2025 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) | 979-8-3315-3553-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/PerComWorkshops65533.2025.00164

2025 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events
(PerCom Workshops)

SandDune: Single ANtenna Device for Detecting User's Natural Eating Habits

Shreyans Jain[†], Yash Bhisikar[†], Surjya Ghosh[†], Timothy J. Pierson[§], Sougata Sen[†]
[†]BITS Pilani Goa Campus, India, [§]Dartmouth College, NH, USA.

***Abstract*—Over the years, researchers have explored various approaches for capturing and monitoring the eating activity, one among which is via Wi-Fi channel state information (CSI). CSI-based approaches commonly rely on multi-antenna systems for the capturing and monitoring tasks. With the advent of low-cost, single-antenna IoT devices with CSI measuring capabilities, a question that arises is whether these inexpensive devices can monitor human activities? In this paper we present the *SandDune* system that demonstrates the possibility of monitoring one human activity – eating – using only inexpensive single-antenna Wi-Fi devices. SandDune is an infrastructure-based system that continuously monitors CSI information to detect the eating activity occurring in its vicinity. When it detects an eating activity, it scrutinizes the signals further to identify all hand-to-mouth eating gestures in the eating episode. We tested SandDune and observed that SandDune can distinguish eating from other activities with an F1-score of 85.54%. Furthermore, it can detect the number of hand-to-mouth gestures that occurred in the eating episode with an error of ±3 gestures. Overall, we believe that a SandDune-like system can enable low cost, unobtrusive eating activity detection and monitoring with potential use-cases in several health and well-being applications.**

***Index Terms*—Wi-Fi CSI, Single Antenna System, ESP32-S**

I. Introduction

There has been a recent interest in using the pervasive RF channel (e.g., Wi-Fi) for sensing various human activities [1]. The RF channel properties are studied and monitored using Channel State Information (CSI). Specifically, CSI provides an understanding of the characteristics of the communication channel as the signal propagates from a wireless transmitter to a receiver [2]. Since Halperin's tool release in 2011 [3], CSI has been available for Wi-Fi. Because of Wi-Fi's ubiquity and availability in numerous wearable and Internet of Things (IoT) devices, researchers have used these devices for various human activity recognition (HAR) tasks, including the eating detection task [4]. CSI-based eating detection primarily focuses on using expensive multi-antenna devices [5]. Because of their size and cost, such systems might not be usable in everyday scenarios. Thus, low-cost systems are needed that can be deployed to pervasively and unobtrusively capture information related to a person's eating habits in everyday scenarios. To fulfill this need, we have developed *a low-cost, usable, ubiquitous Wi-Fi-based system – SandDune – that uses CSI information to detect fine-grained eating-related details.*

The SandDune system consists of two or more low-cost (each costing less than $5) ESP32-S devices [6]. One device transmits a Wi-Fi frame, and one device receives the frame. The transmitter-receiver pair are placed 1 meter apart. The receiver continuously captures the transmitted Wi-Fi packets to extract the channel state information. The receiver then passes this data to an *eating detector* module that determines whether a person is eating. When the module detects eating, the temporal data is passed to the *H2M Detector*, which determines the number of hand-to-mouth gestures in the eating episode. The system details of SandDune is presented in Figure 1.

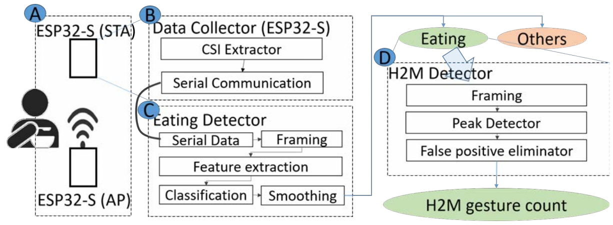

Fig. 1: High-level system overview of SandDune. SandDune consists of a coarse-grained eating detection module, and a fine-grained hand-to-mouth (H2M) gesture counting module.

Developing a CSI-based eating detection system has its challenges. First, having multiple antennas at the transmitter and receiver allows capturing information in spatial, temporal, and frequency domains [4]. A single-antenna device can only rely on temporal and frequency variations to detect the eating activity. Second, a person might perform various activities near the devices. SandDune should be capable of distinguishing eating from these other activities. Third, individuals might have diverse postures while eating and can consume food at different speeds. SandDune should be robust to such differences.

We have developed SandDune, a single-antenna device-based fine-grained eating recognition system that addresses these challenges. In this paper, we aim to answer the following research questions: **RQ1**: Can we use SandDune to distinguish the eating activity accurately from other activities? and **RQ2**: Can we use the detected eating moments captured to compute the count of hand-to-mouth gestures? While addressing these challenges we make the following **key contributions**:

- We have developed the SandDune system that is capable of capturing fine-grained details of the eating activity using CSI data. In this paper, we describe the design details and the system-level choices made while realizing the system.
- We conducted a user study to determine the feasibility of detecting the eating activity and obtaining fine-grained details of the eating episode. Overall, we observed that we could distinguish eating from other activities with a

2766-8576/25/$31.00 ©2025 IEEE
DOI 10.1109/PerComWorkshops65533.2025.00164

F1-score of 85.54%. Within the eating activity, we could detect the hand-to-mouth feeding gestures with an error of ±3 gestures in an eating episode.

II. Related Work

To detect the eating activity automatically, researchers have experimented with various techniques, including the Wi-Fi based approach [5], [7]–[9]. The possibility of extracting CSI using various toolkits has increased the number of Wi-Fi CSI-based sensing applications [3]. This is further proliferated due to easy access to hardware such as the (now discontinued) Intel 5300 Wi-Fi card that provides CSI. Researchers have used Wi-Fi CSI for detecting activities that the humans are performing [4]. These activities are either general everyday physical activities or more complex activities such as shopping activity monitoring [10]. Cominelli et al. studied CSI-based Wi-Fi to understand sensing capabilities and limitations [11]. They explored several prior studies that performed human activity recognition, and then proposed a common multi-activity dataset. This dataset did not cover the eating activity, however. Lin et al.'s work on Wi-Fi CSI-based eating activity monitoring demonstrates the possibility of detecting the eating activity, and extracting fine-grained details about the activity [5]. Their work, however, relies on a multiple-antenna setup, along with a smartphone for the detection. Deployment complexity of the system and its cost makes it difficult to deploy this system in a free-living study. Researchers have also explored expensive multi-modal sensing approaches for HAR, including CSI [12]. In comparison to these systems, SandDune, with two single-antenna microcontrollers is a low cost eating detection solution.

III. SandDune: Design and Working

The architecture of the eating detection system – SandDune – is shown in Figure 1. We envision the system to consist of multiple pairs of low-cost, single-antenna microcontroller devices, the ESP32-S, capable of collecting CSI continuously from the Wi-Fi signal. One such pair is shown in block Ⓐ.

The ESP32-S devices' Wi-Fi connectivity can be configured in two modes – Wi-Fi Access Point (WiFiAP) mode and Wi-Fi Station Device (WiFiSTA) mode. In the WiFiAP mode, the device can receive incoming connections from other devices. In SandDune, for any pair of ESP32-S devices, one ESP32-S is configured as a WiFiAP – the *AP device*, and the other ESP32-S device as a WiFiSTA – the *STA device*. The WiFiSTA device is attached to a more powerful edge device – a laptop. It can extract the CSI data and transfer the data to the laptop. Block Ⓑ presents the details of the WiFiSTA module.

The edge device executes the eating detector module (a classic machine learning pipeline) on the CSI data to determine whether a frame represents the eating activity. SandDune currently uses a shallow learning model with hand-crafted features for eating detection at a frame level. Multiple continuous eating frames are collected together to determine an eating episode. Block Ⓒ in Figure 1 presents the eating detection module.

An eating episode is then passed over to the H2M detector module (hand-to-mouth detector module). This module analyzes the amplitude of various subcarriers of the signal. It performs peak detection on the data (a peak represents the feeding activity). Via this module, the system obtains the fine-grained details of the eating activity. This information can then be passed to the concerned person (e.g., the individual, or a clinician). Block Ⓓ in Figure 1 presents the H2M detector.

IV. Methodology

SandDune's goal is to detect fine-grained details of the eating activity that occurs between the ESP32-S device pairs.

Data Collector: The Espressif IoT Development Framework (ESP-IDF) allows programs to obtain the CSI data. SandDune uses a modified version of Hernandez et al.'s toolkit for collecting the CSI data [13]. The toolkit provides 128 non-STBC HT-LTF subcarrier CSI information in the 40 MHz channel, from which we used data from the 114 important subcarriers between $+57$ and -57 subcarriers.

The toolkit allows configuring the WiFiAP and WiFiSTA in both active transmitting and passively receiving modes. The device configured as WiFiSTA continuously sends requests to the WiFiAP device at 200 frames per second. On receiving the request, the WiFiAP sends a frame with pre-determined pilot symbols. The WiFiSTA captures this frame and extracts the CSI data. The data received by the WiFiSTA is transmitted through the serial port to a more powerful computing device, where it is further processed.

Eating Detector: A code running on the computing device extracts the phase and amplitude information from the raw CSI data for each subcarrier. We observed that we could not derive any meaningful information from the phase information, and thus SandDune does not use it for further computation. Instead, SandDune uses only amplitude.

Framing: Human activities have a temporal aspect. Creating windows of sequential data allows capturing this temporal aspect of the activity. SandDune uses the data from the S subcarriers to create F windows of n seconds each with $k\%$ overlap between windows. We use the value $S = 114$, $n = 5$, and $k = 50\%$ in our implementation. Thus, at the end of this framing step, SandDune obtains F windows of length 5 seconds, and width of 114 subcarriers.

Subcarrier-wise feature extraction: Next, SandDune computes four statistical features of mean $(\overline{x})$, median $(\widetilde{x})$, standard deviation (σ), and slope (m) for each subcarrier of a window $f \in F$ separately. Thus, at the end of this step, SandDune obtains a vector of size $114 \times 4 = 456$ features for each f. It uses these 456 features for further computation.

Classification: The prepared data is then fed into the classification module that performs a binary classification between eating and other activities. We experimented with both shallow learning and deep learning approaches to evaluate SandDune. For shallow learning, we used Random Forests, Support Vector Machines and XGBoost. The scikit-learn's implementation of Random Forest and Support Vector Machine (tested with four different kernels – linear, sigmoid, polynomial and radial-basis function(RBF)) was utilized. For the deep learning model, we implemented an Multilayer Perceptron (MLP) using building blocks from scikit-learn. MLP is a feedforward artificial neural networks consisting of an input layer, one or more hidden layers, and an output layer. Each layer contains neurons with nonlinear activation functions. Our model consisted of 5 layers, with first input hidden layer of size 256, followed by hidden layer sizes of 128, 64 and 16 respectively. ReLU activations is applied on the outputs.

Smoothing: Eating is a longitudinal process; a person might perform non-eating gestures in between the eating gestures. The process of smoothing allows detecting the entire eating episode duration, even when non-eating gestures occur in between the eating gestures. $\forall f \in \{x, ..., y | y > x\}$ if at least $t\%$ windows are classified as eating, SandDune identifies the entire duration $[f_x,\ f_y]$ windows as eating.

Hand-to-mouth (H2M) detector: All $[f_x,\ f_y]$ windows identified as 'eating windows' are passed to the H2M detector module. H2M aims to determine the number of hand-to-mouth gestures that occur during the eating episode. We observed that the hand-to-mouth gesture resulted in more disturbance to the subcarriers, causing observable peaks in the amplitude of the signal. Thus, we suspect (and as evident from our evaluation in Section VI) that counting the number of peaks would allow detecting the number of hand-to-mouth gestures.

SandDune computes the amplitude peaks for the 114 subcarriers. It uses an off-the-shelf peak detection algorithm that returned peaks in each subcarrier signal, based on the height threshold and prominence [14]. The height threshold for each user is calculated as the mean of the amplitude values computed by applying a simple moving average to a window size empirically chosen for our data set.

To determine peaks in the signal, the peak detection algorithm first identifies local maxima and retrieves their indices, along with the left and right edges of these peaks. It applies a height threshold to focus on significant peaks, effectively filtering out those caused by noise. Next, the algorithm calculates the prominence of the remaining peaks and further filters them using a prominence threshold of 1.5 (chosen empirically for our dataset). Finally, it ensures that the peaks are separated by the defined minimum distance before returning the indices of the remaining peaks along with their associated properties including $peakheights$, $prominences$, $leftbases$ and $rightbases$. However, sometimes two peaks might reside close by in time. SandDune groups all those peaks which occurs within a threshold distance together into one single eating gesture. At the end of this step, SandDune returns the number of peaks in each subcarrier.

SandDune determines the number of peaks in the eating episode by two approaches – the mean peak approach, and the median peak approach. For the mean peak approach, it computes the statistical mean of the peak output of the 114 subcarriers, while for the median peak approach, it computes the median of those subcarriers. Overall, SandDune determines the number of episode-wise eating gestures based on the output of the mean peak approach or the median peak approach, as discussed in Section VI.

V. Data collection and Evaluation approach

We conducted a controlled study where participants performed various everyday activities (Eating, Sitting, Walking, and Using Phone) in a controlled setting. Overall, we collected 594,734 data frames in the user studies.

A. Deployment and Data Collection

Data collection was performed in a controlled setting – a 1.1 meter × 1.4 meter table in the home of one of the authors. We placed two ESP32-S at a fixed position on the table, separated from each other by 1 meter; both placed 10 centimeters from the edge of the table and powered by the USB port of a laptop. One ESP32-S executed the WiFiAP code, while the other executed the WiFiSTA code. A script running on the laptop continuously gathered data from the serial port to which the WiFiSTA was connected. The same script also started the camera of the laptop so that we could capture the video of the person for ground truth purposes.

For the study, we recruited 10 participants (9 right handed, aged between 25 and 30 years). The participants visited the location where the SandDune system was deployed. Participants were instructed that they would perform four activities – sit comfortably on the chair between the deployed ESP32-S pair, use a phone while sitting on the chair, walk in the vicinity of the two ESP32-S, and eat rice or noodles from a plate using a spoon. Participants performed each activity for 7 minutes. We, however, removed the first and last 30 seconds of data to filter out the disturbances caused from movements to start and stop the experiment, resulting in a 24 minutes of data collection for each study participant. From the laptop's videos, we observed that one participant did not eat with the spoon, but rather fed themselves using their hand. We discarded that participant's data, leaving us with data from 9 participants.

B. Evaluation Technique and Metric

RQ1: To determine whether SandDune could distinguish eating from not eating, we performed leave one person out cross validation using Synthetic Minority Over-sampling TEchnique (SMOTE) because there was a class imbalance – our dataset consists of 25% eating activity and 75% not-eating activities. We report the F1-score, precision and recall for eating activity detection in Section VI.

RQ2: To determine the number of hand to mouth gestures, we computed the mean and median of peaks identified by the 114 subcarriers. We compute the difference between the ground truth (obtained from the video data) and the mean/median peak to determine the difference in performance.

VI. Results

As mentioned in Section IV, we collected the data obtained from the receiving ESP32-S. Empirically, we observed that the receiver received an average of 60 readings in a second (of the 200 frames transmitted).

Classifier	% Precision	%Recall	%F1-Score
SVM (Linear)	**87.24**	**86.51**	**85.54**
SVM (Poly)	82.18	81.73	79.81
SVM (RBF)	80.05	80.26	75.54
SVM (Sigmoid)	63.22	62.32	55.25
Random Forest	85.31	80.42	77.12
XGBoost	84.10	78.46	74.33
Multi Layer Perceptron	83.5	76.64	73.02

TABLE I: Performance metrics for classifiers in leave-one-person-out validation using SandDune.

1) [RQ1] Eating versus other activities: We performed a leave-one-person-out cross validation on the dataset to determine the performance. In our dataset, we have 25% eating activity data, and 75% other activity data. After applying the SMOTE technique, our model's class distribution was balanced.

To evaluate the performance of classifier we created windows of 5 seconds each and formed sliding windows of 30 seconds. For the smoothing step, we used the threshold value as $t = 20\%$ i.e. if number of eating instances are more than the threshold that window is classified as eating. Table I presents the performance of various classifiers for distinguishing eating from other activities using a leave one person out cross validation. Out of the shallow learning classifiers, SVM with linear kernel performs best for identifying eating activity with precision 87.24%, recall 86.51%, and F1-score 85.54%. We used the MLP deep learning classifier and it resulted in precision 83.5%, recall 76.64%, and F1-score 73.02%. The lower performance might be because of the small dataset or because of the currently chosen architecture for model building.

2) [RQ2] Detecting number of hand-to-mouth gestures: We passed all sub-carriers of each eating episode in the dataset to the peak detection algorithm to estimate the number of peaks. As mentioned previously, we computed two types of peaks – the mean and median peaks. Figure 2 presents the estimated (both mean-peak and median-peak) and actual hand-to-mouth gestures detected for each participant. From the figure it is visible that the median number of peaks detected for a participant for all the 114 subcarriers is more comparable to the number of actual gestures. The mean absolute error for number of peaks estimated using the median peaks approach is ±3, indicating that it is indeed possible to use a single antenna device to detect fine-grained details of the eating activity.

VII. Conclusion

In this paper, we demonstrate that rather than using expensive and complex systems, low-cost, single-antenna microcontrollers can detect the eating activity. We present the design of our developed system, SandDune. Overall, we observed that CSI data from a single-antenna device can enable detecting the eating activity with an F1-score of 85.54%. Furthermore, the system can count the number of hand-to-mouth gestures with an error of ±3 gestures. This shows that such a system can enable low cost unobtrusive eating activity monitoring. In future, we will explore approaches to improve the performance of the system by deploying multiple devices.

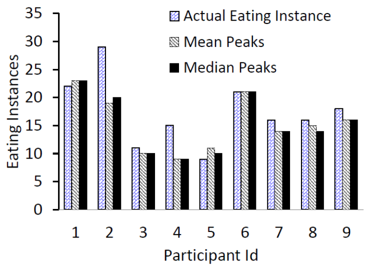

Fig. 2: Comparison of median peak detection approach and mean peak detection approach with the ground truth.

Acknowledgment

This research results partly from BITS Pilani's Cross Disciplinary Research Projects C1/23/173 and C1/23/152, and partly from the SPLICE research program, supported by a collaborative award from the SaTC Frontiers program at the National Science Foundation under award number CNS-1955805. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors. Any mention of specific companies or products does not imply any endorsement by the authors, by their employers, or by the sponsors.

References

[1] H. F. T. Ahmed, H. Ahmad, and C. Aravind, "Device free human gesture recognition using Wi-Fi CSI: A survey," *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103281, 2020.

[2] T. J. Pierson, T. Peters, R. Peterson, and D. Kotz, "Closetalker: Secure, short-range ad hoc wireless communication," in *MobiSys*, ACM, 2019.

[3] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication review*, vol. 41, no. 1, pp. 53–53, 2011.

[4] Y. Ma, G. Zhou, and S. Wang, "WiFi Sensing with Channel State Information: A Survey," *ACM Computing Surveys*, vol. 52, Jun 2019.

[5] Z. Lin *et al.*, "WiEat: Fine-grained device-free eating monitoring leveraging Wi-Fi signals," in *International Conference on Computer Communications and Networks (ICCCN)*, 2020.

[6] Espressif, "Esp32 modules." https://www.espressif.com/en/products/devKits. Accessed 2024-12-31.

[7] S. Sen, V. Subbaraju, A. Misra, R. Balan, and Y. Lee, "Annapurna: An automated smartwatch-based eating detection and food journaling system," *Pervasive and Mobile Computing*, vol. 68, p. 101259, 2020.

[8] S. Bi *et al.*, "Auracle: Detecting eating episodes with an ear-mounted sensor," *Proceedings of ACM on Interactive, Mobile, Wearable, and Ubiquitous Technology (IMWUT).*, vol. 2, sep 2018.

[9] T. Vu, F. Lin, N. Alshurafa, and W. Xu, "Wearable food intake monitoring technologies: A comprehensive review," *Computers*, vol. 6, no. 1, 2017.

[10] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Analyzing shopper's behavior through wifi signals," in *Proceedings of the 2nd Workshop on Workshop on Physical Analytics*, WPA '15, p. 13–18, 2015.

[11] M. Cominelli, F. Gringoli, and F. Restuccia, "Exposing the CSI: A Systematic Investigation of CSI-based Wi-Fi Sensing Capabilities and Limitations," in *Internation Conference on Pervasive Computing & Communications (PerCom)*, IEEE, 2023.

[12] V. K. Singh *et al.*, "WiFiTuned: Monitoring Engagement in Online Participation by Harmonizing WiFi and Audio," ICMI, 2023.

[13] S. M. Hernandez and E. Bulut, "Lightweight and Standalone IoT Based WiFi Sensing for Active Repositioning and Mobility," in *"A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, 2020.

[14] The Scipy Community, "Prominence Based Peak Detection." https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html [Accessed: 2024-10-02].