Wahab, A., Bouraqia, K. 2024. Quality of experience assessment in virtual/augmented reality serious games for healthcare: A systematic literature review. Technology and Disability Preprint, 1–12.

Laghari, A.A., Sun, Y., Alhussein, M., Aurangzeb, K., Anwar, M.S., Rashid, M. 2023. Deep residual-dense network based on bidirectional recurrent neural network for atrial fibrillation detection. Scientific Reports, 13, 15109.

Lei, Z., Rong, B., Jiahao, C., Yonghong, Z. 2024. Smart city healthcare: Non-contact human respiratory monitoring with WiFi-CSI. IEEE Transactions on Consumer Electronics, 70, 333–341.

Lin, Q., Yue, Y. 2015. Device-free passive human detection using Wi-Fi technology: Current state and future trend. Proceedings of the IEEE 12th International Conference on Ubiquitous Intelligence and Computing, 1717–1723.

Liu, J., Liu, H., Chen, Y., Wang, Y., Wang, C. 2020. Wireless sensing for human activity: A survey. IEEE Communications Surveys and Tutorials, 22, 1629–1645.

Liu, M., Lin, Z., Xiao, P., Xiang, W. 2023. Human biometric signals monitoring based on WiFi channel state information using deep learning. arXiv preprint, arXiv:2203.03980.

Liu, S., Zhao, Y., Chen, B. 2017. WiCount: A deep learning approach for crowd counting using WiFi signals. Proceedings of the IEEE International Symposium on Parallel and Distributed Processing with Applications and IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), 967–974.

Ma, Y., Arshad, S., Muniraju, S., Torkildson, E., Rantala, E., Doppler, K., Zhou, G. 2021. Location and person-independent activity recognition with WiFi, deep neural networks, and reinforcement learning. ACM Transactions on Internet of Things, 2, 1–25.

Moshiri, P.F., Shahbazian, R., Nabati, M., Ghorashi, S.A. 2021. A CSI-based human activity recognition using deep learning. Sensors, 21, 7225.

Narayana, S., Prasad, R.V., Rao, V.S., Prabhakar, T.V., Kowshik, S.S., Iyer, M.S. 2015. PIR sensors: Characterization and novel localization technique. IEEE Sensors Journal, 14, 142–153.

Qian, K., Wu, C., Yang, Z., Liu, Y., Jamieson, K. 2017. Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi. Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 1–10.

Qian, K., Wu, C., Zhang, Y., Zhang, G., Yang, Z., Liu, Y. 2018. Widar2.0: Passive human tracking with a single Wi-Fi link. IEEE Sensors Journal, 16, 350–361.

Rangdale, S., Jadhav, N., Solav, S., Gayake, N., Nanware, S., Dipmala, S., Tekade, P. 2023. A survey on crowd detection and management using deep learning. Proceedings of the International Conference on Computer Vision and Internet of Things (ICCVIoT'23), Coimbatore, India, 224–232.

Shah, A.S., Nasir, H., Fayaz, M., Lajis, A., Shah, A. 2019. A review on energy consumption optimization techniques in IoT-based smart building environments. Information, 10, 108.

Shi, Z., Cheng, Q., Zhang, J.A., Xu, R.Y.D. 2022. Environment-robust WiFi-Based human activity recognition using enhanced CSI and Deep learning. IEEE Internet of Things Journal, 9, 24643–24654.

Sun, J., Bian, X., Li, M. 2024. Non-contact heart rate monitoring method based on Wi-Fi CSI signal. Sensors, 24, 2111.

Wang, C.-Y., Lin, F.-S. 2024. AI-driven privacy in elderly care: Developing a comprehensive solution for camera-based monitoring of older adults. Applied Sciences, 14, 4150.

Wang, D., Yang, J., Cui, W., Xie, L., Sun, S. 2021. Multimodal CSI-based human activity recognition using GANs. IEEE Internet of Things Journal, 8, 17345–17355.

Wang, X., Yang, C., Mao, S. 2020. Resilient respiration rate monitoring with realtime bimodal CSI data. IEEE Sensors Journal, 20, 10187–10198.

Want, R. 2006. An introduction to RFID technology. IEEE Pervasive Computing, 5, 25–33.

Winkler, T., Rinner, B. 2014. Security and privacy protection in visual sensor networks: A survey. ACM Computing Surveys, 47, 1–42.

Wu, D., Zeng, Y., Gao, R., Li, S., Li, Y., Shah, R.C., Zhang, D. 2023. WiTraj: Robust indoor motion tracking with WiFi signals. IEEE Transactions on Mobile Computing, 22, 3062–3078.

Xiao, F., Guo, Z., Ni, Y., Xie, X., Maharjan, S., Zhang, Y. 2019. Artificial intelligence empowered mobile sensing for human flow detection. IEEE Network, 33, 78–83.

Zeng, Y., Pathak, P. H., Mohapatra, P. 2016. WiWho: WiFi-based person identification in smart spaces. Proceedings of the 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), 1–10.

Zhang, J., Li, Y., Xiong, H., Dou, D., Miao, C., Zhang, D. 2022. HandGest: Hierarchical sensing for robust-in-the-air handwriting recognition with commodity WiFi devices. IEEE Internet of Things Journal, 9, 19529–19544.

Zhou, H., Zhang, Y., Temiz, M. 2023. High-resolution indoor sensing using channel state information of WiFi networks. Electronics, 12, 3931.